The Role of Visual Scenes in Spoken Language Comprehension: Evidence from Eye-Tracking

Dissertation
zur Erlangung des akademischen Grades eines
Doktors der Philosophie
der Philosophischen Fakultäten
der Universität des Saarlandes
vorgelegt von
Pia Stefanie Knoeferle
aus Ingolstadt

Saarbrücken, 2005

Dekan: Univ.-Prof. Dr. Dr. h.c. Wolfgang Schweickard

Berichterstatter: Prof. Dr. Matthew W. Crocker, Prof. Martin J. Pickering, Prof. Gerry T. M. Altmann

Tag der letzten Prüfungsleistung: 18.02.2005

Contents

List of Figures

List of Tables

Abstract

How do we understand utterances incrementally in visual environments? As a first hypothesis, we might assume that the presence of visual environments does not affect core language comprehension processes such as the incremental structuring of an utterance (Fodor, 1983). Crucially, however, language refers to things in the world. It has further been demonstrated that reference to entities is established rapidly during comprehension, and that referentially relevant information in the scene rapidly influences how the linguistic input is structured (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995).

While Tanenhaus et al. (1995) provide convincing evidence for the influence of scene information on comprehension, their findings are limited to scenes containing simple objects. We know little about what other kinds of information in scenes can affect on-line language comprehension. As a result, we might not have a complete view of the influence of scene information on comprehension. A better understanding of its influence is, however, the first step in specifying how comprehension proceeds in a setting that provides both linguistic and scene information.

The primary goal of this thesis is thus to examine whether the existing insights concerning the role of scene information during comprehension generalize and extend to other types of information in scenes. To achieve this we explore the effects that depicted actions and events can have on the on-line structuring and incremental thematic interpretation of an utterance. We propose that depicted event scenes might play an even more important role in comprehension than previously examined referential contexts, since depicted event scenes provide propositional in addition to referential information.

Investigating the influence of various types of information in scenes on comprehension is an important first step in specifying how comprehension proceeds in environments that provide both scene and utterance as informational sources. Alone, however, it reveals little about how scene information is incrementally integrated with linguistic/world knowl-

edge and with the utterance. The secondary goal of this thesis is thus a more detailed specification of the *precise interplay* between the visual perceptual system and utterance comprehension.

To investigate these issues, we have carried out a series of head-mounted eye-tracking experiments that examined the incremental comprehension of spoken German and English sentences in visual scenes. Taken together, the findings from our studies provide support for the claim that depicted event scenes play a highly important role for spoken language comprehension. Importantly, their influence was not limited to one language or sentence type.

In more detail, the results from the individual experiments have revealed the following: Findings from four studies demonstrated a rapid influence of depicted events on structural disambiguation and incremental thematic role-assignment in structurally ambiguous German and English sentences. Concerning the nature of the interplay between scene perception and utterance comprehension we observed a close temporal coordination between when a cue in the utterance identified a depicted action, and when perception of the action and its associated thematic relations enabled structural disambiguation of the utterance. This finding proved robust both when the verb (an important cue in making available the depicted actions and events for comprehension) and when only soft linguistic cues could be used to identify the depicted events.

A fifth experiment explored the *relative importance* of depicted events in comparison with another important, linguistic type of information (stored thematic role knowledge). Our findings revealed a greater importance of depicted events relative to stored thematic role knowledge for incremental thematic role-assignment. This confirms the strong influence of depicted events on comprehension. We draw, however, attention to the fact that the effects of depicted events crucially depended upon the utterance that made them available for comprehension processes in the first place. Based on these findings, we provide a detailed sketch of how to develop an account of on-line sentence comprehension that characterizes the dynamic coordination between scene perception and incremental language comprehension.

Acknowledgements

Writing a thesis is an exploratory journey of the mind. Many interesting discoveries on the way to the final draft are made by deviating from the straight path and topic of the thesis. At the same time, you know that there is a path, journey, and goal lying ahead of you - a thought which ensures that you do not go to far astray. In the course of this academic journey, many people have contributed ideas and comments to the present thesis. I would like to thank all of them for making it a wonderful and enriching enterprise.

First and foremost I would like to thank my primary supervisor, Matthew W. Crocker for his unwavering support, advice, and patience. There was always a moment (and often much longer) to spare for discussing questions that arose in the course of my work. Many of these discussions helped in solving problems and ensured that deviations from the main path did not take longer than desirable on a three-year journey with a goal to reach. My work has certainly greatly profited from these conversations.

Many thanks go also to my second supervisor, Martin J. Pickering who passed on many intellectually stimulating ideas and was always available for questions despite the geographical distance between Saarbruecken and Edinburgh which email helped to bridge. In addition to frequent email exchanges between these two cities, one of the positive aspects of my PhD was a 6-month stay in the Department of Psychology at the University of Edinburgh. I experienced this stay as pleasant and would like to thank people at the Department of Psychology for their warm welcome that made my stay both an intellectually enriching and socially enjoyable experience. Among them are Jacqueline Thomson, Jamie Ward, Sandie Clelland, Scott McDonald, Martin Corley, Sarah Heath, Jools Simner, Ciara Catchpole, Mikihiro Tanaka, Susanna Flett, and Annabel Harrison.

Further thanks go to people in the Department of Psychology at Glasgow University and in Dundee where I ran experiments that are part of this PhD thesis. In particular I have profited greatly from help by Christoph Scheepers and Patrick Sturt who ensured

laboratory access for conducting the experiments and provided insightful comments on the design and analysis of individual experiments. Thanks go also to Simon Garrod, Lorna Morrow, and Roger van Gompel. People in the broader psycholinguistic community have also at various times provided helpful comments. Among them are Gerry Altmann, Yuki Kamide, and Mike Tanenhaus. Among the people in the psycholinguistics group at Saarland University my thanks go to Claudia Verburg for taking care of many administrative chores and to Kerstin Hadelich for a highly enjoyable office atmosphere. Furthermore, the International Graduate Colleague of which I was a member for the duration of my PhD has provided an interdisciplinary and stimulating research environment. I would moreover like to acknowledge the financial support that has made the current work possible: a PhD scholarship awarded by the German Research Foundation (DFG).

Last but not least the continual support of my parents has been an invaluable encouragement in pursuing the present research. They made sure that I took breaks from work and they frequently spent 10 hours in the car to from Ingolstadt to Saarbruecken and back in one day just to take me out for lunch and an afternoon coffee.

CHAPTER 1

INTRODUCTION

In many cases, people will find themselves in situations where both spoken language and an immediate scene context are available and relevant. Imagine watching a movie where a train is speeding along. The engine driver is inattentive while in charge of the train. When approaching a station it escapes his notice that a car is sitting on the rails. The shrieking of an elderly lady *Look out, there is a red ...* draws his attention to the rails. One glance and before the lady even finishes her exclamation (*... car sitting on the rails!*), perceptual processing might already have communicated the relevant threatening event (a car sitting on the rails) to the driver's comprehension system. He pulls the brake and the train narrowly avoids collision with the car. Alternatively, the driver might wait for the lady to finish her utterance before taking heed of the scene. In this case he would probably pull the brake too late, and the train would collide with the car.

Clearly, cognitive processes such as understanding an unfolding utterance and perceiving an immediate scene may occur simultaneously. How does utterance comprehension proceed when an utterance is about the immediate scene? Do listeners make rapid use of the immediate scene in incrementally understanding an unfolding utterance?

1.1 FROM MODULARITY TO INTERACTION

There are at least two answers to the above questions. One is that listeners might be able to make rapid use of scene information during incremental understanding when a scene is available and relevant. Alternatively, it is possible that scene information cannot directly influence the ongoing processing of the language system.

To further investigate these possibilities we consider factors that might determine whether perceptual processing of scene information can influence language comprehension. Comprehension in the situation sketched above probably involves the combination of informational sources (utterance and scene) that are perceived and processed by distinct cognitive systems such as language and vision. Prior research has suggested that whether utterance and scene can rapidly combine depends to a great extent on the *architecture* (i.e., the static arrangement) of the systems that perceive and process utterance and scene (Fodor, 1983) (for a more detailed discussion of the term 'architecture' see, e.g., Crocker, 1999; Pickering, Clifton, & Crocker, 2000).

Fodor (1983) postulated strong architectural restrictions on the informational interaction between distinct cognitive systems such as language and vision. He proposed that the mind was organized into transducers, input modules, and a central system. An illustrative sketch of this type of architecture is provided in Figure 1.1 (see Coltheart, 1999). Transducers convert physical stimuli into neural activity. This neural signal is passed on to the input modules that interpret the transduced signal. Examples for input modules are, for instance, the language or visual perception systems. Importantly, the characterizing property of input systems is that they are modular. While the concept of modularity is defined by several characteristics, one of the most important properties of a modular system is, according to Fodor, that it is *informationally encapsulated* (Fodor, 1983, p. 71). This means that distinct input modules such as language and visual perception have only access to the *output* of another distinct module, but cannot influence its *internal processes*. In Figure 1.1 this is indicated by the lack of direct arrows between the input modules.

In consequence, in a Fodorian framework the answer to the question of whether scene information can influence the incremental interpretation of an unfolding utterance is that the output of the visual perceptual system can only combine with the output of the language system. Ongoing scene perception cannot, however, directly influence core linguistic processes that are internal to the language system such as the incremental syntactic structuring of an utterance. Rather, Fodor considered the structuring of an utterance as an informationally encapsulated process. As a result of such an assumption, direct consideration of the influence of scene perception on such core comprehension processes is precluded.

We suggest that under the influence of a Fodorian view of the mind, many psycholinguistic theories of on-line language comprehension have been developed on the basis of the assumption that the mechanisms underlying language comprehension can be examined in

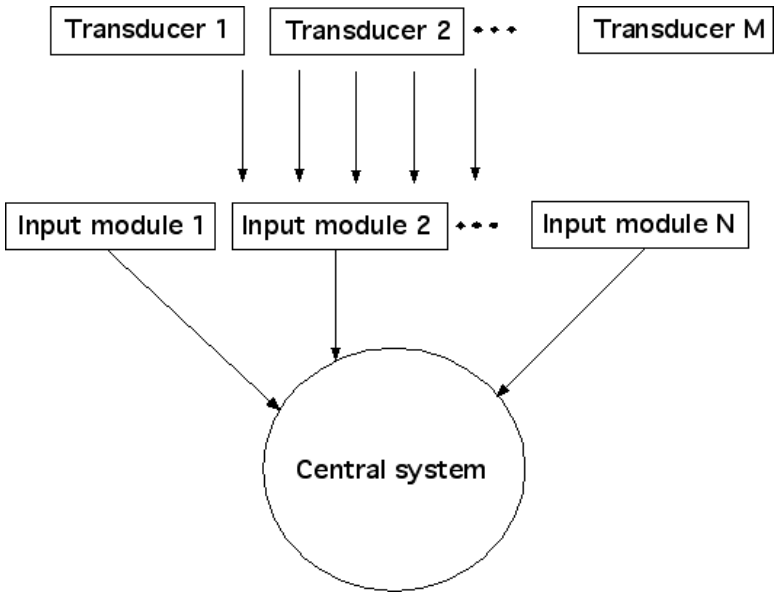


Figure 1.1: *Schematic sketch of the Fodorian organization of the mind (Coltheart, 1999, p. 116)*

isolation from the perception of scene information. Scene information has, for instance, not been explicitly included in most psycholinguistic theories or frameworks of on-line language comprehension (e.g., Crocker, 1996; Forster, 1979; Frazier & Fodor, 1979; Frazier & Clifton, 1996; E. Gibson, 1998; Gorrell, 1995; MacDonald, Pearlmutter, & Seidenberg, 1994; Mitchell, Cuetos, Corley, & Brysbaert, 1995; Pickering, Traxler, & Crocker, 2000; Townsend & Bever, 2001; Trueswell, Tanenhaus, & Kello, 1993; Trueswell, Tanenhaus, & Garnsey, 1994). Furthermore, the visual perceptual system has not been explicitly included as a cognitive system that might proffer important information for comprehension processes in these approaches, which are therefore unable to account for its potential influence. We argue that as a result such theories do not provide a complete account of language comprehension for situations in which language is about the scene.

A sub-group of these theories *by definition* implicitly exclude the influence of scene information on the initial structuring of a sentence through restricting the informational sources that can influence this process to syntactic information (e.g., Crocker, 1996; Frazier & Fodor, 1979; Frazier, 1987; Gorrell, 1995) or coarse-grained structural frequencies (e.g.,

Mitchell et al., 1995). In contrast to this restricted position (see Pickering, Clifton, & Crocker, 2000), scene information could, in principle, influence the initial structuring of a sentence in unrestricted interactionist frameworks (e.g., MacDonald et al., 1994; Trueswell et al., 1993, 1994). Such frameworks propose that any available and relevant informational source can influence the initial structuring of a sentence. However, many unrestricted interactionist accounts - just as restricted psycholinguistic theories - do not explicitly include scene information as an informational source. More generally, there has been relatively little investigation of how language comprehension proceeds in situations in which language is about the immediate scene, and where scene information is relevant for comprehension.

Only a few interactionist accounts of on-line sentence comprehension *explicitly* include scene information as an informational source and characterize an alternative view to Fodorian modularism (e.g., Tanenhaus et al., 1995). In contrast to Fodor (1983), the unrestricted interactionist position argues for the unrestricted interaction of informational sources that derive from distinct cognitive systems such as language and vision. This view has received limited but strong support by experimental findings from the investigation of auditory language comprehension in visual scenes. Tanenhaus et al. (1995) examined whether information types issuing from distinct cognitive subsystems such as language and vision interact rapidly on-line. They argued that if non-linguistic information (e.g., scene context) influences core linguistic processes such as the incremental structuring of an utterance, then on-line sentence comprehension is clearly a highly interactive process. Such finding would support the claim that the visual perceptual and comprehension systems are not informationally encapsulated in the Fodorian sense.

Tanenhaus et al. (1995) revealed that core comprehension processes such as the structuring of an utterance can rapidly be influenced by a visual referential context (e.g., Tanenhaus et al., 1995; Spivey, Tanenhaus, Eberhard, & Sedivy, 2002). In particular, they investigated sentence comprehension in visual contexts that contained one versus two objects (e.g., one apple, two apples). Tanenhaus et al. (1995) found effects of a visual referential context on the on-line resolution of local structural ambiguity. 'Structural ambiguity' refers to situations where a sentence can be structured in two distinct ways. The resolution of structural ambiguity is a much-investigated process in psycholinguistic research on on-line sentence comprehension.

One reason for the interest in structural ambiguity resolution is that it can reveal insights about how the processing system structures a sentence. An example for a structural ambiguity is the prepositional phrase attachment ambiguity from Tanenhaus et al. (1995):

(1.1) Put the apple on the towel in the box.

In instructions such as *Put the apple on the towel in the box*, the phrase *on the towel* can be temporarily analysed as modifier and interpreted as the location of the *apple* (identifying which apple) or it can be attached to the verb phrase and interpreted as a destination (where to put the apple). Tanenhaus et al. (1995) found that the initially adopted structure and interpretation in sentences such as 1.1 is dependent upon the visual context. In a scene containing one apple, the phrase *on the towel* was preferentially interpreted as destination. In a scene with two apples people preferentially analysed the phrase *on the towel* as a modifier of the apple, interpreting it as a location. Such finding demonstrates the existence of a rapid interaction of the type of visual context and the way in which the sentence comprehension system chooses between two options in structuring an unfolding utterance. Importantly, the finding by Tanenhaus et al. (1995) has demonstrated that the informational integration between the language and vision systems is not informationally encapsulated in the Fodorian sense. Psycholinguistic theories that assume that core language comprehension processes such as the structuring of a sentence can be investigated in isolation from the immediate visual scene and from vision will not be able to straightforwardly account for such effects.

As a result of the above and other related findings, recent linguistic frameworks of the language system have begun to take into account the fact that scene information can influence core comprehension processes such as the structuring of an utterance. They explicitly embed the language system in relation to the other perceptual systems and allow for informational interaction between the ongoing processes of such diverse cognitive systems and between such systems and the immediate scene (e.g., Jackendoff, 1997, 2002). While procedural informationally encapsulated modularity is no longer assumed in this theory, representational modularity (i.e., that distinct information kinds such as syntax and semantics are represented by distinct mental representations) is still maintained. An example for the organization of cognitive systems in this type of framework is presented in Figure 1.2 (Jackendoff, 1997, p. 44). Figure 1.2 characterizes that representationally distinct linguistic structures (e.g., syntactic and conceptual structures) can interact with distinct spatial and visual representations. Spatial representations provide

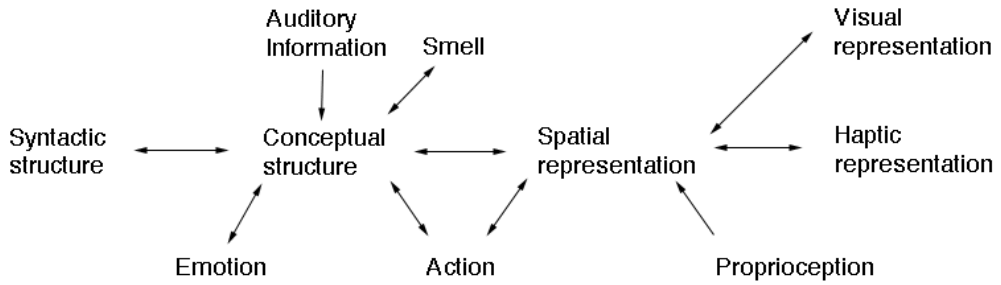


Figure 1.2: *Schematic sketch of a Jackendovian organization of cognitive systems (Jackendoff, 1997, p. 44)*

information about shape and location of objects and are the *'upper end' of the visual system* (Jackendoff, 2002, p. 346). The arrows represent interface modules that provide for the communication between distinct structures. Through communication via the interface modules, information from the immediate visual scene can influence language comprehension and the structuring of a sentence. Thus, such a framework should, in principle, be able to account for the rapid influence of visual scenes on the on-line structuring of an utterance that was observed by Tanenhaus et al. (1995).

Despite the convincing evidence Tanenhaus et al. (1995) provide for the influence of scene information on sentence comprehension, their findings are limited to scenes that contain objects. We do not know what kind of information in scenes can affect on-line language comprehension and how strong its influence is. We know, for instance, little about the influence of scenes that contain actions and events on on-line comprehension. These kinds of scenes provide propositional information as well as a referential context. As a result, they could be of even greater importance for core sentence comprehension processes such as the structuring of an utterance than scenes that contain only objects.

The primary goal of this thesis is to generalize and extend the existing insights on the influence of scene information during on-line sentence comprehension. To achieve this we explore the influence that depicted actions and events can have on the on-line structuring of an utterance.

1.2 SENTENCE COMPREHENSION: BEYOND INTERACTION

The insight *that* scene information can influence core comprehension processes such as the structuring of an utterance does not enable us to fully characterize how comprehension proceeds. It reveals little about precisely how perception of scene information is integrated with the unfolding utterance and with linguistic knowledge. In addition to investigating whether a type of information in scenes (depicted actions and events) influences incremental utterance comprehension, we thus set out to specify in more detail the *precise interplay* of the visual perceptual system with comprehension. In this respect we aim to go beyond frameworks that provide for the rapid interaction of visual perception and comprehension but fail to specify the nature of the interaction (e.g., Jackendoff, 2002). The Jackendovian framework is underspecified with respect to the precise nature and time-course of the interplay between the visual and language systems. Crucially, it makes no explicit predictions about the *mechanisms* that describe this interplay (or, as a matter of fact, comprehension in general). Mechanisms characterize the dynamic aspects of comprehension (see also Crocker, 1999):

> A description of mechanisms is, in effect, a claim about the events in which the architectural components participate and about the causal structure of these events.
>
> (Pickering, Clifton, & Crocker, 2000, p. 1)

A full description of the mechanisms of the interplay between the visual perceptual system and incremental language comprehension is beyond the scope of this thesis. Pickering, Clifton, and Crocker (2000) suggest, however, that we can approximate a description of the mechanisms by attempting to understand the processes that take place. We follow this proposal and investigate in more detail the processes that occur in integrating utterance, linguistic knowledge, and scene information.

Since a Jackendovian framework is underspecified with respect to how the integration of utterance, linguistic knowledge and scene information proceeds, we consider whether

unrestricted interactionist frameworks provide a better specification of on-line comprehension processes (e.g., McRae, Spivey-Knowlton, & Tanenhaus, 1998; Tanenhaus et al., 1995; Trueswell et al., 1993, 1994; Spivey-Knowlton, 1996). The only proposal among these accounts that is explicitly concerned with the interaction of visual and linguistic information is Tanenhaus et al. (1995). Their account details that the scene can influence processing. It does not describe more precisely the on-line interplay between utterance, linguistic knowledge, and scene information.

A more fully specified and implemented model within the interactionist framework is the competitive-integration model by Tanenhaus, Trueswell, and Hanna (2000) (see Spivey-Knowlton, 1994). We propose that such a model might, in principle, be extended to incorporate information that has been derived from a visual scene. We briefly describe the model here to discuss the kinds of specifications it provides. A sketch of this type of model is provided in Figure 1.3.

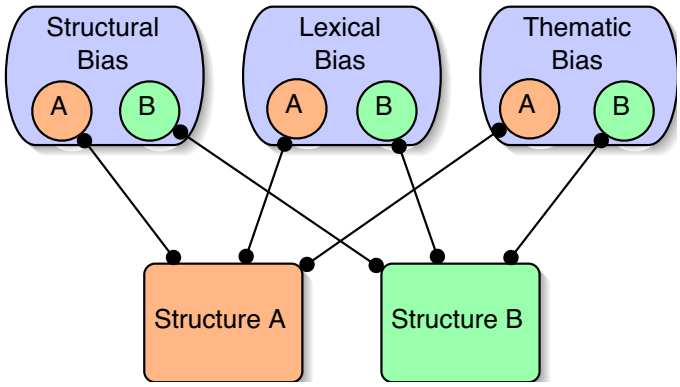


Figure 1.3: *Schematic sketch of a constraint-based model*

There are several key parts to the model: informational biases (e.g., structural, lexical, and thematic), their weights (represented by the lines), and sentence structures (e.g., Structure A and Structure B in Fig. 1.3). A 'bias' is an informational source in this model. Each bias is characterized by the extent to which is supports an alternative structure (A, B) of a structurally ambiguous sentence. Processing steps in the model detail how the biases are combined, a process which simulates structural ambiguity resolution. In the first processing step the activations for the two nodes (A, B) of each bias are normalized. In a second step the activation of the two structures (A, B) is determined by integrating

their respective corresponding bias nodes using a weighted sum. A third step ensures that structure nodes can send feedback to the biases depending on how strongly a bias has activated that structure. This type of model thus provides a specification of how diverse informational biases combine. This is a substantial achievement and should not be underestimated.

We think, however, that while the model details how *informational biases* combine, it is not a model of full *sentence comprehension*. Information processing may be determined through probabilistic algorithms alone. The recovery of a complete interpretation for a sentence, however, requires more than that. Specifically, we take the view that it requires a description of how *mental representations* of diverse kinds of information are derived, and how they combine, an aspect that is entirely missing from the model by Tanenhaus et al. (2000). Algorithms such as the ones used in the competitive-integration model, for instance, do not determine how grammatical function and thematic role are assigned to a currently processed phrase. If we want to model not just how to decide between two alternative structural analyses, but if we aim at actively building a structure, then we need a theory that in some form or other permits us to specify conceptual structures and mechanisms for combining mental representations.

We argue that unrestricted accounts have shown how language comprehension does *not proceed* in falsifying proposals such as the modularity hypothesis. They are, however, less informative about how comprehension *does proceed*. The secondary goal of this thesis is to go beyond such unrestricted interaction accounts. Specifically, we propose that exploring the precise nature of the interplay between the visual perceptual system and comprehension could lead to a more fully specified account (and ultimately theory) of on-line sentence comprehension in visual scenes. In pursuing an account that goes beyond unrestricted interaction we therefore take as our starting point the following observation:

> We agree that constraint-based models need to be more explicit about the nature of the constraints and how they combine [...].
>
> (Tanenhaus et al., 2000, p. 94)

1.3 THE INTERPLAY BETWEEN LANGUAGE AND VISION

To investigate more fully how available informational sources such as language and scene information combine, we now consider in more detail the nature of the interplay between

the visual perceptual system and on-line comprehension. One way of specifying what we term the 'nature' of the interplay between visual perception and comprehension is by detailing how mental representations combine. This might, for instance, involve specifying their origin and importance, and furthermore the dynamic mechanisms that govern how they combine. In a setting where scene and utterance are available, crucially, sentence comprehension can be informed by mental representations deriving from both of these informational sources. Further, in deriving and combining mental representations of utterance and scene, an important aspect is whether the mechanisms of the systems that process utterance and scene are coordinated or not.

Linguistic and visual processing might, for instance, be *asynchronous*. 'Asynchronous' in communication research means that two communication devices do not require a common timing reference in order to communicate (Webster, 1994). Alternatively, the interaction of distinct cognitive processes might be of a more coordinated nature. Linguistic and visual processing could, for instance, be tightly *synchronized* (see Zelinsky & Murphy, 2000). 'Synchronized' in communication research refers to a communication technique between two systems that requires a common time signal in order to coordinate informational exchange between the systems. To explore these two possibilities (asynchronous and synchronous/coordinate), we consider findings from studies by Zelinsky and Murphy (2000) and Bergen, Narayan, and Feldman (2003). While none of these studies directly investigate sentence comprehension in visual scenes, their findings are nonetheless relevant for the question of how tightly coordinated the language system and the visual perception system are during linguistic processing.

Zelinsky and Murphy (2000) investigated the relationship between the number of syllables in the name of an object and visual inspection of that object in two tasks (short-term recognition and visual search). They examined whether the inspection duration of a face depended upon the number of syllables (one or three) of a face's name. They suggested that finding evidence for such a dependency would support the claim that linguistic and visual processing are dependent upon one another. In the short-term recognition task, people were presented with a display depicting four faces. They studied these for subsequent recognition. They were then presented with a face and had to decide whether or not the face had been part of the previous set of faces. In the visual search task, the order of presentation was reversed. Participants first saw a target face, and then had to decide whether or not it was present in a set of four faces. Zelinsky and Murphy (2000) found that people inspected faces with three-syllable names significantly longer than faces

with a one-syllable name. This effect was only observed in the short-term recognition task that encouraged linguistic processing, but not in the visual search task. The finding supports the claim that for *linguistic processing*, inspection of an object was shorter when processing the object's name (one-syllable name) required less time. When encoding an object's name took longer (three-syllable names), visual processing of the object continued until encoding of its name was achieved. The authors conclude that linguistic and visual processing are highly synchronized, and that *the faster of these two processes waits for the slower process to be completed* (Zelinsky & Murphy, 2000, p. 130).

Further support for the claim that processes of the language system can be tightly coordinated with visual perceptual processing comes from research on the embodiment of language (e.g., Bergen et al., 2003; Buccino et al., 2001; Feldman & Narayanan, 2003; Martin, Haxby, Lalonde, Wiggs, & Ungerleider, 1995; Pulvermüller, Härle, & Hummel, 2001). An emerging body of research shows that brain areas that are distinct from areas which have been identified as responsible for language comprehension are active during the comprehension of words (e.g., Bergen et al., 2003; Buccino et al., 2001; Martin et al., 1995; Pulvermüller et al., 2001), or sentences (Stowe, Paans, Wijers, & Zwarts, 2004). Regions in the human pre-motor cortex, for instance, are involved in the observation and execution of actions performed by the hand or leg. It has been suggested that this *mirror system*, which is involved in performing and recognizing an action, is also involved in understanding a verb describing an action (e.g., Bergen et al., 2003, p. 139).

Bergen et al. (2003) carried out an experiment where people had to decide upon whether a depicted action that they had just seen matched a subsequently presented written verb. Their hypothesis was that processing of action verbs relies on the same motor circuitry as recognizing an action. If this were the case, interference should be caused when a depicted action (e.g., scratch) was followed by a verb that did not match the action (e.g., *hold*), but that required the same effector (e.g., a hand is an effector for a scratching and a holding action). This condition was compared to two other conditions. In one, the depicted action was followed by a matching verb (e.g., *scratch*); in the other, it was followed by a non-matching verb (e.g., *stumble*) that required a different effector than the previously presented action. They found that when the action and the verb shared the same effector (a scratching-action and *hold*) participants took longer to reject a no-match than when depicted action and verb did not share the same effector (a scratching-action and *stumble*). This finding supports the claim that effector-specific neural circuitry of the human mirror system is activated during the understanding of action verbs.

The conclusion that we can draw from these and other findings on the interplay between the language and vision systems is that perceptual processing of the immediate visual scene could be an influential factor for on-line comprehension processes. Findings such as those by Bergen et al. (2003) provide strong support for the claim that comprehension of language about actions is closely coordinated with perception of the immediate scene that depicts the actions. Furthermore, their findings draw attention to the possibility that cognitive processes other than linguistic ones are involved in language processing, and might, in principle, contribute important information to language comprehension.

Based on these findings and those of Tanenhaus et al. (1995) we argue that the interplay between the comprehension and visual perceptual system is tightly coordinated, and that visual perception may play an important part in language comprehension. Experiments on auditory comprehension in visual scenes have demonstrated that upon hearing a word, people inspect the object that the word identifies (e.g., Dahan, Magnuson, & Tanenhaus, 2001; Tanenhaus et al., 1995; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1996). It has further been shown that the type of referential context can influence the structuring of an utterance (Tanenhaus et al., 1995). However, little attention has so far been given to the possibility that utterance and scene might mutually constrain each other's influence. Their interplay, we suggest, is a coordinated interaction: Comprehension of words in an utterance can direct attention in a visual scene and thus be the causal factor for when which part of a scene becomes available for comprehension. Importantly, such language-mediated perception of the immediate visual scene might in turn be an important guiding factor for interpretation of the utterance. We propose that this interplay constrains on-line language comprehension and has presumably shaped the architecture of the language system and comprehension mechanisms (see section 6.2.3).

1.4 A RESEARCH PROGRAM

Ultimately, a theory of on-line sentence comprehension will have to account for comprehension both when the immediate scene is relevant for comprehension and when it is irrelevant. We have argued that existing psycholinguistic theories of on-line sentence comprehension, while accounting for the effects of linguistic knowledge, are not able to account for the effects of scene information on comprehension. This is due to the fact that they neither *explicitly* relate the comprehension system to distinct cognitive systems such as the visual perception system nor explicitly and formally situate the process of on-line language com-

prehension within the immediate visual scene (e.g., Jackendoff, 2002). Since so far only *frameworks* (rather than fully specified processing theories) have situated comprehension in the immediate scene, a further issue is that theoretical psycholinguistic research has not yet specified the precise processes and mechanisms that govern the interplay between scene processing and comprehension processes.

We suggest that the lack of attention that theoretical research has paid to the influence of scene information is also reflected in experimental psycholinguistic research. Indeed, psycholinguists in the field of on-line sentence comprehension have only recently recognized the importance of investigating processing at the language-vision interface (e.g., Altmann, 2004; Henderson & Ferreira, 2004). Most of the ongoing psycholinguistic studies that explicitly investigate spoken sentence comprehension during the inspection of visual scenes have paid little attention to the influence of visual scene information on the structuring of an utterance. Previous research has rather emphasized how interpretation of the linguistic input affects incremental or even anticipatory determination of reference to entities in the world during on-line comprehension (e.g., Altmann & Kamide, 1999; Chambers, Tanenhaus, Filip, & Carlson, 2002; Hanna, Tanenhaus, & Trueswell, 2003; Kaiser & Trueswell, in press; Kamide, Altmann, & Haywood, 2003; Kamide, Scheepers, & Altmann, 2003; Runner, Sussman, & Tanenhaus, 2003; Sedivy, Tanenhaus, Chambers, & Carlson, 1999; Sussman & Sedivy, 2003).

Such an enterprise is interesting, and has revealed a great deal about how utterance and knowledge of language influence language comprehension, i.e. attention in the scene *passively* reflects unfolding sentence interpretation. We argue, however, that in addition research on auditory comprehension in visual scenes must pay attention to the way in which a perceived scene actively influences core comprehension processes. Specifically, we know little about which types of information in visual scenes may influence core comprehension processes and to what extent. Further, we propose attention should be paid to the nature (i.e., coordinated or not) of the interplay between the visual perceptual system and on-line sentence comprehension. Among the questions we address are:

(a) Which types of information in scenes can influence on-line sentence comprehension processes such as structural disambiguation and incremental thematic role assignment? What is the time-course of their influence?

(b) Does the utterance in turn constrain the rapid on-line influence of visual scenes on such comprehension processes?

(c) During comprehension, what is the relative importance of scene information in comparison to stored linguistic and world knowledge?

Based upon an investigation of these questions, the present thesis aims to extend existing theories/models of language comprehension (e.g., MacDonald et al., 1994; McRae et al., 1998; Tanenhaus et al., 2000) that seek to account for the interaction of diverse types of linguistic knowledge on sentence comprehension. We aim to develop an account of on-line sentence comprehension that can serve as a basis for accommodating the effects of scene information on comprehension in addition to accommodating the influence of linguistic knowledge on comprehension processes.

1.5 THESIS OVERVIEW

In the following chapter, we summarize relevant insights from psycholinguistic research on comprehension during reading, and during auditory sentence comprehension in visual scenes. The review will discuss in detail the importance of stored linguistic/world knowledge and of visual scenes in current psycholinguistic research on on-line sentence comprehension with respect to comprehension processes such as the structuring of an utterance and thematic role assignment. Based on this review, we motivate the research issues that we investigated.

Chapters 3 to 5 presents findings from five eye-tracking experiments. All of the experiments investigated auditory sentence comprehension in visual scenes. The first four of these experiments explore how depicted action scenes, which have not yet been directly investigated in on-line sentence comprehension, affect comprehension processes. The fifth experiment examined how diverse informational sources such as scene information and stored linguistic/world knowledge combine during on-line comprehension.

In chapter 3, we present findings that indicate great importance of immediately depicted events in the structuring and thematic role assignment of initially ambiguous German verb-second main clauses. In chapter 4, we report experiments that generalize this finding to two further sentence constructions (German verb-final and English reduced relative clause sentences), and another language (English). We further explore the extent to which the influence of depicted events on comprehension is modulated by the word order of the unfolding utterance. Chapter 5 presents the fifth experiment. This study explicitly contrasts two types of information (stored thematic role knowledge and immediately

depicted events) in on-line sentence comprehension. Findings revealed a preference for immediately depicted events in the on-line assignment of thematic roles over stored thematic role knowledge. In chapter 6, we summarize the main findings, relate them to prior research, and integrate them within an account of sentence comprehension that describes how the interplay between utterance comprehension and attention in the scene constrains comprehension. A conclusion in chapter 7 discusses the contribution of the findings from the present thesis to psycholinguistic research on on-line sentence comprehension. Chapter 8 provides a German summary of the present thesis.

Chapter 2

Incremental Thematic Role Assignment

The discussion in chapter 1 has drawn attention to the fact that under the influence of Fodorian philosophy, researchers have paid little attention to the potential influence of visual perceptual processing on incremental comprehension processes such as the structuring of an utterance. While many studies have investigated language comprehension in situations where a visual scene was irrelevant (e.g., reading of isolated sentences), only few have examined comprehension of language that is related to the immediate scene. As a result, current psycholinguistic theories of sentence comprehension have almost exclusively been developed on the basis of findings from comprehension of language that is not related to the immediate visual scene. We outlined that Fodorian modularity has since been largely falsified by research that has provided evidence for the influence of a visual referential context on the incremental structuring of an utterance (Tanenhaus et al., 1995). Findings from studies by Tanenhaus and colleagues have revealed that the syntactic processes of the language system are not informationally encapsulated in the relevant Fodorian sense. Rather, on-line language comprehension processes can interact with ongoing visual perceptual processes. Since the precise nature of the interaction between language comprehension and visual perception is so far unclear we proposed to further specify the nature of this interaction with the goal of providing an account of comprehension in situations where the scene is relevant for comprehension. We view such an account as complementary and adding to a description of the mechanisms of comprehension when the scene is irrelevant for comprehension.

As a first step in this direction, consider an example experiment on on-line sentence comprehension in a visual environment that contains entities (objects, animals, and people). Participants are presented with a visual scene. They inspect the scene entities for a short time and can acquire information about them. After this preview time (often around 1000 ms), they listen to a sentence that is related to the scene. The sentence can be an instruction which participants have to carry out. Alternatively, it describes (part of) the scene from a third-person perspective while participants are either engaged in a question-answering or a passive listening task. All of these tasks have in common that they encourage participants to understand the utterance. During scene inspection and utterance comprehension, a camera monitors people's eye-movements to objects in the immediate scene. Psycholinguists are particularly interested in the record of eye-movements to scene objects during utterance comprehension. This is because attention to scene objects during auditory comprehension has been identified as an indicator of on-line language comprehension processes such as establishing reference (e.g., Cooper, 1974; Tanenhaus et al., 1995).

From the above example it becomes clear that there is a range of informational sources which might inform language comprehension in visual scenes. Among them are:

(2.1) The linguistic context provided by the utterance (and also the larger discourse/dialogue context whose importance we acknowledge but do not deal with here)

(2.2) The listener's stored linguistic and world knowledge

(2.3) Information from the immediate visual scene

The comprehension system accesses information from utterance and scene via the perceptual systems (e.g., Tanenhaus et al., 1995). Language comprehension is hence not the only process that takes place in an experimental setting such as the one described above. Rather, at least two processes occur:

(2.4) Language comprehension

(2.5) Scene perception

We know little about the interplay between available informational sources (e.g., 2.1 to 2.3) and distinct cognitive processes (e.g., 2.4 to 2.5) in such a setting since only few studies

have directly investigated the role of perceptual processing of scene information during on-line utterance comprehension (e.g., Tanenhaus et al., 1995; Sedivy et al., 1999; Spivey et al., 2002). Similarly, relatively few studies, have investigated comprehension during reading when scene information was available (e.g., Carroll, Young, & Guertin, 1992; Rayner, Rotello, Stewart, Keir, & Duffy, 2001; Underwood, Jebbett, & Roberts, 2004). In contrast, more is known about the interplay between available informational sources (various types of stored linguistic/world knowledge) and comprehension mechanisms when we read a sentence in the absence of a relevant scene.

Stored linguistic and world knowledge has been identified as an important informational source for comprehension during reading in a number of studies. Prior psycholinguistic research has, for instance, shown that stored linguistic and world knowledge rapidly influence the structuring and incremental interpretation of a sentence during reading (e.g., Hyönä & Hujanen, 1997; Traxler & Pickering, 1996; Trueswell et al., 1993). In particular, stored verb-based knowledge of who-does-what-to-whom has been found to have a strong and rapid influence on structural disambiguation and incremental thematic role assignment during reading (e.g., McRae, Ferretti, & Amyote, 1997; McRae et al., 1998; Taraban & McClelland, 1988; Trueswell et al., 1994). In addition, research on auditory comprehension has demonstrated its rapid influence on the incremental interpretation of an utterance that described a scene (e.g., Altmann & Kamide, 1999; Kamide, Scheepers, & Altmann, 2003). An example for this type of knowledge is, for instance, that a noun phrase such as *the cop* is a typical thematic agent and a noun phrase such as *the crook* a typical thematic patient role filler of the verb *arrest* (see McRae et al., 1997, p. 153). We henceforth refer to this type of knowledge as 'stored thematic role knowledge'.

In reviewing relevant literature we will outline that prior research has investigated the effects of such stored thematic role knowledge as an informational source that permits determining who-does-what-to-whom in structuring and interpreting a sentence (section 2.1). We highlight relevant insights on the influence of stored thematic role knowledge during on-line language comprehension. The influence of *scene information* on the structuring and thematic interpretation of a sentence has, in contrast, so far received much less attention. Direct investigation of the influence of scenes has furthermore been limited to a *referential* visual context in psycholinguistic research on on-line sentence comprehension (see section 2.2). The influence of scenes that establish *who-does-what-to-whom* in on-line utterance comprehension has not been directly investigated in psycholinguistic studies on on-line sentence comprehension. While existing findings allow us to characterize the inter-

play between processes that are involved in language comprehension when language is not related to an immediate scene, the interplay between scene perception and comprehension is relatively unexplored. To investigate it further, we identify factors that might be responsible for the influence of scene information on comprehension in existing studies (see section 2.3). Based on this analysis and on other relevant experimental findings discussed in section 1.3, we then motivate the experiments that will be presented in chapters 3 to 5 (see sections 2.4 and 2.5). Among other questions we ask which type of information in scenes can rapidly influence incremental structural disambiguation. Further to examining the influence of one informational source on on-line comprehension, we ask precisely how (i.e., coordinated or not) the visual perception systems interacts with on-line utterance comprehension.

2.1 STORED THEMATIC ROLE KNOWLEDGE

The present section discusses the influence of stored thematic role knowledge on structural disambiguation and incremental thematic role assignment. On-line language comprehension is a rapid process where the structure and interpretation of a sentence must be computed in real-time. One decisive factor for the incremental influence of stored thematic role knowledge in on-line comprehension is whether it can be retrieved rapidly on-line. Ferretti, McRae, and Hatherell (2001) found that verbs immediately activated stereotypical knowledge of agents (*arresting-cop*) or patients (*arresting-criminal*). They concluded that this type of world knowledge is part of thematic-role knowledge, and immediately activated during lexical access upon encountering the verb (see also McRae et al., 1997). Such knowledge is thus available to inform the assignment of thematic roles to sentence constituents on-line. In the following section, we discuss findings on the influence of stored thematic role knowledge on incremental thematic role assignment and structural disambiguation and outline how they have influenced theory development. Particular emphasis will be given to the time-course with which stored linguistic and thematic role knowledge can affect on-line sentence comprehension.

2.1.1 EXPERIMENTAL EVIDENCE

In an eye-tracking study on the reading comprehension of initially structurally ambiguous reduced relative clause sentences, Trueswell et al. (1994) investigated the influence of animacy on ambiguity resolution. In the main clause (MC)/reduced relative clause (RR)

ambiguity, the initial noun phrase is role ambiguous. It can either be the agent of a simple main clause, receiving only one thematic role. Alternatively, it can be the patient of a reduced relative clause, and the agent of the main clause that contains the reduced relative clause, thus receiving two thematic roles (e.g., *The evidence examined by the lawyer turned out to be unreliable*) (see chapter 4 for a more detailed description of this ambiguity). The structural ambiguity hinges upon an ambiguity of verb form. If the post-nominal verb is ambiguous between a simple past and a past participle form of the verb, then the initial noun phrase can temporarily be either assigned an agent (main clause) or a patient (reduced relative clause) role. There is typically a strong bias to initially assume a main clause structure in isolated sentences (e.g., Bever, 1970). As a consequence, when readers encounter the post-verbal *by*-phrase that disambiguates towards a reduced relative clause structure, they experience processing difficulty. This usually manifests itself in increased reading times at the disambiguating region. An example item from Trueswell et al. (1994) is provided below. For both sentences 2.6 and 2.7, the initial noun phrase *the witness* and *the evidence* is the patient of the reduced relative clause *examined by the lawyer*. At the same time, the first noun phrase is the agent of the main clause *turned out to be unreliable*.

Trueswell et al. (1994) investigated whether verb-based semantic knowledge (e.g., *examine* requires an animate noun as its agent) rapidly influences the assignment of thematic role and alleviates the difficulty that is typically associated with the reduced relative clause structure. Sentences 2.6 and 2.7 provide two example sentences used in their study. In sentence 2.6, the first noun phrase is animate, and thus a correct thematic agent for an examining-action. In sentence 2.7, in contrast, the initial noun is inanimate, and thus can only be a thematic patient, as opposed to agent of the verb *examined*. If the comprehension system can rapidly establish the semantic fit between the animacy of a role filler and verb-related semantic knowledge, then the first noun phrase in 2.7 should receive a patient role once the verb has been encountered. This should lead to less processing difficulty when the *by*-phrase disambiguates towards a reduced relative clause in sentence 2.7 as opposed to 2.6.

(2.6) The witness examined by the lawyer turned out to be unreliable.

(2.7) The evidence examined by the lawyer turned out to be unreliable.

Trueswell et al. (1994) found that the semantic fit between the animacy of the first noun phrase and the following verb rapidly influenced people's decisions about which

thematic role to assign to the sentence-initial noun phrase. In studies investigating the comprehension of sentences such as 2.7, the first noun phrase was rapidly assigned the thematic role of a patient despite a strong main clause preference.

Evidence for this was obtained from first and second pass reading times on the region following the verb. First pass reading times were obtained by summing all the left-to-right fixations in a region including regressions within that region. Second pass times reflect any re-reading of a region. Both first and second pass reading times for the *by*-phrase that disambiguated the local ambiguity towards a reduced relative clause were longer for sentence 2.6 in comparison with 2.7 (Trueswell et al., 1994). When the semantic fit between the inanimate first noun phrase (*the evidence*) and the ambiguous verb (*examined*) indicated that the first noun phrase could only be the patient of the verb in 2.7, the difficulty associated with the reduced relative clause was diminished. The fact that this facilitation effect was observed in first pass reading times provides support for the claim that comprehension processes made rapid use of stored knowledge about the appropriateness of a noun phrase as the thematic agent/patient of a verb, once such knowledge had been made available by the words in the sentence.

McRae et al. (1997) built on and extended this line of research. They investigated the influence of stored thematic role knowledge on comprehension. In a self-paced reading study, McRae et al. (1997) tested whether such experience-based knowledge of who-does-what-to-whom could be applied rapidly to inform structural disambiguation and incremental thematic role assignment. They used adjectival features to bias the concept of an initial noun (*gambler*) towards being a good agent (*shrewd heartless*) or a good patient (*young naive*) of the past participle (*manipulated*) of a reduced relative clause (*The shrewd heartless/young naive gambler manipulated by the dealer had bid more than he could afford to lose*). As mentioned above, people typically favor a main clause analysis of the initially ambiguous reduced relative clause. Accordingly, they should experience processing difficulty when the *by-phrase* disambiguates towards a reduced relative clause. McRae et al. (1997) demonstrated that a thematic-fit manipulation modified this comprehension difficulty. Reading times on the *the dealer*-region were increased for agent-biased (*shrewd heartless*) over patient-biased (*young naive*) sentences, indicating greater difficulty with reduced relative clauses for agent-biased sentences. This finding shows that in patient-biased sentences, the patient-bias was rapidly used to interpret the first noun phrase as the patient of the clause. This then led to decreased reading times on the disambiguating second noun phrase for this condition (*young naive*) over the agent-biased condition.

Even stronger evidence for the use of such verb-specific thematic role knowledge is provided by McRae et al. (1998). They carried out a self-paced reading experiment that investigated the influence of the thematic fit between a verb (*arrest*) and a pre-verbal noun phrase (either *cop* or *crook*). They compared reading times for sentences such as *The cop arrested by the detective was guilty of taking bribes* with reading times for sentences such as *The crook arrested by the detective was guilty of taking bribes*. Unambiguous control sentences were constructed by inserting *who was* before the *ed*-verb.

The decisive manipulation was that for the structurally ambiguous sentences, the fit between *cop* and *arrested* should create an agent bias. This should increase reading difficulty upon disambiguation of the clause as a reduced relative clause instead of a main clause (*arrested by*). For *crook*, a good patient of the verb *arrested*, the thematic fit between verb and noun should in contrast reduce the difficulty that is typically associated with the disambiguating *by*-phrase. At the verb-region (*arrested by*), McRae et al. (1998) found the expected interaction between thematic fit and reduction of comprehension difficulty. The interaction at the verb-region was brought about by a more substantial reduction effect for the sentences with good patients of *arrested by* (see Fig. 2.1).

The above findings of the rapid interaction of verb-based thematic role knowledge with other available informational sources (e.g., structural biases) during reading provide evidence for the rapid influence of such knowledge on incremental structural disambiguation. In all of these studies, this rapid influence occurred rapidly after the verb had made relevant stored thematic role knowledge available.

The finding that such thematic role knowledge is important for incremental thematic role assignment generalizes to spoken utterance comprehension in visual scenes. Indeed, studies examining the use of verb-related knowledge on incremental thematic interpretation in visual scenes provide further strong support for the rapidity with which such knowledge can be used. In two experiments Altmann and Kamide (1999) investigated the influence of verb selectional restrictions on incremental interpretation of structurally unambiguous sentences. They make clear that they consider 'verb selectional restrictions' as little different from the verb-based thematic role knowledge investigated by McRae et al. (1997) (Altmann & Kamide, 1999, p. 260). Altmann and Kamide (1999) found strong evidence for the highly rapid use of such verb-based knowledge during on-line comprehension in visual scenes. In their experiments, people listened to spoken utterances, while inspecting a scene. An example image contained a boy, a cake, a toy train, a car, and a ball. Altmann and Kamide (1999) contrasted sentences where the verb indicated only one

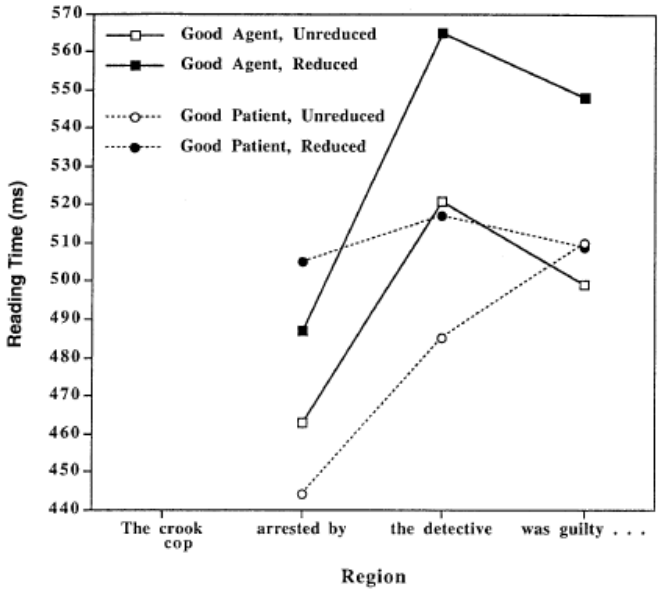


Figure 2.1: *Self-paced reading times from the experiment by McRae et al. (1998)*

object as an appropriate goal with sentences where verb selectional restrictions allowed all available objects as targets. For sentences where only one object (a cake) was compatible with the verb's selectional restrictions such as *The boy will eat ...*, Altmann and Kamide (1999) observed anticipatory eye-movements to the one appropriate edible object (a cake) (i.e. 'anticipatory' means looks to an object that occur before a word directly refers to the object). No such anticipatory eye-movements to the cake were found when the verb's selectional restrictions applied to several objects (a cake, a ball, a toy train, and a car) as was the case for sentences like *The boy will move ...* . Their finding reveals that stored verb-based knowledge can be used rapidly to restrict the post-verbal domain of reference. While findings by Altmann and Kamide (1999) still leave room for an interpretation in terms of a strong association between the verb (*eat*) and available scene objects (a cake), subsequent studies by Kamide, Scheepers, and Altmann (2003) exclude this possibility.

Kamide, Scheepers, and Altmann (2003) showed that unambiguous case-marking and verb-related knowledge of plausible thematic role fillers influence expectations of post-verbal arguments and available referents. In German, a case-marked article can determine the grammatical function and thematic role of the noun phrase it modifies. Both SVO (subject-verb-object) and OVS (object-verb-subject) orders are grammatical. Participants inspected images showing a hare, a cabbage, a fox and a distractor object while hearing sentences such as *Der Hase frisst gleich den Kohl* ('The hare (subj) eats soon the cabbage (obj)') and *Den Hasen frisst gleich der Fuchs* ('The hare (obj) eats soon the fox (subj)'). Subject and object case-marking on the first noun phrase together with verb-based thematic role knowledge allowed determination of the correct post-verbal referent. This was evidenced by anticipatory eye-movements to the cabbage after participants had heard 'The hare (subj) eats ...' and to the fox after having encountered 'The hare (obj) eats ...'. Anticipation of the appropriate patient role filler after 'The hare (subj) eats ...' and of the correct agent role filler after 'The hare (obj) eats ...' required compositional integration of morphosyntactic case-marking on the first noun with verb plausibility. A mere association between the first noun phrase ('the hare') and the verb ('eat') would have resulted in a main effect of inspections to the cabbage in both SVO and OVS conditions. These studies on auditory comprehension in visual scenes confirm that there is indeed a close time-lock between when the utterance makes relevant thematic role knowledge available, even when that knowledge is modulated by morpho-syntactic cues (case-marking), and when such knowledge can inform interpretation of the utterance. A separate issue with studies by Altmann and Kamide (1999) and Kamide, Scheepers, and Altmann (2003) is whether stored thematic role knowledge was provided by the utterance alone, or by both the utterance and the referential context. We discuss this issue in section 2.3.

2.1.2 THEORIES OF THEMATIC ROLES

The above discussion of experimental findings has drawn attention to the importance of stored thematic role knowledge for structural disambiguation and incremental thematic role assignment. Since this type of thematic role knowledge is influential in comprehension, we suggest that theories of thematic *role assignment* should, in principle, be able to account for the effects of this type of knowledge. This means that they must consider the above findings in deriving their inventory of thematic roles. Specifically, theories of thematic role assignment require an inventory of lexically specified thematic roles to be able to account

for the effects of such knowledge. The question of what the inventory of thematic roles is, and how they are best characterized has been the concern of much research in theoretical linguistics (e.g., Dowty, 1991; Fillmore, 1968, 1985; Foley & Van Valin, 1984; Jackendoff, 2002) and (to a lesser extent) in psycholinguistic research on language comprehension (e.g., McRae et al., 1997; Tanenhaus, Carlson, & Trueswell, 1989).

In what follows we provide a brief survey of three types of thematic role theories and the kind of inventory of *thematic roles* they provide (also *theta-role* in Government-Binding (GB), or *thematic relation* in Jackendoff (1990)). In particular, we briefly sketch the approaches by Dowty (1991), Fillmore (1985) and McRae et al. (1997), as well as Davidson (1967) and Davidson (1986b) in as they are of importance within the context of this thesis, while pointing out relevant differences between the approaches. One important aspect for this thesis is the purpose with which a theory of thematic roles has been developed. We suggest here that it is important for theories of thematic role assignment in comprehension to be developed in close relation to experimental evidence from comprehension. Many theories of thematic roles (e.g., Davidson, 1967; Dowty, 1991) have, however, been developed for different purposes (e.g., linguistic abstraction or formalized description of thematic roles) and cannot be straightforwardly used as theories of thematic role assignment during comprehension. We further propose that the purpose with which theories of thematic roles have been developed influences the kind of inventory of thematic roles that they provide. This is an important issue since (from a language comprehension perspective) the inventory of thematic roles determines whether a theory of thematic roles can account for the effects of a particular type of information (e.g., stored thematic role knowledge) on thematic role assignment.

Dowty (1991, pp. 571ff.) classifies thematic relations at a relatively coarse-grained level, collapsing across finer-grained thematic-role distinctions in the search of a higher-level generalization of the notion of 'thematic role'. He distinguishes between two prototypical role clusters, *Proto-Agent* and *Proto-Patient*. These two clusters each comprise finer-grained thematic role concepts that share certain characteristics. The Proto-Agent cluster comprises thematic roles with a more agentive character (e.g., properties such as volitional involvement in event or state, causing an event, movement). Proto patient, in contrast, contains thematic roles that have higher number of patient-features (e.g., undergoes change of state, causally affected by another participant, stationary).

The theory by Dowty (1991) has thus been developed with a specific goal in mind (i.e., to provide an abstract linguistic description of the inventory of thematic roles).

This is an entirely valid enterprise that has certainly revealed important insight about how to best capture a large subset of thematic role concepts in a set of two general concepts such as Proto-Agent and Proto-Patient. However, we propose that a theory whose aim is to provide an *abstract* characterization of the inventory of thematic roles is unlikely to capture the often fine-grained semantic differentiations that matter in on-line language comprehension (e.g., the kinds of thematic role knowledge activated when we hear a sentence fragment such as *The cop arrested ...*) (e.g., McRae et al., 1997). We suggest that theories of the type Dowty (1991) proposes are thus little suited to accommodating findings from language comprehension (e.g., Altmann & Kamide, 1999; Kamide, Scheepers, & Altmann, 2003; McRae et al., 1997, 1998; Trueswell et al., 1994).

A further type of thematic role theory is represented by truth-conditional semantic theories (e.g., Davidson, 1986a, 1986b; Parsons, 1990). Foundational contributions in developing this type of theory were made by Quine (1953) who introduced the use of the Tarskian theory of truth in semantic research (see Katz, 1996). According to Katz (1996) the motivation for developing this type of theory was to provide a formal account of semantics in terms of the conditions under which an utterance is true. 'Truth' in this type of theory is relative to a model, situation, or world. It characterizes a relation between a sentence and the world. Davidson (1967) and elsewhere developed such a referential semantic theory using the formal tools provided by the Tarskian theory of truth. In his theory, semantic interpretation of sentences are referential relations from the world to the semantic values that expressions in the world have (e.g., Jackendoff, 1990; Katz, 1996). The inventory of thematic roles in this theory is thus influenced by the range of objects in the world that are referred to by referential expressions (i.e., things, although this type of theory also encompasses mental representations for events, see Davidson (1986b)). This formal type of semantic theory fails, however, just as the theory by Dowty to consider many aspects of language *comprehension* such as fine-grained semantic differences (e.g., conceptual meaning and plausibility) in the use and understanding of individual lexical items. Specifically, Fillmore (1985) argued that a set of truth conditions may not always be sufficient to account for the meaning of a word or sentence if we consider meaning from the perspective of language understanding (for further details see, for instance, the examples provided by Fillmore (1985) on the use of prepositions such as *on* and *in*).

We propose that in contrast to the Dowtyan and Davidsonian approaches, the theory by Fillmore (frame semantics) is informed to a greater extent by criteria of language understanding (Fillmore, 1985). Fillmore refers to the type of theory he proposes as

'U-semantic theory' where the 'U' stands for 'understanding' and is contrasted with 'T-semantic theory' (truth-conditional semantics). In contrast to the above-sketched truth-conditional semantic theory that emphasizes the notion of referential 'truth' of a sentence with respect to a world for semantic analysis, the theory by Fillmore (1985) highlights the importance of language understanding in developing a semantic theory:

> The phenomenologically primary data for language theory are taken to be the data of 'understanding' rather than such theory-defined derivative data limited to conditions under which *sentences* can be described as 'true'.
>
> (Fillmore, 1985, p. 235)

As a result of its different objectives, the frame semantics approach pursued by Fillmore (1985) provides an entirely different inventory of thematic roles in comparison with Dowty's approach, and focuses on other aspects of semantics and thematic roles than a Davidsonian account. Specifically, Fillmore (1985) takes into account more fine-grained lexical semantic differentiations of meaning. He introduces the notion of *frame* as a cognitive organizational device for knowledge and experience to which individual words are linked. A well-known example is the Commercial Transaction Frame that includes, for example, a buyer, a seller, money, and goods. Verbs that are linked to this frame are, for instance, *buy*, *sell*, *cost*, or *pay*. The idea is that individual words (e.g., verbs) can evoke a frame and the knowledge related to it. Each of the verbs in the Transaction Frame emphasizes different aspects of the frame. While *buy* highlights the buyer and the goods, *sell* emphasizes the seller and the goods (Fillmore, 1982).

We suggest that Fillmore's frame semantics theory is to a great extent compatible with research on thematic role assignment presented by McRae et al. (1997). The similarity between the two approaches is that both rely on thematic roles that are specific to individual lexical items. McRae et al. (1997, p. 137) define thematic roles as *verb-specific concepts*. In contrast to Fillmore (1985), their account is moreover specifically motivated by the goal to provide an account of the influence of this type of thematic role knowledge on on-line language comprehension. They argue that conceptual and world knowledge associated with verbs are rapidly available for comprehension processes such as incremental thematic role assignment. Their findings from on-line thematic role assignment during reading comprehension emphasize that it is important for a theory of incremental role assignment to take

into account the information that is provided in the form of experience-based verb-derived knowledge.

From the above discussion it has become clear that while an approach of the type that Dowty pursues has its appeals as a linguistic theory of thematic role generalization, and a Davidsonian theory as a formal characterization of truth-reference, neither of these theories provide an inventory of thematic roles that allows us to fully describe the influence of lexically-specific stored thematic role knowledge (e.g., that a cop is a typical agent of an arresting action) on incremental thematic role assignment. In contrast, theories of the type that Fillmore (1985) proposes, or the account of incremental thematic role assignment put forward by McRae et al. (1997) offer an inventory of lexically-specified thematic roles that permits description of the influence of stored thematic role knowledge on on-line comprehension. It should be noted that such a lexically specific thematic role theory is broadly compatible with other lexical semantic theories such as the one proposed by, for instance, Jackendoff (1983), Jackendoff (1990), and Jackendoff (2002).

SUMMARY

Taken together, results from the above studies provide convincing evidence that verb-based stored thematic role knowledge can contribute highly rapidly to inform comprehension processes such as structural disambiguation (e.g. Trueswell et al., 1994; McRae et al., 1998) and incremental thematic interpretation on-line (e.g., Altmann & Kamide, 1999; Kamide, Scheepers, & Altmann, 2003). In characterizing the influence of stored thematic role knowledge on comprehension the reviewed findings revealed a close time lock between when relevant thematic knowledge was made available by the utterance during lexical access and when it had an effect on comprehension. Evidence for this close time-lock was provided in particular by the eye-tracking data from studies by Trueswell et al. (1994). This indicates a high degree of coordination between the accessing of stored linguistic/world knowledge and on-line comprehension. We suggest that we might also find such a coordinated interplay between visual perception and comprehension in situations when processing of the scene is directly relevant for comprehension.

The rapid lexically-mediated effect of stored thematic role knowledge on comprehension provides support for the claim that such knowledge - including world knowledge that is associated with thematic roles - is tightly linked to lexical items, and should be considered in a theory of incremental thematic role assignment. A discussion of various thematic

role theories has drawn attention to the fact that only few thematic role theories can account for the experimental findings discussed in section 2.1.1. To describe the influence of stored thematic role knowledge that was observed in studies by Trueswell et al. (1994) and McRae et al. (1998), we require a thematic role theory that encompasses lexically-specified thematic roles in its inventory of thematic roles. Furthermore, from the point of view of language comprehension, a theory of incremental thematic role assignment should specify the processes and mechanisms that determine how thematic roles are assigned on he basis of such knowledge (see section 6.2.2).

2.2 THE IMPORTANCE OF THE IMMEDIATE SCENE

The above review of experimental research on the influence of stored thematic role knowledge about who-does-what-to-whom has drawn attention to the importance of this type of knowledge for thematic role assignment and for theories of thematic role assignment. While studies that examined the influence of *stored knowledge* on comprehension have in detail investigated thematic information and processes of thematic role assignment, psycholinguistic studies that explicitly investigated the influence of *scene information* have mainly considered the effects of a *referential context* established by a scene. Research on the influence of scene information on comprehension has investigated the time-course with which such a referential context influences the structuring and semantic interpretation of an utterance.

An essential pre-requisite for the rapid influence of information from the immediate scene on incremental thematic role assignment is its on-line availability. Stored thematic role knowledge associated with the words in an utterance becomes rapidly available in the course of lexical access (see section 2.1). For scene information this might not necessarily be the case since not all scene information is part of stored long-term knowledge that can rapidly be retrieved during lexical access (e.g., the immediate location of an object). However, we suggest that scene information can rapidly be made available in at least two ways.

First, in experiments investigating comprehension in visual scenes, the typical procedure ensures that people have time to inspect the scene prior to utterance onset. This time of free viewing permits people to inspect most of the available objects and to acquire available information from the scene. Such information should then, in principle, be available from episodic short-term memory to inform sentence comprehension processes while

the utterance unfolds (see Altmann, 2004). With utterance onset, scene information can furthermore be identified through words in the utterance. Upon hearing a word, the comprehension system can rapidly retrieve its meaning from memory (e.g., Dahan et al., 2001; Tanenhaus et al., 1996). Word meaning might then be used to search the scene, or a partially active episodic representation of the scene acquired during the preview time. Once reference from a word to an object has been established, scene information is available to inform language comprehension processes (e.g., Tanenhaus et al., 1995). We propose that, since scenes in existing psycholinguistic experiments on sentence comprehension typically contain relatively few objects, searching episodic memory or the immediate scene will be a rapid process. Scene information which is identified by the utterance should thus be rapidly available to inform comprehension processes.

Evidence from experiments by, for instance, Tanenhaus et al. (1996) provide indeed support for the claim that reference from nouns to scene objects is established rapidly during comprehension in visual scenes. They further demonstrate that referential ambiguity can delay identification of a target object. When a word (e.g., *candy*) was temporarily compatible with two scene objects (e.g., candle, candy), then both words (*candle/candy*) were considered during lexical access at the point when people had heard *can...*. Importantly, inspection of the target object (candy) was delayed when there was a competitor (candle), in comparison with when there was no such competitor (e.g., Tanenhaus et al., 1996). The important insight from these findings for our research question is that they show that unambiguous reference to a scene object can speed up the process of establishing reference, and thus ensure that scene information becomes rapidly available.

2.2.1 THE INFLUENCE OF A VISUAL REFERENTIAL CONTEXT

To further discuss the influence of scene information on comprehension, we re-consider studies by Tanenhaus et al. (1995) in more detail (see chapter 1 for a short summary of their findings). Example sentences from their study are given in 2.8 and 2.9. Recall that their findings revealed that the type of referential context (one-referent, two-referent context) can influence the incremental structuring of a temporarily structurally ambiguous utterance (2.8). The one-referent context in their studies contained one apple on a towel, an empty towel, a box, and a further object. In the two-referent context, the scene contained two apples (one on a towel, one on a napkin), an empty towel, and a box. Tanenhaus et al. (1995) investigated the effects of these two types of context on the resolution of a

prepositional phrase attachment ambiguity. In the ambiguity that they investigated, the prepositional phrase *on the towel* can temporarily be either attached to the verb phrase (indicating a destination) or be analysed as modifying the noun phrase (indicating a location) (see sentence 2.8). As a baseline for the examination of the ambiguous sentences, they included sentences in which the prepositional phrase was unambiguously attached to the noun phrase (see sentence 2.9). While the structuring of the utterance in the ambiguous condition (2.8) might differ depending on visual context, this should not be the case for the structurally unambiguous condition (2.9).

(2.8) Put the apple on the towel in the box.

(2.9) Put the apple that's on the towel in the box.

The experimental sentences in the studies by Tanenhaus et al. (1995) were instructions that participants had to carry out. To perform the task indicated in the instruction (moving an object to a destination), people had to first establish reference to the object identified by the instruction, and then to find the appropriate location to which they should move the object. We suggest that performing this type of task (carrying out instructions) requires that people have understood the utterance. An act-out task thus ensures that the ongoing cognitive processes are comprehension processes.

Ensuring that the cognitive processes reflected in the eye-movements were comprehension processes was important since Tanenhaus et al. (1995) used a then-new method for investigating the influence of scene information on comprehension (see also Cooper, 1974). They monitored people's eye-movements during scene inspection. In pilot studies, they found that while participants listened to the instructions, they rapidly inspected an object after hearing a relevant word in the utterance. From this close temporal link between utterance and inspection of an object, Tanenhaus and colleagues deduced that eye-movements to scene objects do indeed provide insight into the cognitive processes (e.g., establishing reference from a noun to an object) that take place during comprehension of the instruction.

During comprehension of the ambiguous sentence type (2.8) Tanenhaus et al. (1995) found the following eye-gaze pattern: When there was only one apple, people looked first at the apple, then to the empty towel after they had heard *Put the apple on the towel*. This gaze pattern revealed that people had interpreted the structurally ambiguous sentence fragment *on the towel* as a destination, attaching it to the verb phrase. In contrast,

after people had heard *Put the apple on the towel* in the two-referent context they first looked back and forth between the two apples, and then rapidly issued eye-movements to the correct destination (the box) without inspecting the empty towel. This revealed interpretation of the ambiguous prepositional phrase as a location that helped to distinguish two otherwise similar objects (an apple on a towel and an apple on a napkin). Comparison with the unambiguous sentence type baseline enabled interpretation of these findings. For the unambiguous sentence type (2.9) Tanenhaus et al. (1995) found that people's eye-gaze pattern did not change between the one-referent and the two-referent contexts. The time-course of establishing reference to objects for unambiguous sentences was the same as for the ambiguous ones in the two-referent condition. This comparison with the unambiguous sentences (where the prepositional phrase was unambiguously attached to the noun phrase) enabled interpretation of the eye-gaze pattern for ambiguous sentences in the two-referent condition as revealing initial attachment of the prepositional phrase to the noun rather than to the verb phrase.

Recall that when reviewing experimental evidence for the influence of stored thematic role knowledge, we highlighted a close time-lock between lexical access that made available stored linguistic and world knowledge and the time when this knowledge affected comprehension processes. We furthermore anticipated finding a similarly close temporal coordination during comprehension of utterances in a setting where scene information was relevant for comprehension (see the summary in section 2.1). Indeed, in the experiments by Tanenhaus et al. (1995) the time between when people heard a word and when they inspected an object was closely coordinated (see also Tanenhaus et al., 1996). This provides support for the claim that for processes of establishing reference to an object, utterance comprehension, and attention in the scene are tightly coordinated. What is less clear is whether the utterance also determines when perceptual processing of the scene has an effect upon the structuring of an utterance. Since eye-movements in the visual context differ from the very start between the two types of contexts, it is not possible to determine the precise moment when scene information affected comprehension.

To further investigate this interplay we consider studies by Sedivy et al. (1999). Their experiments demonstrated that visual contexts affect interpretation of temporary referential ambiguity in structurally unambiguous utterances. Specifically, they found that the time course of establishing reference to objects in a scene depended on whether there was referential contrast between two same-category objects (two glasses) or not. By 'visual referential context' we understand a context that contains two objects (e.g., two glasses)

whose referential uniqueness depends upon one differentiating property (e.g., their size - tall or small). The task was the same as in the studies by Tanenhaus et al. (1995). An example for the kind of instruction that people received is provided in example (2.10).

(2.10) Pick up the tall glass and put it below the pitcher.

In the referential contrast condition, an example scene contained two tall objects (a glass and a pitcher), a small glass, and a key. In the no-referential-contrast condition, the scene contained two tall objects (a glass and pitcher), a key, and a file folder, however, no contrasting object of the category 'glass'. The utterance beginning *Pick up the tall ...* was temporarily referentially ambiguous between the tall glass and the tall pitcher as target of the instruction. Sedivy et al. (1999) found that after people had heard *Pick up the tall*, they looked more quickly at the target referent (the tall glass) than at the other tall object (a pitcher) when the visual context displayed a contrasting object of the same category as the target (a small glass) than when it did not. The account which Sedivy et al. (1999) provided for their findings was that referential contrast between two same-category objects facilitated the incremental semantic interpretation of the utterance. Importantly, the above-discussed findings demonstrate that referentially relevant non-linguistic information influences rapidly how the linguistic input is structured (Tanenhaus et al., 1995) and how a sentence is interpreted semantically (Sedivy et al., 1999).

As argued above, Tanenhaus et al. (1995) provided evidence for a close temporal coordination between the lexical access of a noun with the time when people inspected an object. Studies by Sedivy et al. (1999) further extend this findings to adjectives that restrict the domain of reference in the scene. Importantly, in the studies by Sedivy et al. (1999) (in contrast to those by Tanenhaus and colleagues) the point in time when *scene information* influences the semantic interpretation of the utterance can be determined more precisely. Support for this view is provided by the observation that eye-movements to the target object (the tall glass) did not differ in the time graphs between the contrast and no-contrast conditions prior to the onset of the critical adjective *tall* (no inferential analyses were, however, provided for the time before adjective onset). It was only once people had processed the adjective *tall* that the difference in the type of context (contrast, no contrast) influenced processes of establishing reference to the target object as evidenced by the latency of eye-movements to the target object.

2.2.2 THE REFERENTIAL THEORY OF SENTENCE PROCESSING

In explaining the influence of scene information on comprehension, Tanenhaus et al. (1995) and Sedivy et al. (1999) have drawn upon the *Referential Theory* of sentence processing (e.g., Altmann & Steedman, 1988; Crain & Steedman, 1985). This theory has aimed at providing an account of structural ambiguity resolution which explains the choice of a syntactic analysis in terms of the referential relations between the words in a sentence and a referential context. Specifically, Crain and Steedman (1985) argued that the choice of analysis in the case of structural ambiguity depends on the *Principle of Parsimony*:

(2.11) *Principle of Parsimony*

> If there is a reading that carries fewer unsatisfied but consistent propositions or entailments than any other, then, other criteria of plausibility being equal, that reading will be adopted as most plausible by the hearer, and the presuppositions in question will be incorporated in his or her model.

This means that in the case of an ambiguity for which two or more syntactic analyses are possible, the comprehension system should choose the analysis that has the fewest unsatisfied presuppositions. According to Crain and Steedman (1985), a noun phrase which is postmodified by, for instance, a prepositional phrase presupposes that there is a set of entities of the same category as the noun phrase over which the modifier (the prepositional phrase) restricts (see also Altmann & Steedman, 1988).

More concretely, if there is, for instance, a definite noun phrase which could either be analysed as a simple noun phrase or as a complex noun phrase (e.g., being postmodified), then an analysis of the noun phrase as a simple noun phrase is supported by a context that contains one potential referent for the noun phrase. No further specification of the noun phrase is necessary since it uniquely identifies a referent. In contrast, in a context that contains two potential referents (e.g., two apples), a complex noun phrase should be built, since establishing unique reference requires that we distinguish between the two potential referents. A postmodifier of the noun phrase might precisely permit us to further specify which referent the noun phrase refers to.

For an ambiguous sentence such as the one in example 2.8, in which a prepositional phrase can temporarily either be attached to a noun phrase or to a verb phrase, this means the following: The prepositional phrase should only be analysed as a noun phrase modifier if there is more than one entity of the same category as the entity to which the initial

noun phrase refers (i.e., more than one apple). In contrast, a single noun phrase has fewer presuppositional requirements and can be felicitously analysed in a context where there is only one referent of the same category as the simple noun phrase.

SUMMARY

The discussion in the present section outlined that a visual referential context can rapidly influence the incremental structuring and semantic interpretation of an utterance. We further made clear that the effects of a referential visual context on sentence comprehension can straightforwardly be accounted for by the Referential Theory of Sentence Processing. In addition to providing a summary review of prior research on the influence of scene information we specifically paid close attention to the temporal interplay between on-line language comprehension and scene perception in the studies that we reviewed. Based on the observation of a close temporal coordination in the accessing of stored thematic role knowledge and its influence on comprehension processes (section 2.1) we explored the interplay between *scene perception* and utterance comprehension. We found that the temporal coordination of hearing a word and inspecting an object in the scene was well attested in both of the reviewed experiments (and indeed in many other studies such as Kamide, Altmann, and Haywood (2003), Kamide, Scheepers, and Altmann (2003), and Tanenhaus et al. (1996)). In contrast, the reverse temporal coordination was only present in the studies by Sedivy et al. (1999): In their studies alone we found a close temporal coordination between the time when the utterance drew attention to entities in the scene and when scene information related to that entity in turn influenced interpretation of the utterance.

2.3 IMPLICIT RELATIONS BETWEEN ENTITIES IN A SCENE

Despite convincing evidence for the influence of scene information, psycholinguistic research that explicitly investigates the influence of scene information on on-line sentence comprehension has so far only examined the effects of one type of information in scenes (a referential context that established object identity). Clearly, establishing the referential identity of an object is an important process in language comprehension. However, research on the influence of stored thematic role knowledge provides support for the view that it is in addition essential to establish who-does-what-to-whom. Indeed, thematic relations have been the focus of much theoretical investigation both in linguistic and psycholinguis-

tic research (see section 2.1). In contrast to the great attention that utterance-derived stored thematic role knowledge has received in establishing thematic relations in comprehension during reading, the effects of a visual context that provides thematic role in addition to referential information has so far not been *directly* investigated. This means that an important informational source - if not dimension of our life (i.e., the immediate environment) - for establishing thematic relations has been omitted from investigation. The impression which prior research has rendered of the influence of scene perception and of its role in the interplay with comprehension might thus not be complete. Specifically, we cannot be sure to know the full extent to which scene information might determine this interplay since only the influence of one type of information in scenes on comprehension has been investigated. To explore further whether scenes can establish important relations between entities, and whether these relations influence comprehension, we consider studies by Tanenhaus et al. (1995) Sedivy et al. (1999), and Kamide, Altmann, and Haywood (2003) from this point of view.

2.3.1 VISUAL REFERENTIAL CONTRAST

In reviewing the studies by Tanenhaus et al. (1995) and Sedivy et al. (1999) we have so far not analysed in more detail precisely which kind of information in their scenes influenced comprehension. Here we take the view that a more detailed analysis of their findings is necessary in order to better understand which kinds of information in a scene can influence the structuring and interpretation of an utterance. In the experiments by Tanenhaus et al. (1995) the authors suggested that the reason why the scene influenced the structuring of the utterance was the type of context (i.e., a referential context with two same-category objects made the noun phrase attachment of the prepositional phrase parsimonious). We propose, however, that it is not clear whether the presence of two-same category objects alone (i.e., two apples) would have influenced the structuring of the utterance. If the two apples had not been distinguishable from one another by means of their location (on a towel and on a napkin), or by some other property that uniquely established their respective identities, people could not have resolved the referential ambiguity. This means that the *contrast* between these two objects was an important characteristic of the scene in determining the structuring of the utterance. We consider contrast to be one type of relation between objects that a scene can establish and through which it can influence comprehension. For further support of this claim re-call the experiment by Sedivy et al.

(1999). In their studies, it was a differentiating attribute between two glasses (a small and a tall glass) that enabled rapid interpretation of the utterance. We propose that such minimal contrast between object attributes (e.g., tall/small; on a towel/not on a towel) is one instance of how the a perceiver can establish (non-linguistic) relations between otherwise similar entities. We propose that such contrastive relations arise through comparison by a perceiver, rather than being overtly realized in the scene, and hence represent an implicit relationship between entities.

2.3.2 ENTITY AFFORDANCES

In the above-discussed experiments, the scene has contributed information about an *immediate* context that is probably not part of long-term stored linguistic/world knowledge (e.g., the immediate location of an apple) towards utterance interpretation. An important point which we have so far omitted from our discussion is the possibility that scenes might furthermore provide *stored* knowledge such as thematic role knowledge during comprehension. Prior research that has directly investigated the influence of stored linguistic/world knowledge during comprehension in visual scenes (e.g., Altmann & Kamide, 1999; Kamide, Scheepers, & Altmann, 2003) has paid little attention to the possibility that both utterance and scene (rather than the utterance alone) might make stored knowledge available for comprehension. The present section discusses this issue.

To explore this idea consider a study by Kamide, Altmann, and Haywood (2003). They investigated how compositional integration of a sentence-initial agent noun phrase and an ensuing verb restricts expectations about post-verbal goals. An example image contained a girl, a man, a motorbike, a carrousel, a candy jar, and a beer tankard. An item sentence set from the experiment is presented in examples 2.12 to 2.15. Consider first the examples in 2.12 and 2.13. After people had heard *The girl will ride...*, Kamide, Altmann, and Haywood (2003) observed more anticipatory eye-movements to the one appropriate goal for a girl-riding action (the carrousel), than to the other goal of a riding action (the motorbike). When the sentence beginning was instead *The man will ride ...*, the motorbike as the most plausible goal of a man-riding action received more anticipatory eye-movements than the carrousel. To exclude an interpretation of these findings in terms of a low-level association account Kamide, Altmann, and Haywood (2003) introduced an additional two sentences (2.14 and 2.15). If eye-movements during the post-verbal region resulted solely from an association of the first noun phrase *the man* (2.13) with available

plausible scene objects (e.g., a motorbike or beer) rather than from compositional semantic interpretation of the first noun phrase and the verb (e.g., *ride*, 2.13), then the motorbike should receive equally many inspections after *The man will ride ...* as after *The man will taste ....* Eye-gaze patterns showed that this was not the case. Rather, an analysis of the eye-movement patterns during the post-verbal region showed that comprehension processes were influenced by the combination of the semantics of the first noun phrase, verb plausibility, and verb-based world knowledge. The authors interpret these findings as revealing that the combinatory interpretation of the first noun phrase and the ensuing verb determined expectations about which of the available scene objects would most likely fill the thematic role of the post-verbal goal in the sentence.

(2.12) The girl will ride the carrousel.

(2.13) The man will ride the motorbike.

(2.14) The girl will taste the sweets.

(2.15) The man will taste the beer.

While the baseline sentences (2.14 and 2.15) exclude an association account, they do not exclude an account that attributes the findings to the combined effects of compositional semantic interpretation of the linguistic input and entity *affordances* (e.g., a carrousel affords a child-riding action, see J. Gibson (1966); see further Steedman (2002a) and Steedman (2002b) for a formal description of object and event affordances). Kamide, Altmann, and Haywood (2003) acknowledged the possibility that the referential context provided by the scene in their studies might have contributed to their findings. Here, we reviewed their studies to draw attention to the potential importance of entity affordances as a further way in which a referential scene context can establish relations between entities. Based on this review we argue that the above findings by Kamide, Altmann, and Haywood (2003) fit in well with our account that what makes scene information influential in comprehension is its ability to establish relations between entities. In studies by Kamide, Altmann, and Haywood (2003), presumably thematic/world knowledge from both utterance and scene enabled perceivers to establish thematic relations between a man/girl and a plausible goal of a man-/girl-riding action (a motorbike and a carrousel respectively). Just like in the studies discussed in section 2.3.1, the scene did not explicitly depict the relations between the objects. Rather it was knowledge associated with a noun phrase/verb or a scene object that enabled utterance interpretation.

SUMMARY

To identify further types of information in scenes that might be influential in incremental comprehension, we analysed the causes that underlie the influence of scene information (section 2.3). We have proposed that what makes scene information relevant for comprehension processes is its ability to establish meaningful relations between entities. One possibility that we identified was that the scene can establish relations between entities through world knowledge that is associated with an entity, and that might become available upon inspection of the entity (e.g., a carrousel affords a child-riding action). A further option that we identified was that the scene can establish relations between entities through contrast. We characterize these two ways in which the scene can provide information about relations between entities as *implicit* relations. Such relations arise through comparison of objects in the eyes of a perceiver. Based on the discussion of findings by Tanenhaus et al. (1995), Sedivy et al. (1999), and Kamide, Altmann, and Haywood (2003), we conclude that such important relations between entities can influence structural disambiguation and incremental interpretation of an utterance.

2.4 EXPLICITLY DEPICTED AGENT-ACTION-PATIENT EVENTS

The preceding sections (2.1, 2.2, and 2.3) have provided relevant background knowledge for the suggestions that we will put forward in the present and ensuing sections. Recall that we outlined two goals of the present thesis in the introductory paragraphs of this chapter and in chapter 1. One was to investigate the types of information in scenes that can influence comprehension processes such as structural disambiguation and incremental thematic role assignment. Our second aim was to consider in more detail the nature of the interplay between visual perception and comprehension. The present section deals with the first of these two goals. Section 2.5 will discuss the second goal.

Before considering the first objective, we briefly recapitulate the most important insights that we gained from the previous sections. Section 2.1 has outlined that there is evidence for the rapid influence of stored thematic role knowledge on incremental structural disambiguation and thematic role assignment. In reviewing the insights that prior research has provided about the influence of scene information, we pointed out that there is convincing evidence for the importance of scene information in the incremental structuring and interpretation of an utterance (see section 2.2). However, direct investigation

of the influence of scene information has so far been limited to visual referential contexts that provide referential information about objects (i.e., their location or size). Accordingly, the focus of investigation has been on the time-course with which referential scene information influences processes of establishing reference to an entity. The effects of scene information on *incremental thematic role assignment* have not yet been directly investigated in psycholinguistic research on on-line sentence comprehension. Researchers in this area have not explicitly examined whether other types of information in scenes (e.g., about the thematic role of an entity) can rapidly affect comprehension. An analysis of potential causes underlying the effects of a visual referential context on comprehension revealed that an important reason for its influence might be the fact that scenes can establish relations between entities. We identified two such types of (implicit) relations between entities (referential contrast and entity affordances, see section 2.3).

We importantly propose that there are further ways in which the scene can establish relations between entities and contribute to incremental thematic role assignment. We argue that visual scenes crucially can explicitly establish important relations between entities in addition to identifying them. Indeed, scene information such as depicted actions and events might have the potential to establish *thematic relations* between entities. Through these relations, the visual context may establish, for example, an entity's role with respect to other objects. It is suggested events are one ontological category that enables visual scenes to provide information about role-relations between entities. This proposal receives support from the fact that events are a central component of thematic role theories such as frame semantics or Davidsonian semantics (e.g., Fillmore, 1982; Davidson, 1986a, 1986b).

There is little research on the on-line influence of depicted events on comprehension processes such as incremental thematic role assignment, and little is known about how reference to depicted events is established on-line. Indeed, world-language relations other than how language establishes reference to entities have received little attention in psycholinguistic research on spoken sentence comprehension. One reason for this is that environments in preceding experiments in this research area only contained inaminate entities (such as objects, animals, and characters). Clearly, a scene displaying only entities represents a highly restricted setting, severely limiting the influence that visual information can exercise on incremental comprehension.

We propose that just as language can refer to entities, it can also refer to other ontological categories such as actions or events. While the definition of the term reference has generally only been extended to cover reference from nouns to the ontological category

of things (Crystal, 2003), the experiments reported in this thesis are compatible with a notion of 'reference' that includes reference from linguistic expressions like verbs and sentences to a broad variety of corresponding distinct ontological categories such as actions and events (e.g., Jackendoff, 1983). We propose reference to these ontological categories can be established on-line.

If on-line comprehension mechanisms exploit a wider range of ontological categories than just entities (e.g., actions and events), then we need to establish reference to them through, for example, verbs and sentences in order to describe how depicted actions and events can be exploited for on-line sentence comprehension. Theories of on-line sentence processing and incremental thematic role assignment would in this case require a suitably rich array of ontological categories and corresponding referential expressions, as well as mental representations for both. In particular, they would need to be based upon a theory that allows reference to a variety of ontological categories such as actions and events in addition to entities. One framework, which fulfils this requirement, is Jackendoff's semantic theory (1983, 1990, 2002). It allows us to describe reference to both actions and events and thus to account for how we access information about role-relations in the immediate visual context in a setting that offers combined linguistic and visual information. We return to this issue in sections 6.2.1 to 6.3.1 where we further explore the Referential Theory, current thematic role theories and the Jackendovian framework under this perspective.

For the reminder of the thesis, we concentrate on agent-action-patient events. Actions in such events determine a character's role (carrying out an action for agents), and establish role-relations between characters. If we see a woman greeting a man, the event information tells us that the woman is the agent of the greeting action, and the man the patient. Eye-movement studies of sentence production, for example, have already revealed a close link between visual attention in depicted action scenes and incremental production processes (e.g., Griffin & Bock, 2000). It should therefore be possible to understand role-relations established through actions when they are referred to by linguistic expressions such as verbs during comprehension. Making agent-action-patient events available in a scene might thus permit a more "active" influence of visual information on on-line comprehension processes, since such events provide propositional as well as referential information.

2.5 COMPREHENSION AND VISUAL PERCEPTION

Crucially, the fact that one informational source affects comprehension processes reveals only a limited amount of insights about how comprehension proceeds. In a situation where diverse types of information are available and relevant, the nature of comprehension processes is rather determined by how the available kinds of information combine (as opposed to the fact that they influence comprehension processes at all).

In chapter 1 we have drawn attention to the fact that there is strong experimental evidence for the claim that diverse informational sources interact rapidly and unrestrictedly with one another (e.g., Altmann & Kamide, 1999; Chambers et al., 2002; Kaiser & Trueswell, in press; Kamide, Altmann, & Haywood, 2003; Kamide, Scheepers, & Altmann, 2003; MacDonald et al., 1994; Sedivy et al., 1999; Sussman & Sedivy, 2003; Tanenhaus et al., 1995; Tyler & Marslen-Wilson, 1977). Recent results from modelling research (e.g., McRae et al., 1998; Tanenhaus et al., 2000) further support the view that the integration of various kinds of information (e.g., lexical, structural and thematic) is best modelled as an unrestricted interaction of constraints. In theory, such a model could be extended to incorporate information that has been derived from a visual scene (see chapter 1).

But is such an interaction account really a fully specified theory or model that details the mechanisms which govern the construction of mental representations in sentence comprehension? In chapter 1 we have argued that this is not the case. We have reasoned that constraint-based accounts do not specify how mental representations are derived and how they combine in the interpretation of a sentence. We further proposed that the interplay between visual perception and comprehension might be of a more coordinated nature than unrestricted constraint-based accounts suggest (see section 1.3 for discussion). Recall further that we observed a close temporal coordination between when stored linguistic and thematic role knowledge was retrieved during lexical access and when it informed comprehension processes such as the incremental assignment of thematic roles (e.g,. Altmann & Kamide, 1999; Kamide, Scheepers, & Altmann, 2003) and structural disambiguation (e.g., McRae et al., 1998; Trueswell et al., 1994) (see section 2.1.1). In what follows we explore whether there might be a similarly close *temporal* coordination between on-line comprehension processes and perception of the scene, and in turn between processing of scene information and core comprehension processes such as incremental thematic role assignment and structural disambiguation (see section 2.2 for a more detailed discussion).

Recall the comprehension situation that we sketched at the beginning of this chapter.

People inspect a scene, then a relevant utterance is presented. In this situation, distinct informational sources such as utterance, stored linguistic/world knowledge, and scene information are available. While, for instance, inspection of the scene prior to utterance onset is relatively unconstrained, the situation changes with utterance onset. With onset of the utterance, the words in the unfolding utterance strongly influence inspection of the scene. Prior research has demonstrated that people rapidly direct their attention to those objects in the scene that have just been mentioned (e.g., Tanenhaus et al., 1995), or to objects they expect to be mentioned next (e.g., Altmann & Kamide, 1999). Although this attentional link can be broken in simple discrimination tasks (see Rayner, 1998), there is probably a tight connection between the locus of attention and the locus of fixations in complex tasks such as language comprehension.

The utterance could thus, by directing or failing to direct attention to scene parts, determine *which* parts of the scene or of stored thematic role knowledge are available for comprehension of the utterance. Reference from a noun in the utterance to a scene entity, for instance, directly identifies the entity and makes it available for comprehension. Words in the utterance lead to the rapid retrieval of lexical meaning and of thematic role knowledge that is associated with the meaning of a word. In addition to identifying available information, the incremental order with which words in the utterance are encountered could moreover determine *when* an informational source is identified as relevant for comprehension.

We propose that scene information such as depicted events can rapidly influence incremental thematic role assignment and structural disambiguation if the depicted events are directly identified as relevant by the utterance and closely linked to when their identification occurs. Upon encountering, for instance, a verb in the utterance that directly identifies a relevant depicted action in the scene, the comprehension system should, in principle, be able to rapidly avail itself of the depicted thematic relations associated with the actions for interpretation of the utterance.

What, however, if the scene does contain relevant thematic role relations, yet the utterance does *not directly* identify these relations as relevant? A number of studies that we discussed in section 2.1 have provided evidence for the claim that referential identification of an object can be constrained in other ways than through direct noun - object reference. Among utterance-derived kinds of information that have been found to constrain the domain of reference are, for instance, prepositions, verb selectional restrictions, or case-marking that determined the grammatical function of an entity. Here, we build on the

insight that entities in the scene can be identified through utterance constraints in an indirect way. Could depicted events also be accessed through more subtle linguistic cues as opposed to being identified through direct reference from, for example, a verb to an action? If so, then this would indeed be evidence for an active contribution of scene information such as depicted events to on-line language comprehension.

The utterance emerges from this discussion as a crucial factor in determining the time-course with which available informational sources can affect language comprehension. In the two situations sketched above the utterance determines the use of relevant thematic role information from the scene through either direct reference, or subtle linguistic cues. Does this mean that it is an invariant guiding force in language comprehension? And what does such a paramount status of the utterance entail for the role of depicted events in incremental thematic role assignment? Let us explore these questions in more detail.

An even stronger test of the importance of depicted events than the above two cases would be a situation in which the utterance does not uniquely identify which informational source (e.g., depicted events or stored thematic role knowledge) is relevant for the construction of mental representations in incremental thematic role assignment. In the examples that we discussed so far, the utterance always uniquely identified the informational source on the basis of which the comprehension system should build a mental representation of the unfolding utterance. In a situation where the utterance does not have this important constraining effect we should, in principle, be able to observe which informational source the comprehension system chooses itself for interpreting a sentence. What is the *relative influence* of stored thematic role knowledge and depicted event information in a situation where diverse types of information such as scene and stored knowledge are both available, and are both identified as relevant by the utterance? Does stored thematic role knowledge determine interpretation of the unfolding utterance, or does the comprehension system rely on thematic role information provided by the immediate visual scene?

The research question of the relative importance of two informational sources has to the best of our knowledge so far not been directly investigated. As a result, no existing psycholinguistic account or theory appears to make a direct prediction on the matter. It seems that there are at least three potential outcomes. Either we find a priority of stored thematic role knowledge, or of depicted events, or neither of the two. In the absence of explicit predictions, we associate different strands of research with these three potential outcomes.

One possibility is the finding of a relative priority of stored thematic knowledge in

incremental thematic role assignment. As detailed in section 2.1.1, research by Altmann and Kamide (1999) and Kamide, Scheepers, and Altmann (2003) has shown that stored linguistic and world knowledge can enable listeners to form expectations as to which scene object will be mentioned next. We suggest that the basis for the expectations that they observed in anticipatory eye-movements might be mental representations of thematic relations that have been built on the basis of stored linguistic and world knowledge. It should be noted that the studies by Altmann and colleagues do not demonstrate a priority of stored thematic role knowledge. However, they do reveal the rapid incremental influence of stored thematic role knowledge and might thus be associated with a great importance of this type of knowledge.

A further possible finding is a greater importance of depicted events relative to stored thematic role knowledge. While the work by Tanenhaus and colleagues does not show a priority of depicted information in comparison with stored thematic role knowledge, the strong influence of scene information in their studies points to a great importance of scene information for comprehension. In their studies, the process of building mental representations of the unfolding utterance could not be informed by structural relations derived from the utterance. Utterances were structurally ambiguous and neither utterance nor scene provided stereotypical knowledge. What influenced the structuring of the utterance was rather the type of referential context established by the scene.

It should be noted that the precise timing of the influence of scene information on the structuring of an utterance in the studies by Tanenhaus et al. (1995) did not appear to be closely linked to incremental utterance comprehension (see section 2.2 for discussion). We view their research as compatible with an account where the effects of scene information are not tightly coordinated with on-line comprehension. As a variant of this account we identify an approach that we will henceforth refer to as 'coordinated interplay' account. The difference between the two approaches is precisely that in a coordinated interplay account we expect a close temporal coordination between utterance comprehension and the effects of scene information on interpretation of the utterance. Under this view, lexical access makes available word meaning and associated linguistic/world knowledge that can then be used to rapidly identify relevant parts of the scene. The scene parts can, once identified, in turn inform incremental comprehension.

Finally, in contrast to the above options, a constraint-based interactionist framework such as the one by Tanenhaus et al. (2000), which we introduced in chapter 1, would predict no preference for either stored thematic role knowledge or depicted events (provided

that these two types of information are comparable in strength). If they are, the model expects comprehension difficulty. Competition between these two constraints should be high and none of them should, in principle, be able to rapidly inform comprehension in preference over the other.

To summarize the main conclusions from the discussion in sections 2.4 and 2.5: Section 2.4 has argued that to extend existing insights on the influence of a visual referential context during comprehension, we must investigate the effects of scenes that provide propositional in addition to referential information. We suggested depicted events are one type of scene information that does provide propositional information in addition to identifying entities. Our experiments will thus focus on investigating the influence of depicted events on incremental comprehension. Section 2.5 explored the nature of the interaction between the perception of scene events and on-line utterance comprehension. From this discussion it became clear that the utterance might play an important role in determining the timing with which scene information and stored thematic role knowledge inform on-line comprehension. Our experiments will examine whether this is indeed the case. Further to investigating the temporal coordination between the perception of depicted events and their influence on incremental thematic role assignment, we ask how important depicted events are precisely for on-line comprehension. Are they "just another" informational source among many, or do they have a certain priority for on-line sentence comprehension processes? Investigating this issue should, in principle, enable us to decide between the various accounts of the relative importance of depicted events in comparison with stored thematic role knowledge that we sketched further above. Taken together, we identify three situations that require consideration in answering the research questions that the discussion in sections 2.4 and 2.5 has motivated.

(a) The verb in the utterance directly identifies depicted events early in the utterance.

(b) Depicted actions are only referred to directly by the verb late in the utterance. Subtle linguistic cues, however, bias early on towards the use of relevant depicted events.

(c) The utterance identifies two informational sources - a stereotypical agent and an agent of a depicted event. Both scene and stored knowledge offer relevant information about relations between scene entities.

We propose that investigating these research questions will enable us to determine in

more detail the nature of the interplay between visual perception and on-line language comprehension. We now discuss each of the research questions in turn, motivate in more detail why they are relevant cases of investigation, and provide a brief outline of how we will investigate them experimentally. A more specific introduction to individual experiments can be found at the beginning of each of the experimental chapters (chapters 3 to 5).

THE VERB DIRECTLY IDENTIFIES DEPICTED EVENTS EARLY: EXPERIMENTS 1 AND 2

In chapter 3, two studies are presented that explore the importance of events in depicted scenes for on-line language comprehension. Recall that in section 2.2, we outlined that Tanenhaus et al. (1996) had found that establishing reference to scene objects was delayed when a word in the utterance did not uniquely identify one scene object. For our inquiry, this means that we must uniquely identify depicted actions and events in the scene so as to make them rapidly available for comprehension processes. This could happen through direct identification of the action by means of a verb in the utterance. If we are able to also - just as in the case of objects - establish reference to scene *events* on-line, then relevant thematic role information from the scene should be available rapidly during comprehension. The rapid availability of scene events should be further facilitated through the fact that most of the scene is readily present in episodic memory from inspection of the scene prior to utterance onset.

In Experiments 1 and 2 in chapter 3, the verb in the utterance directly identified the relevant depicted event early during the utterance. We examine whether perception of event relations between entities allows the rapid recovery of mental representations such as thematic roles in on-line comprehension of initially ambiguous sentences, and actively influences incremental resolution of initial structural and role ambiguity in the linguistic input.

Participants inspected agent-action-patient events (e.g., princess-washing-pirate, fencer-washing-princess) while listening to German verb-second sentences with initial structural and role ambiguity. The experiments investigated the time course with which listeners could resolve this ambiguity by relating the verb in the utterance to the depicted events. Depending on the verb people heard (*wash/paint*), the scene identified the princess as an agent (princess-washing) or as the patient of an event (fencer-painting-princess). Such verb-mediated visual event information should, in principle, allow early disambiguation on-line if thematic role information that is present in the immediate visual environment is

available rapidly enough so as to influence incremental thematic role assignment on-line.

SOFT LINGUISTIC CUES AND SCENE EVENTS: EXPERIMENTS 3 AND 4

Direct identification of scene objects and events is only one possibility for making scene information available during comprehension. A further option is through more indirect and subtle cues in the utterance. If we only indirectly refer to depicted events, we make it presumably more difficult for the comprehension system to avail itself of the depicted event information. Hence, testing whether the comprehension system is able to use depicted events for incremental thematic role assignment under these conditions provides a particularly strong test of how actively they influence incremental thematic role assignment. Most experiments on comprehension in visual scenes have investigated how interpretation of the utterance constrains reference to objects in the scene. We examine whether the influence of thematic relations in the scene can determine thematic interpretation and structuring of the utterance, and this even if depicted events are only identified as relevant through subtle linguistic cues.

In chapter 4, experiments 3 and 4 consider the influence of subtle linguistic cues on the use of depicted event scenes for incremental thematic role assignment and structural disambiguation. They moreover generalize findings from the first two experiments to two other sentence constructions (German verb-final, English main verb/reduced relative constructions) and to another language (English).

Experiment 3 on structurally ambiguous German verb-final active/passive sentences used the same images as Experiment 1. We investigated the time course of incremental thematic role assignment when the verb was sentence-final and thus did not establish early reference to the depicted events. Delaying the verb - the essential cue that made available the depicted actions - allowed us to test how active a role visual scenes play in on-line thematic role assignment. We tested whether in this case an active/passive bias occurring before the second argument and main verb would make available the depicted events for thematic role assignment. An active bias should identify the princess as the agent of the washing event, while a passive bias should identify her as the patient of the fencer-painting event.

Experiment 4 examined the comprehension of the English main verb/reduced relative clause ambiguity. We explored whether the findings for the German SVO/OVS sentences in Experiment 1 generalize and extend to initially ambiguous English main verb and

reduced-relative clauses. The fourth experiment further tested the use of subtle linguistic cues (auxiliaries) that preceded the main verb of the sentence in incremental thematic role assignment.

THE RELATIVE IMPORTANCE OF INFORMATIONAL SOURCES : EXPERIMENT 5

Experiments 1 to 4 examined whether depicted events can influence incremental thematic role assignment both when directly referred to by the verb and when identified more indirectly through subtle linguistic biases. Experiment 5 (presented in the fifth chapter) goes beyond investigating the influence of a single informational source on language comprehension. Rather, it aims at comparing the relative importance of diverse kinds of information. Specifically, Experiment 5 examined the relative importance of stored thematic knowledge in comparison with depicted events for incremental thematic interpretation.

Within a single study we examined two research questions. First, we examined whether depicted events and stored thematic role knowledge can each inform incremental thematic interpretation when the verb *uniquely* identifies either a stereotypical agent or an agent of depicted event. We contrasted this with a situation where the utterance did *not determine uniquely* whether stored thematic role knowledge or depicted events provided relevant information for thematic role assignment. In this case, it should be possible to observe which type of knowledge (stored thematic role knowledge or depicted scene event) the comprehension system prefers to rely on in interpreting the utterance. Observing a relative priority for one over the other kind of information would reveal how much importance the comprehension system really attributes to one such informational source in comparison with another. We argue that establishing the relative priority of stored knowledge versus depicted events constitutes an important step towards developing a more fully specified account of on-line sentence comprehension in visual scenes.

CHAPTER 3

THE INFLUENCE OF DEPICTED EVENTS

The first two experiments[1] reported in this thesis investigate the time course with which verb-mediated agent-action-patient events allow disambiguation of the local structural and thematic role ambiguity that arises due to constituent order variability in German.

German has a rich case marking system where grammatical function is usually indicated by case morphemes. Still, there is some case ambiguity. The subject case (NOM) and the object case (ACC) are identical for feminine noun phrases in German. While both SVO and OVS ordering is possible, SVO is easier to process (e.g., Hemforth, 1993), and is the canonical order used in the description of an event in isolated sentences. Case ambiguity and word order variation lead to grammatical function and thematic role ambiguity. A sentence-initial feminine noun phrase in the nominative/accusative case could thus be either the subject or the object of the sentence.

Furthermore, in German SVO/OVS main clauses, the verb does not always disambiguate role and grammatical function of an ambiguous noun phrase that precedes it. People thus often realize only during a second noun phrase marked for subject-case that their subject-analysis of an ambiguous initial noun phrase was wrong. This has been revealed by reading times on the second noun phrase for OVS in comparison with SVO sentences. When people had read a sentence fragment in which the initial noun phrase

[1] Experiments 1 and 2 in the present chapter were previously published as part of the journal article "The influence of the immediate visual context on incremental thematic role assignment: evidence from eye-tracking in depicted events", see Knoeferle, Crocker, Scheepers, and Pickering (2005)

was ambiguously case-marked (i.e., could be the subject or object of the sentence), reading times on a second noun phrase where higher when the second noun phrase disambiguated the sentence towards an OVS than when it disambiguated the sentence towards a SVO structure (Hemforth, 1993). This has been taken as support for the claim that participants usually interpret a sentence-initial ambiguous noun phrase as subject (for related experiments see, e.g., Mecklinger, Schriefers, Steinhauer, and Friederici (1995), and Schriefers, Friederici, and Kühn (1995)). The experiments presented in this chapter build on the studies by Hemforth (1993). They investigate the same initially ambiguous sentence structure (German SVO/OVS main clause sentences). In contrast to the studies by Hemforth (1993) which examined the comprehension of such sentences during reading, our experiments examine the auditory comprehension of such sentences during the inspection of event scenes. In particular, we examine whether early disambiguation, prior to case-marking on the second noun phrase, is possible in the presence of disambiguating depicted events.

Experiment 1 examined the comprehension of initially structurally ambiguous spoken German SVO/OVS main clauses of the form NP1-V-ADV-NP2 (see Table 3.1). Sentences described event scenes showing a role-ambiguous character, acting and being acted upon (e.g., a princess), and two role-unambiguous characters that were agent and patient (e.g., a fencer and a pirate) respectively (see Fig. 3.1 a). In contrast to previous studies investigating the influence of visual contrast between entities on how we establish reference to objects (e.g., Sedivy et al., 1999; Tanenhaus et al., 1995) (see section 1.2 and 2.2.2), reference to an entity was never ambiguous on the images in our experiments. Rather, sentences and images contained role-ambiguity. The first noun phrase of the sentence and its referent in the scene were always role-ambiguous, and the initial structural and role ambiguity in the linguistic stimuli persisted up to the second noun phrase, which was unambiguously case-marked as subject or object.

In contrast to studies by Kamide, Scheepers, and Altmann (2003) (see section 2.1), information about thematic role-relations between the sentential arguments was not provided by the linguistic input prior to the second noun phrase in the first two experiments presented in this chapter. Sentences for the two experiments presented in this chapter were always initially structurally and role ambiguous. We ensured further that stereotypical relationships were absent from our materials. The relation between agents and actions (princess-washing), between agents and patients (princess-pirate), and between actions and patients (washing-pirate) were kept non-stereotypical (see Fig. 3.1 and Table 3.1). The absence of typical noun-verb relationships, and the presence of depicted

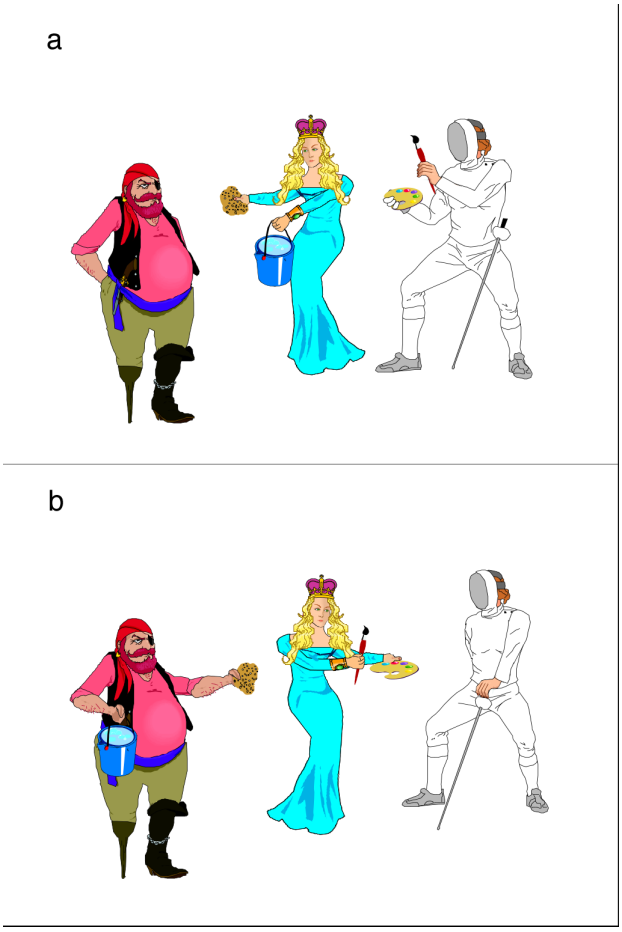


Figure 3.1: *Example image for sentences in Table 3.1*

actions that established thematic role-relations in the immediate visual context, crucially differentiates our experiment design from previous experiments by Altmann and Kamide (1999), Kamide, Altmann, and Haywood (2003), and Kamide, Scheepers, and Altmann (2003). In their studies, the scene contained only objects, and thematic role assignment was enabled through linguistic and world (including stereotypical) knowledge derived from the utterance. In our experiments, in contrast, thematic role-information was provided

by the scene events alone (prior to the unambiguously case-marked second noun phrase). Early disambiguation and role assignment would only be possible through the combined influence of sentence and the immediate visual context determining a character's role. The depicted actions differentiate between the event in which the princess is an agent (princess-washing-pirate) as described by SVO sentences, and the event in which she is a patient (fencer-painting-princess) as described by OVS sentences (see Fig. 3.1 a, Table 3.1 (1a) and (1b)). Once the verb is encountered, the depicted actions that are necessary to establish role-relations between agents and patients in the events could be accessed.

If such use of verb-mediated depicted events influences incremental thematic role assignment, then we argue that psycholinguistic theories of sentence comprehension must incorporate richer semantic inventories than have been explicitly incorporated into theories of sentence processing to-date (e.g., Jackendoff, 1983, 1990, 2002). A rich inventory of mental representations that include actions and events would allow us to describe reference to both actions and events and thus to account for how we access information about role-relations in the immediate visual context in a setting that offers combined linguistic and visual information.

If verb-mediated identification of depicted events rapidly enables early thematic role assignment in Experiment 1 then expectations due to thematic role assignment should be revealed in eye-movements in the event scenes. Based on findings from Altmann and Kamide (1999), and Kamide, Scheepers, and Altmann (2003), eye-movements for early disambiguation should be anticipatory; that is, they should occur before participants heard the respective second noun phrase. For the SVO condition, participants hear *Die Prinzessin wäscht ...* ('The princess (amb.) washes ...'), and the scene shows a princess-washing event with the pirate as the patient (Fig. 3.1 a). We expect that on encountering the verb 'washes', participants interpret the pirate as the patient of the washing-event and hence direct more anticipatory eye-movements to the pirate for SVO than for OVS sentences. In the OVS condition, participants hear *Die Prinzessin malt ...* ('The princess (amb.) paints ...'). In this case, the verb ('paints') does not match the action performed by the princess. Rather, the scene identifies her as the patient of the fencer-painting event (see Fig. 3.1 a). Expectations of the fencer as the appropriate agent and thematic role interpretation of the princess as the patient of the painting event should be revealed by a higher percentage of fixations on the fencer for OVS than for SVO sentences. We expect this pattern to continue during the second noun phrase, which presents a further point of disambiguation through object or subject case-marking in the linguistic stimuli. Experi-

ment 2 was aimed at replicating the findings from the first experiment while examining the influence of intonation.

3.1 EXPERIMENT 1

3.1.1 METHOD

PARTICIPANTS

Twenty-eight German native speakers with normal or corrected-to-normal vision were paid five euro for taking part in the experiment.

MATERIALS

We created 48 images using commercially available clipart and graphic programs. Images were pre-tested to ensure that participants were able to accurately recognize the agent-action-patient events and to discriminate between the two actions on an image. Sentences used in the pretest only differed from the experimental items in that they were all in SVO order. For each depicted action twenty participants saw both a matching and a non-matching sentence. For Figure 3.1 a, participants read *Die Prinzessin wäscht/malt offensichtlich den Pirat* ('The princess (amb.) washes/paints apparently the pirate (ACC)') and *Der Fechter malt/wäscht offensichtlich die Prinzessin* ('The fencer (NOM) paints/washes apparently the princess (ACC)'). Participants were asked to judge if a sentence felicitously described an event on the image (yes) or not (no). The overall percentage of correct answers was 98.60.

A set of 24 items was created. Each item consisted of four spoken sentences and two pictures (see Table 3.1 and Fig. 3.1). An example image (Fig. 3.1 a) showed a princess, a pirate and a fencer. The princess was washing the pirate, and the fencer was painting the princess. Thus, the princess was role-ambiguous (agent and patient); the pirate was the patient of the princess-washing action, and the princess was the patient of the fencer-painting action. To ensure that each of the role-unambiguous entities (i.e., the pirate and the fencer) was an equally good patient as agent, we created two versions of each image, which only differed in the roles of the depicted characters (the pirate and the fencer); depiction of actions (washing, painting) as such did not change.

For the scene in Figure 3.1a, two sentences were recorded: *Die Prinzessin wäscht offensichtlich den Pirat.* ('The princess (amb.) is apparently washing the pirate (obj.)') and

Die Prinzessin malt offensichtlich der Fechter ('The princess (amb.) is apparently painted by the fencer (subj.)') (see Table 3.1, (1a) and (1b) and Fig. 3.1 a). One sentence per image thus had a canonical SVO and the other an OVS word order. Grammatical subject and object corresponded to agent and patient on the images respectively. Conditions were matched for length and frequency of lemmas using CELEX (Baayen, Pipenbrock, & Gulikers, 1995) for all the words within an item. All the sentences had an SVO intonation up to the second noun phrase. For OVS sentences (e.g., Table 3.1 (1b)), the sentence beginning of the SVO sentence (e.g., Table 3.1, (2a)) was spliced in before the second noun phrase. This ensured that OVS intonation cues could not contribute to disambiguation.

We expect comprehension processes of establishing reference and assigning thematic roles to manifest themselves in the eye-movement pattern when participants process sentences and scenes. When listeners hear 'The princess', we expect they will identify the princess as the appropriate referent in the scene. Hence, they should look predominantly at the princess. On hearing the verb 'washes' for the SVO sentence, it becomes clear that the princess is the agent of the washing event. Both the orientation of the princess towards the pirate and the depicted action provide information that the pirate is the target of the action. As a result, more eye-movements should be made to the pirate (the patient) than to the fencer (the agent) shortly after the verb in SVO sentences. When the second noun phrase ('the pirate' (obj.)) is encountered, the fixations on the pirate should continue, since expectations of the pirate as the patient of the washing action will have been confirmed by object-case marking on the second noun phrase. For the OVS condition, people hear 'The princess paints'. They presumably realize that the princess is not performing a painting action, and look for the agent of the painting event (the fencer). Shortly after the verb, participants should inspect the fencer more often than the pirate. On hearing 'the fencer', their expectation of the fencer as the agent of the painting action is confirmed by subject case marking on this noun phrase, and eye-movements to the fencer should further increase. Crucially, for SVO sentences more anticipatory fixations should occur to the pirate as the patient of the event described by SVO sentences than to the fencer prior to the second noun phrase. For OVS sentences the fencer as the anticipated agent should receive more inspections than the pirate before the second noun phrase is encountered.

In addition to the 24 items, a set of 32 filler items was constructed. Each filler item consisted of a scene and a sentence accompanying it. Eight started with an adverbial phrase and images showed two characters with only one being involved in an action; eight started with an unambiguously case-marked noun phrase and had four characters, with

Table 3.1: *Example of sentence pairs for images in Figure 3.1*

Image	Condition		Sentences
3.1:a	SVO	(1a)	Die Prinzessin wäscht offensichtlich den Pirat. The princess (subject/object) washes apparently the pirate (object). 'The princess is apparently washing the pirate.'
3.1:a	OVS	(1b)	Die Prinzessin malt offensichtlich der Fechter. The princess (subject/object) paints apparently the fencer (subject). 'The princess is apparently painted by the fencer.
3.1:b	SVO	(2a)	Die Prinzessin malt offensichtlich den Fechter. The princess (subject/object) paints apparently the fencer (object). 'The princess is apparently painting the fencer.'
3.1:b	OVS	(2b)	Die Prinzessin wäscht offensichtlich der Pirat. The princess (subject/object) washes apparently the pirate (subject). 'The princess is apparently washed by the pirate.'

two being involved in an action; eight had scenes that did not contain action depictions, with verb-final sentences describing non-stereotypical actions in the future tense; eight had an unambiguously case-marked initial noun phrase followed by a second coordinated noun phrase with images showing three characters, of which two were involved in an action.

The fillers ensured that the verb was not always in the second position; that there was not always an ambiguous noun phrase in the first position; that images did not always display a depicted action, and that there were not always three characters in the scene. Experimental items were separated from one another by at least one filler item. There were four lists of stimuli, each consisting of 24 experiment and 32 filler items. Each participant saw only one of the four conditions of each item, and the same number of items in each condition, and the order of items was randomized individually for every participant.

PROCEDURE

An SMI EyeLink head-mounted eye-tracker with a sampling at a rate of 250 Hz monitored participants' eye-movements. Images were presented on a 21-inch multi-scan color monitor at a resolution of 1024 x 768 pixels concurrently with the spoken sentences. Although only the dominant eye of each participant was tracked, viewing was binocular. Participants' head movements were unrestricted. Before the experiment, participants were instructed to listen to the sentences and to inspect the images, and to try to understand both sentences and depicted scenes. There was no other task. They were shown two example images and sentences. Next, participants were set up and calibrated manually using a nine-point fixation stimulus. This procedure was repeated after approximately half of the trials. The EyeLink software validated calibration; if validation was poor, the calibration procedure was repeated until validation was good. Three filler items preceded the first experimental item for each participant. Between the individual trials, participants saw a centrally-located fixation dot on the screen, which they were asked to fixate. This allowed the eye-tracking software to perform a drift correction if necessary. The entire experiment lasted approximately 30 min.

ANALYSIS

The procedure for analyzing the output data of the eye-tracker was as follows: The visual scenes were color-coded into distinct regions on bitmap templates (1024 x 786 pixels). The coded regions were the background and each of the three characters (e.g., for images in Fig. 3.1, the princess, the pirate, the fencer, and the background). The eye-tracker software recorded the X-Y co-ordinates of participants' fixations. This output was then converted into distinct codes for the characters and background so that participants' fixations were mapped onto the objects of an image.

Characters were coded depending on their event role for the inferential and descriptive analyses ('amb. char., agent, and patient, see Figs 3.2, 3.3, 3.4, and 3.5). Consecutive fixations within one object region (i.e., before a saccade to another region occurred) were added together, being counted as one 'inspection' (also termed 'gaze'). Contiguous fixations of less than 80 ms were pooled and incorporated into larger fixations. Blinks and out-of-range fixations were added to previous fixations. We report the mean proportions of fixations on scene entities over the course of the entire utterance (Fig. 3.2), as well as inferential analyses and descriptive data of the proportion of inspections for individual

time regions (Figs 3.3 to 3.5). We report proportions of inspections, a measure for which previous studies on auditory sentence comprehension in visual environments have established a close link to on-line comprehension processes (e.g., Tanenhaus et al., 1995; Sedivy et al., 1999).

For the descriptive presentation of the time course of the eye-movement data (Fig. 3.2), the process was a follows: For each time slot and object, the program computed the number of inspections that fell within a given time slot. For example, if an inspection on an object started at 1000 ms and lasted until 1125 ms after display onset, and time slots were 50 ms, then the program scored one inspection on that object for the 1000-1050 ms time slot, one inspection for the 1050-1100 ms time slot, and finally one inspection for the 1100-1150 ms time slot (i.e., the end of the inspection fell still within the 1100-1150 ms slot). For the subsequent slot from 1150-1200 ms the program would score zero inspections unless a new inspection started within that time slot. Figure 3.2 plots the mean of the proportion of inspection counts per time slot, separately for each sentence condition and role of character. The word onsets marked on the graphs represent the average of word onsets for the individual item trials.

In contrast to this, the data plotted for the individual time regions (Figs 3.3 to 3.5) are based on exact trial-by-trial computations of these regions. Word onsets in each item speech file had been marked for the first noun phrase, the verb, the post-verbal adverb, and the second noun phrase. We computed the proportion of cases in a sentence condition for which there were inspections that started within a time-region (see Figs 3.3 to 3.5). This calculation type differs from the computation of inspection counts for the time curve presentation, where an inspection was counted when it fell within a given slot. As a result, there may in some cases be slight differences between the two types of data presentation. When an inspection started in a time slot and lasted into the following slot, it would be counted as an inspection during both periods in the time curve presentation (Fig. 3.2). For the individual word region graphs (Figs 3.3 to 3.5), it would only be counted for the region in which it started.

The main time region for the analysis was the post-verbal adverbial region (ADV). As linguistic and visual information had to be integrated, we expected the disambiguation effect to occur shortly after the verb. The ADV region stretched from adverb onset to the onset of the second noun phrase (e.g., apparently, see Table 3.1) and had a mean duration of 794 ms for the SVO and 810 ms for the OVS condition. The second noun phrase with its case-marking provided corroboratory evidence for disambiguation patterns on ADV and was thus included in the analysis (NP2). The NP2 interval extended from the onset of the case-marked article of the second noun phrase until the offset of the noun. Its mean duration was 831 ms for the SVO and 815 ms for the OVS condition. In addition we also examined inspections during the verb region (VERB), so as to check for potential early

disambiguation effects on the verb (e.g., Altmann & Kamide, 1999). The mean duration of the verb was 491 ms for the SVO and 489 ms for the OVS condition.

For the analysis of inspection proportions within a time region, we used hierarchical log-linear models, which combine characteristics of a standard cross-tabulation chi-square test with those of ANOVA. Log-linear models are adequate for count variables because they neither rely upon parametric assumptions concerning the dependent variable (e.g., homogeneity of variance), nor require linear independence of factor levels (Howell, 2002). Inspection counts for a time region were adjusted to factor combinations of target character (patient, agent), sentence condition (SVO, OVS) and either participants (N = 28) or items (N = 24). We report effects for the analysis with participants as $LR\chi^2$(subj) and for the analysis including items as a factor as $LR\chi^2$(item).

3.1.2 RESULTS AND DISCUSSION

Figure 3.2 gives a detailed impression of the eye-movements over time by showing the mean proportion of inspections to characters during presentation of trials in time frames of 250 ms. The first graph of Figure 3.2 shows the eye-movements to the patient and agent on the images for SVO and OVS sentences from the onset of the first noun phrase (mean onset 1019 ms). The second graph displays inspections to the ambiguous character for both sentence types. Fixations on the background were left out for clarity of presentation, and inspections to the ambiguous character and the two target characters (agent, patient) are presented in separate graphs. The mean proportion of inspections to all entities (ambiguous, agent, patient, background) per sentence condition (SVO, OVS) add up to one hundred per cent.

Figure 3.2 suggests the following: Just before the onset of the verb, when participants have heard 'The princess', they predominantly inspect the ambiguous character (the princess) (see the second graph of Fig 3.2). At the same time, they look equally often at the agent (the fencer) and patient (the pirate), but much less often than at the princess as shown in the first graph (Fig 3.2). The high amount of inspections to the ambiguous character immediately before and while hearing the verb ('washes'/'paints') presumably simply reflects the fact that it has just been mentioned as first noun phrase. While people hear the verb, the pattern of eye-movements to the patient (the pirate) and agent (the fencer) changes. The first graph (Fig. 3.2) shows that shortly after the onset of the verb ('washes'/'paints'), participants start to inspect the patient (the pirate) more

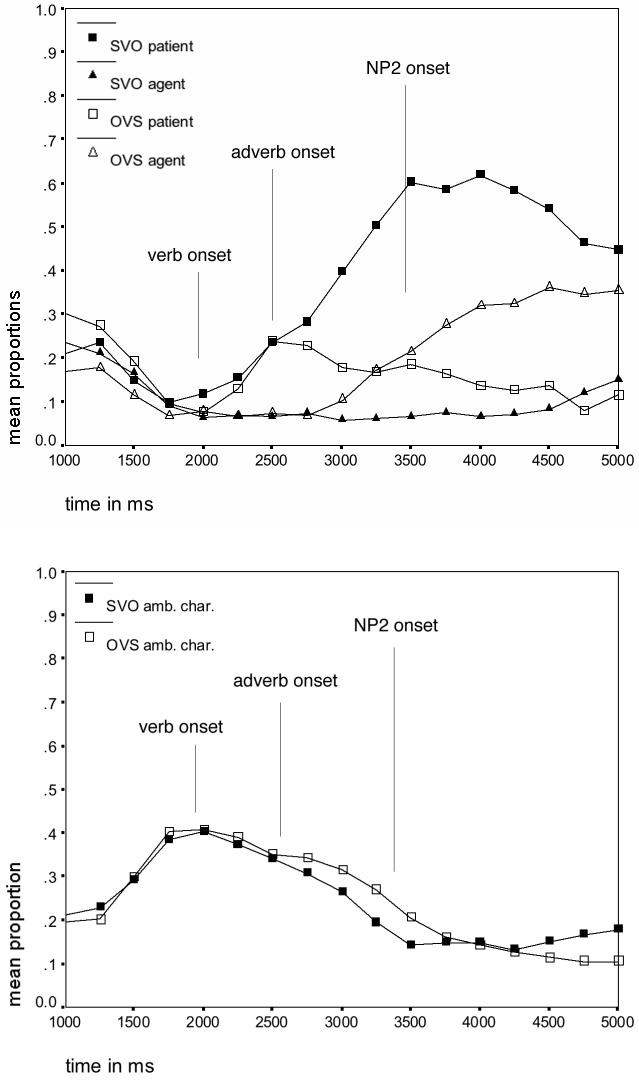


Figure 3.2: *Time course of the eye-movement data for Experiment 1 showing the mean proportion of inspections to characters from the onset of the spoken stimuli in time frames of 250 ms*

often than at the agent (the fencer) for SVO and OVS sentences. This may be due to the interpretation of the sentence fragment, treating the first noun phrase as the agent and starting to look for the patient, but it may also be due to visual factors such as the gaze direction and posture of the ambiguous character towards the patient. Then, the patterns diverge at about the onset of the adverb. While participants who have heard SVO sentences ('The princess washes') continue to look at the patient of the washing-event (the pirate), participants who have heard OVS sentences ('The princess paints') start to look more at the agent of the painting event (the fencer). At the same time inspections to the patient (the pirate) decrease in the OVS condition (first graph of Fig. 3.2). Whereas the initial looks to the patient might be due to the orientation of the ambiguous character (the princess), this cannot be the case once the sentence type has an effect. At this point, we can be certain that the pattern of anticipatory looks reflects processes of identifying the verb with the appropriate depicted action.

During the second noun phrase, which provides another point of disambiguation at its case-marked article, the ambiguous character is looked at less often (see second graph of Fig. 3.2), as its mention occurred some time ago. The fixations on the agent have further increased when people heard 'The princess paints apparently the fencer' and exceed inspections to the patient (the pirate) for OVS sentences as shown in the first graph (Fig. 3.2). The continuation of this fixation pattern during the second noun phrase provides clear evidence for disambiguation processes.

Participants are more likely to fixate the ambiguous character during the adverb time interval for OVS than SVO sentences. This coincides with the time that listeners are switching from fixating the patient (the pirate) to fixating the agent (the fencer). It may be due to participants who have not so far fixated the patient reprogramming their movement to fixate the agent, or it may be due to failed movements from patient to agent (participants reading SVO sentences are simply likely to continue inspecting the patient). Alternatively, it may be due to deliberate re-fixations on the ambiguous character in order to recompute its role from agent to patient. The main finding, however, is the evidence for tremendously rapid looks to the appropriate role filler for SVO and OVS sentences once the verb had been encountered.

We now turn to the inferential analysis. Crucially, the conclusions emerging from the descriptive analysis above hold for the inferential analysis of the data. While the time curves present an overview of the eye-movements to all scene entities, the inferential analyses for each experiment focus only on the theoretically most relevant effects.

The proportion of inspections to all entities (ambiguous, agent, patient, background) per sentence condition (SVO, OVS) add up to one hundred per cent (Figs 3.3 to 3.5).

For the VERB region, log-linear analyses revealed a significant main effect of target character (patient, agent) ($LR\chi^2(subj) = 18.55, df = 1, p < 0.0001; LR\chi^2(item) = 19.21, df = 1, p < 0.001$).There was no interaction of sentence condition (SVO, OVS) and target character (agent, patient) (both $LR\chi^2 s < 1$). The main effect was due to a higher proportion of fixations on the patient than on the agent (see Fig. 3.3).

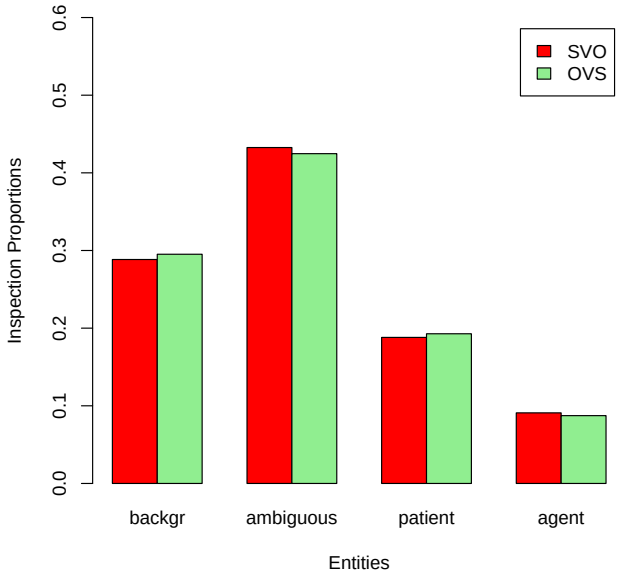


Figure 3.3: *Percentage of inspections to entities for the VERB region in Experiment 1*

Figure 3.4 shows the proportion of inspections to the target characters and background during the ADV time interval for both sentence conditions (SVO, OVS). The interaction between sentence condition (SVO, OVS) and target character (agent, patient) was significant ($LR\chi^2(subj) = 82.30, df = 1, p < 0.0001; LR\chi^2(item) = 83.81, df = 1, p < 0.0001$). Log-linear contrasts confirmed that this was due to a significantly higher percentage of agent inspections in the OVS condition than in the SVO condition ($LR\chi^2(subj) = 39.95, df = 1, p < 0.0001; LR\chi^2(item) = 40.23, df = 1, p < 0.0001$), and to a significantly

higher percentage of inspections to the patient in the SVO condition than in the OVS condition ($LR\chi^2(subj) = 68.15, df = 1, p < 0.0001; LR\chi^2(item) = 67.86, df = 1, p < 0.0001$). For the ADV time interval, contrasts showed in addition that there were a significantly higher percentage of inspections to the ambiguous character for the OVS condition than for the SVO condition ($LR\chi^2(subj) = 4.96, df = 1, p < 0.03; LR\chi^2(item) = 4.95, df = 1, p < 0.03$).

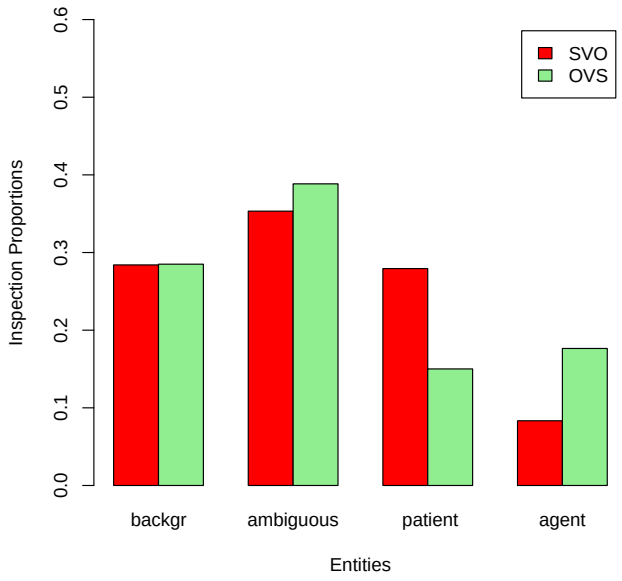


Figure 3.4: *Percentage of inspections to entities for the ADV region in Experiment 1.*

During the NP2 time interval, log-linear analyses revealed a significant interaction of sentence condition (SVO, OVS) and target character (agent, patient) ($LR\chi^2(subj) = 123.64, df = 1, p < 0.0001; LR\chi^2(item) = 125.50, df = 1, p < 0.0001$). The interaction was qualified by a higher proportion of inspections to the patient for SVO than for OVS sentences ($LR\chi^2(subj) = 66.67, df = 1, p < 0.0001; LR\chi^2(item) = 67.43, df = 1, p < 0.0001$), and by a higher proportion of inspections to the agent for OVS than for SVO sentences ($LR\chi^2(subj) = 105.20, df = 1, p < 0.0001; LR\chi^2(item) = 106.67, df = 1, p < 0.0001$) (see Fig. 3.5).

The fixation pattern prior to the second noun phrase could be due to either thematic role assignment processes or a visual search strategy. A visual search account would argue that once people have heard the verb, they check whether it identifies the action performed by the first-mentioned entity (the princess). For SVO sentences, they hear 'washes', and this matches the princess-washing event, and hence they direct their attention to the character the princess faces (the pirate, Fig. 3.1 a). For OVS sentences, they hear the verb 'paints', realize that the princess is not painting, start searching for an appropriate painting-action, and find the fencer-painting event. This account would moreover explain the rapidity with which people locate the appropriate depicted agent. It does not require that the fixations on the fencer for OVS sentences be guided by scene-derived expectations of thematic role-relations.

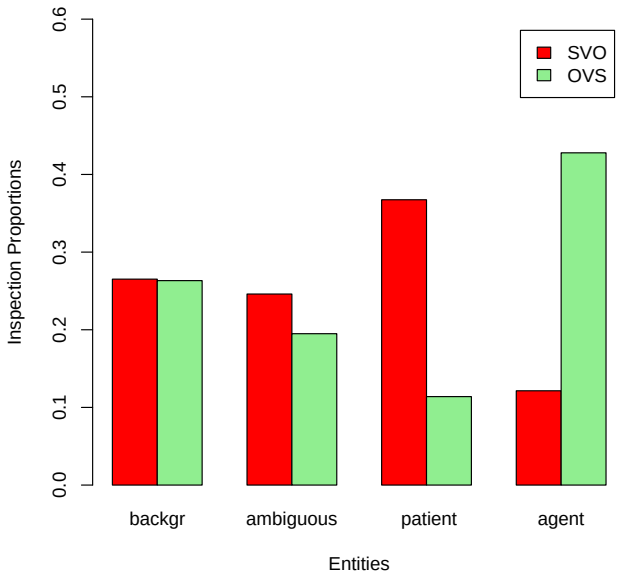


Figure 3.5: *Percentage of inspections to entities for the NP2 region in Experiment 1.*

Alternatively, the observed eye-movements may reflect anticipation of thematic role-fillers, as argued in previous studies which did not explore the effects of depicted events but of case-marking and verb-based world knowledge on thematic role assignment (e.g.,

Kamide, Scheepers, & Altmann, 2003). Such role assignment account of our data would assume that people have extracted information about thematic-role relations from the event scenes prior to hearing the verb. In the SVO condition, when people hear 'The princess washes', they assign an agent role to the first noun phrase, as the princess is the agent of the verb ('wash'). They anticipate a patient role-filler and hence look predominantly at the patient of the washing action (the pirate) rather than the agent (the fencer). For OVS sentences, they hear 'The princess paints'. The assignment of an agent role to the first noun phrase is disconfirmed, as the princess is not doing the painting action. Previously acquired scene information about the fencer as the agent of a painting action, allows anticipating the fencer as the agent, and assigning a patient-role to the princess.

In order to dissociate between these two, possibly complementary, accounts, we consider research on visual and linguistic processing. Findings by Zelinsky and Murphy (2000) support the claim that linguistic properties of an object's name only influenced gaze behavior (frequency and duration of fixations) during object inspection when linguistic processing of objects was encouraged, but not when the task was visual search. Similarly, experiments on visual search and linguistic comprehension in text processing by Rayner and Raney (1996) found that word frequency only affected gaze duration when the task was reading, not during visual search. Accurate interpretation of gaze behavior during text and scene inspection hence necessitates consideration of the instructions given to participants. While the task in the present studies (scene and sentence comprehension) differs from those in the experiments by Zelinsky and Murphy (2000) (short term object recognition vs. visual search), an important commonality is that both scene/sentence comprehension and short-term object recognition encourage linguistic processing of the scene.

The fact that our instruction (understanding scene and sentences) encouraged linguistic processing thus argues for a thematic role assignment rather than visual search account of our data. This line of reasoning is supported by results from a study that was carried out subsequent to the present experiments and with our materials (Scheepers & Crocker, 2004). Their experiment replicated our findings of early visual disambiguation using inspection duration as measure, and further showed that the structure of a sentence presented prior to our stimuli had a priming influence on inspection duration for scene entities. Such priming presumably reflects linguistic expectations based on the preceding sentence structure, and further supports a contribution of thematic role assignment in accounting for our data. Additional research is, however, necessary for a full understanding of how eye-movements in visual scenes depend upon the task at hand.

While a visual search strategy alone does not account for our data, it may nonetheless contribute to the observed gaze pattern. If so, this indicates a tight coordination of visual and linguistic processing, which ensures efficient scene perception and hence enables a most rapid impact of the perceived events on understanding (see sections 1.3, 2.5). Expectation of thematic roles derived from scene inspection prior to hearing the verb may under this view combine with visual search as a rapid mechanism for locating scene objects that correspond to anticipated role-fillers.

3.2 EXPERIMENT 2

It is conceivable that some of the strength of our effects in Experiment 1 is due to a lack of distractor objects such as have been used in previous studies (e.g., Altmann & Kamide, 1999; Kamide, Altmann, & Haywood, 2003; Sedivy et al., 1999; Tanenhaus et al., 1995). After having encountered the ambiguous character, participants just had to choose between two possible options looking at the patient or the agent character), since there were no alternative objects to look at. Further, while partial cross-splicing in Experiment 1 excluded effects of the OVS intonation on disambiguation, we wanted to explore more fully the possible influence of intonation cues on the fixation patterns we observed by carrying out a fully cross splicing on the spoken materials.

3.2.1 METHOD

PARTICIPANTS

Forty further participants from the same population as in Experiment 1 were paid five euro for taking part in the experiment.

MATERIALS

The design, procedure and analysis were the same as for Experiment 1. Two distractor objects were added on each image (see Fig. 4.1). Distractors were always inanimate, and were added to provide some further possibilities for people to look at. All filler items were modified to include distractor objects. While experimental sentences did not refer to the distractors, 12 of the 32 filler images had sentences that referred to them. Linguistic materials differed only in that we included the original version (unchanged intonation) and a cross-spliced version for each item sentence. For OVS sentences (e.g., Table 3.1, (1b)),

the sentence beginning of the SVO sentence (e.g., Table 3.1, (2a)) was spliced in before the second noun phrase. For SVO sentences (e.g., Table 3.1 (2a)), the sentence beginning of the OVS sentence (e.g, Table 3.1, (1b)) was spliced in before the second noun phrase.

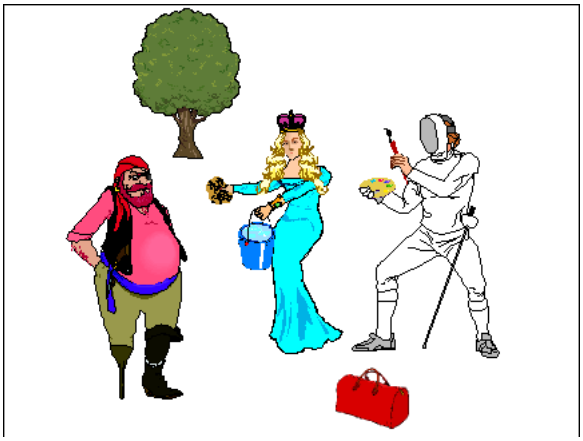


Figure 3.6: *Example image of Experiment 2 for sentences in 3.1*

One item set thus consisted of eight sentences (four original, four cross-spliced) and two images. The cross-spliced versions for SVO and OVS sentences had an OVS and SVO intonation respectively up to the second noun phrase. The full cross-splicing allowed us to examine the influence of intonation cues on all the observed effects in a more complete way than in Experiment 1 where all sentence had SVO intonation up to the second noun phrase. Intonation (original, cross-spliced) was included as a factor in the inferential analyses.

The same word regions were chosen for analysis as in Experiment 1. The mean duration of the VERB interval was 490 ms for SVO, and 488 ms for OVS sentences. The mean duration of the ADV interval was 789 ms for SVO, and 798 ms for OVS sentences. The mean duration of the NP2 interval was 676 ms for SVO and 668 OVS sentences.

3.2.2 RESULTS AND DISCUSSION

Figure 3.7 displays the time course of eye-movements during trials. The first graph shows inspections to patient/agent, and the second graph displays the time course of inspections to the ambiguous characters.

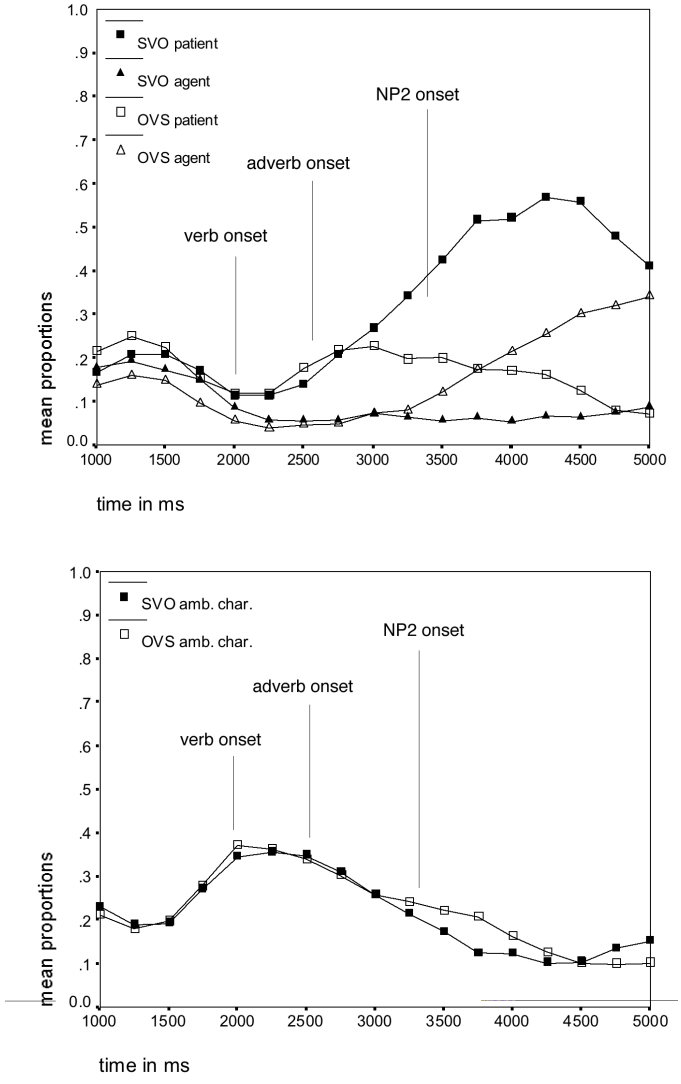


Figure 3.7: *Time course of the eye-movement data for Experiment 2 showing the mean proportion of inspections to characters from the onset of the spoken stimuli in time frames of 250 ms*

Overall, the picture that emerges is similar to the one for Experiment 1. Shortly after the onset of the verb ('washes'/ 'paints'), more inspections occur to the patient (the pirate) than to the agent (the fencer) as a result of either linguistic expectations or visual factors (see Table 3.1 (1a), (1b) and Fig. 4.1). After the onset of the adverb, the fixation patterns diverge. For SVO sentences, participants continue to look at the patient, whereas in the OVS condition they start to look to the agent. The pattern of eye-movements during the second noun phrase after participants had heard the case-marked article was the same as for the adverb time interval, and provided further evidence for disambiguation processes.

We now turn to the inferential analyses. There was no effect of intonation (original, cross-spliced) for the analysis regions (all $LR\chi^2 s < 1$). The distractor objects were hardly looked at and had no visible influence on the eye-movement patterns during any of the time windows (see Figs 3.8, 3.9, and 3.10).

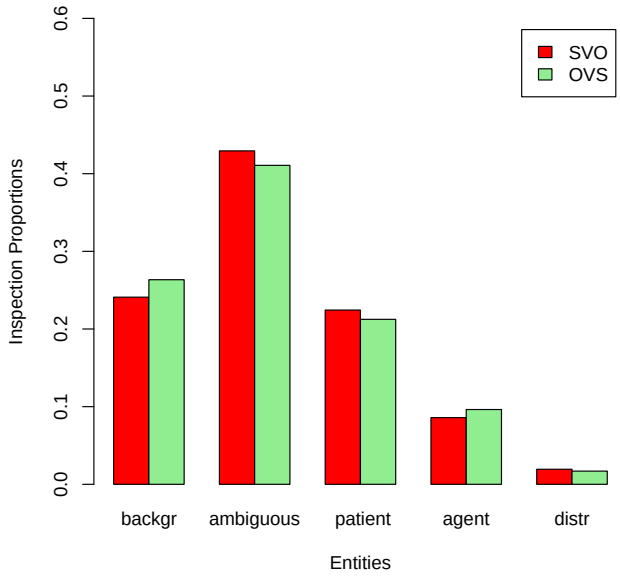


Figure 3.8: *Percentage of inspections to entities for the VERB region in Experiment 2*

For the VERB region, log-linear analyses revealed a significant main effect of target character (patient, agent) ($LR\chi^2(subj) = 27.07, df = 1, p < 0.0001; LR\chi^2(item) =$

29.46, $df = 1, p < 0.0001$) in the absence of a significant interaction (both $LR\chi^2 s < 1$). The main effect was due to a higher proportion of fixations on the patient than on the agent for both sentence conditions (Fig. 3.8).

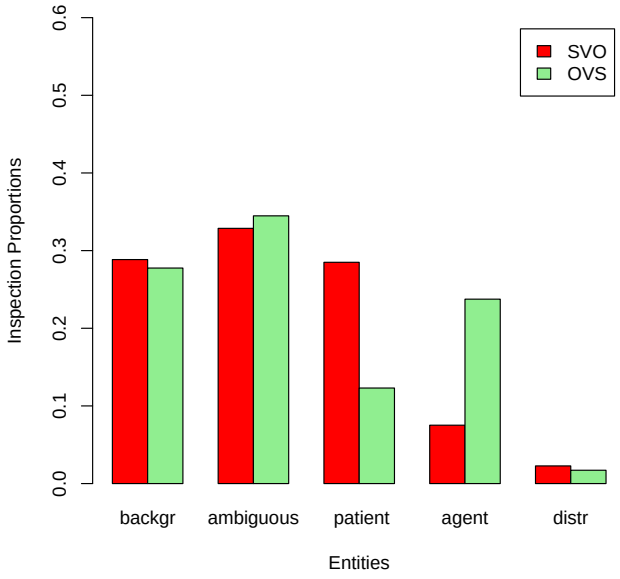


Figure 3.9: *Percentage of inspections to entities for the ADV region in Experiment 2*

Figure 3.9 shows the proportions of inspections to the characters, distractor objects, and the background during the ADV time interval. The interaction between percentage of inspections to entities (agent, patient) and sentence condition (SVO, OVS) was significant ($LR\chi^2(subj) = 106.95, df = 1, p < 0.0001; LR\chi^2(item) = 111.31, df = 1, p < 0.0001$) for this region. As for Experiment 1, log-linear contrasts revealed a significantly higher percentage of inspections to the agent for the OVS condition than for the SVO condition ($LR\chi^2(subj) = 69.12, df = 1, p < 0.0001; LR\chi^2(item) = 69.69, df = 1, p < 0.0001$), and a significantly higher percentage of inspections to the patient in the SVO condition than in the OVS condition shortly after the verb ($LR\chi^2(subj) = 66.94, df = 1, p < 0.0001; LR\chi^2(item) = 65.67, df = 1, p < 0.0001$).

It is unlikely that the delay was due to looks to the distractor objects, which are inspected very little during the adverb time interval (see Fig. 3.9). We assume that it is rather due to participants having to choose one out of five entities rather than one out of three as in Experiment 1.

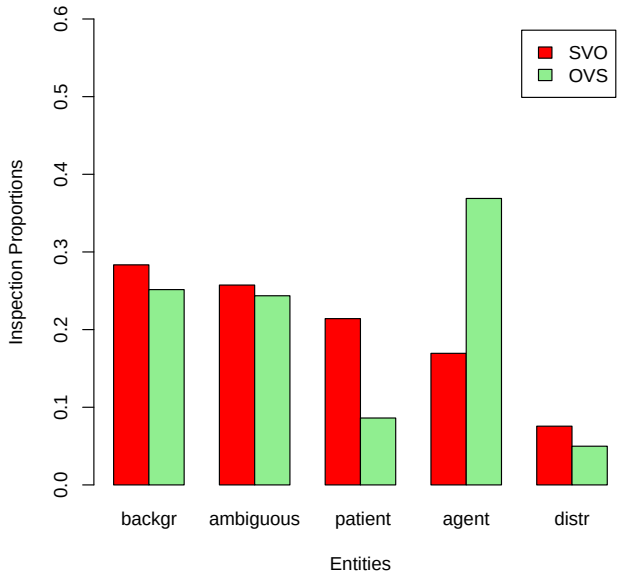


Figure 3.10: *Percentage of inspections to entities for the NP2 region in Experiment 2*

During the NP2 time region, there was a significant interaction of sentence type (SVO, OVS) and target character (agent, patient) ($LR\chi^2(subj) = 176.63, df = 1, p < 0.0001; LR\chi^2(item) = 185.16, df = 1, p < 0.0001$). Contrasts again confirmed a higher proportion of looks to the patient for SVO than for OVS sentences ($LR\chi^2(subj) = 114.27, df = 1, p < 0.0001; LR\chi^2(item) = 117.61, df = 1, p < 0.0001$) and a higher proportion of inspections to the agent for OVS than for SVO sentences ($LR\chi^2(subj) = 126.14, df = 1, p < 0.0001); LR\chi^2(item) = 132.53, df = 1, p < 0.0001$) (see Fig. 3.10).

SUMMARY

In sum, Experiments 1 and 2 provided strong evidence for early verb-mediated visual disambiguation of local structural and role ambiguity in German SVO/OVS sentences. Effects proved robust in the presence of distractor objects and a full cross-splicing. Our findings show that perceived event relations between entities in the immediate visual context allow the rapid recovery of mental representations such as thematic roles in on-line comprehension of spoken sentences, and actively influence incremental resolution of initial structural and role ambiguity in the linguistic input.

This was revealed in Experiments 1 and 2, where anticipatory eye-movements in the event scenes provided evidence for expectations of a patient and agent role filler for initially ambiguous SVO and OVS sentences respectively once the verb had identified the depicted actions. Experiment 2 replicated the early visual disambiguation effect in the presence of distractor objects and investigated more fully the possible influence of intonation. The rapid verb-mediated accessing of the depicted actions suggests further a tight coupling of visual and linguistic processing where a verb-initiated visual search cooperates in tight coordination with scene-derived expectations of thematic role-fillers in incremental interpretation of scene and sentence.

Chapter 4

The Active Influence of Scene Events

In chapter 3, two experiments investigated the verb-mediated influence of depicted events on incremental thematic role assignment and structural disambiguation. We found that as soon as people had heard the verb in the utterance they were able to use it in order to identify depicted actions and their associated thematic relations that were then available for early incremental thematic role assignment. The experiments in this chapter were designed to see whether findings from Experiments 1 and 2 generalize and extend to other sentence constructions and to another language (English). They further examined whether depicted events can enable structural disambiguation and incremental thematic role assignment even when other linguistic cues in the utterance (not the verb) had to be used to identify the depicted events.

Specifically, Experiment 3[1] explored whether soft linguistic cues can make available depicted scene events for structural disambiguation of initially structurally ambiguous German sentences when the verb appeared in sentence-final position. Experiment 4 examined the on-line comprehension of initially structurally ambiguous main clause/reduced relative sentences. It had two key goals. First, it investigated whether the verb might just as in Experiments 1 and 2 on German make available the depicted actions and their associated thematic relations for early structural disambiguation and incremental the-

[1]Experiment 3 in the present chapter was previously published as part of the journal article "The influence of the immediate visual context on incremental thematic role assignment: evidence from eye-tracking in depicted events", see Knoeferle et al. (2005)

matic role assignment. Second, it further explored the research questions of Experiment 3 and examined whether in the case of the English main verb/ reduced relative ambiguity soft linguistic cues could also make the depicted events available for early thematic role assignment even before the lexical verb identified them in the scene. We first focus on Experiment 3 and return to Experiment 4 in section 4.2.

Experiment 3 examined the on-line comprehension of initially structurally ambiguous German active/passive sentences. Consider the example sentence 4.1 from the experiment. Images were the same as in Experiment 1. An example image is repeated here for convenience (Fig. 4.1):

(4.1) Die Prinzessin wird soeben von dem Fechter gemalt.

'The princess (ag./pat.) is currently by the fencer (ag.) painted'.

Figure 4.1: *Example image for sentences in Table 4.1*

The lexical verb ('painted') appears in sentence-final position and hence does not establish early reference to the depicted events. Before the second argument ('by the fencer'), this sentence is structurally ambiguous between an active and a passive sentence. The structural active/passive ambiguity relies on the auxiliary being ambiguous between a future tense auxiliary (*wird*, active 'will' or passive 'will be') and a present tense auxiliary (passive *wird*, 'is'). Hence, if the auxiliary is interpreted as referring to the future tense,

the initial noun phrase can be agent of an active or patient of a passive sentence; if it refers to the present tense it is the patient of a present tense passive sentence. Clear linguistic disambiguation occurred when morphosyntactic marking on the second argument disambiguated the thematic role relations of the sentence. An initial subject noun phrase is disambiguated as agent in an active sentence when accusative (direct object) case-marking assigns a patient role to the second noun phrase. For passive sentences, preposition and dative case-marking disambiguate the thematic role of the second argument as agent, and the initial subject noun phrase receives the role of a patient (see Table 4.1). We use the term second argument to refer to both the noun phrase for active and the prepositional phrase for passive sentences. In contrast to Experiments 1 and 2, where the ambiguity concerned grammatical function and role, resolving the active/passive ambiguity only involves assignment of the correct thematic role to the first noun phrase, and does not require changing its grammatical function. Observation of linguistic disambiguation in Experiment 3 would thus provide strong evidence that disambiguation patterns reflect thematic role assignment processes.

In order to explore other possibilities of early disambiguation, we introduced a further soft adverbial constraint. The hypothesis was that temporal adverbs bias interpretation of a preceding ambiguous auxiliary towards future or present tense, making the depicted events available for early structural and thematic role disambiguation before the second argument and main verb were encountered. We expected an adverb referring to a point in the future (e.g., sogleich, 'soon') to bias participants' interpretation of the preceding *wird* towards an active future structure. While active or passive structures are possible continuations, people tend to opt for an active analysis in the absence of constraints biasing towards the passive. Present tense adverbs (e.g., soeben 'currently') should bias interpretation of the preceding auxiliary towards a passive auxiliary and hence towards a present tense passive structure. Adverbs were pre-tested for their bias and selected by means of sentence completions.

Twenty participants took part in the pretest. They read sentence beginnings such as *Sogleich wird die Prinzessin ...* ('The princess will soon...') and *Soeben wird die Prinzessin ...* ('The princess is currently ...') on a computer screen. Participants were told to complete the sentence fragments. Sentences were then hand-coded according to whether completions were active or passive. We tested ten pairs of adverbs. For each pair, one adverb referred to the future, and one to the present. From these ten, we selected four future/present tense adverbial pairs that were matched for length and fre-

quency (Baayen et al., 1995). These were *sogleich/soeben* ('soon'/ 'currently'), *umgehend/momentan* ('immediately'/ 'currently'), *unverzüglich/im Augenblick* ('immediately'/ 'right now'), *sofort/derzeit* ('at once'/ 'at the moment'). The future tense adverbs used in the experiment resulted in 74 percent active completions, and 26 percent passive completions. The present tense adverbs resulted in 39 percent active completions, and 61 percent passive completions. This was the strongest passive bias we were able to obtain under the additional constraints of length and frequency matching.

Based on results of Experiments 1 and 2, we predicted anticipatory inspections to the patient for both active and passive sentences on the auxiliary region and the ensuing adverb. We expected early disambiguation due to the adverbial bias before morphosyntactic marking on the second argument was available for disambiguation, if the manipulation was strong enough. Disambiguation of the initial role-ambiguity should definitely occur after participants heard the case-marked article/preposition of the second argument.

4.1 EXPERIMENT 3

4.1.1 METHOD

PARTICIPANTS

Twenty-six further participants from the same population as Experiment 1 were paid five euro for taking part in the experiment.

MATERIALS

Images were the same as in Experiment 1. Spoken stimuli differed from those in Experiment 1 in two respects. We used a different type of construction (verb-final active and passive sentences) and introduced an additional adverbial manipulation. An example sentence pair for one image version (see Fig. 4.1) is given in Table 4.1. We ensured that intonation did not influence disambiguation prior to the second argument. For the passive condition, all respective item sentences were recorded also with an active structure. The active sentence beginning was then spliced together with the passive continuations.

PROCEDURE

The procedure and analysis were the same as for Experiment 1. The time interval (AUX-ADV) chosen for the analysis of the initial interpretation effect extended from the onset

of the auxiliary (*wird* 'will'/'is') to the offset of the adverb (e.g., sogleich 'soon', soeben 'currently'). The rationale was similar to Experiments 1 and 2, in that anticipatory eye-movements to the patient were expected shortly after the first noun phrase. Given that the auxiliary region was very short, we included the adverbial region in the analysis. The mean duration of the AUXADV time region was 811 ms for the active and 788 ms for the passive condition.

For early disambiguation by means of the temporal adverbs, we expected the effect to rapidly show up in the eye-movement pattern. The reason for this was that the adverbs did not refer to anything in the scene, and should provide a direct link to linguistic biases. We thus expected the effect to start once participants had heard most of the adverb. As inserting another filler word after the adverbs was not possible (it would have resulted in highly unnatural sentences), we ensured there was a short pause (mean duration of 200 ms) between adverb offset and onset of the second argument.

The time region for the inferential analyses corresponded thus to the late adverb (from 200 ms before the offset of the adverb for each item until the onset of the second argument) (POSTADV). Its mean duration was 401 ms for active and 422 ms for passive sentences. In order to ensure that early disambiguation (more looks to the agent as compared to patient) in the passive as opposed to the active condition could not be attributed to the additional 21 ms for the POSTADV region in passive sentences, we subtracted 21 ms from the onset of the second argument for every item in the passive condition. In the active condition the region extended up to the second argument for every item as

Table 4.1: *Example of sentence pairs for images in Figure 4.1 for Experiment 3*

Image	Condition		Sentences
4.1	active	(1a)	Die Prinzessin wird sogleich den Pirat waschen. The princess (ag./pat.) will soon the pirate (pat.) wash. 'The princess will soon wash the pirate.'
4.1	passive	(1b)	Die Prinzessin wird soeben von dem Fechter gemalt. The princess (ag./pat.) is currently by the fencer (ag.) painted. 'The princess is currently painted by the fencer.'

described above. The POSTADV time region captured a time frame in which the adverbial constraints were available to the listener. At the same time, case-marking/preposition of the second argument was not yet available for disambiguation. As all sentences had active intonation up to the second argument, intonation cues could not influence disambiguation processes prior to the second argument. The region of the second argument (SECOND ARGUMENT) extended from the onset of the noun phrase/prepositional phrase to the onset of the sentence-final verb.

4.1.2 RESULTS AND DISCUSSION

Figure 4.2 gives an impression of the time course of the eye-movements. The first graph displays looks to the agent and patient for both conditions, the second one displays inspections to the ambiguous character. The first graph shows that there are more inspections to the patient than to the agent for both active and passive condition during the AUXADV time interval (from a mean auxiliary onset of 2047 ms to a mean adverb offset at 2847 ms). The observed initial fixation pattern could, as in Experiments 1 and 2, be due to either visual factors or linguistic expectations of an active sentence construction as evidenced by past research (e.g., Kamide, Scheepers, & Altmann, 2003; Ferreira, 2003; MacWhinney, Bates, & Kliegl, 1984).

Between the offset of the adverb and the onset of the second argument (mean onset 3060 ms), the patterns begin to diverge. When hearing *Die Prinzessin wird gerade* ('The princess is currently'), participants start to look more at the agent for passive sentences, while looks to the patient start to fall. They look more to the patient for active sentences when hearing *Die Prinzessin wird gleich* ('The princess will soon'), while simultaneously looks to the agent decrease. Disambiguation is clearly visible in the fixation patterns of the first graph during the second argument. The second graph of Figure 4.2 shows furthermore that during the second argument region participants looked more often to the ambiguous character for the active than for the passive condition around the onset of the second argument.

Log-linear analyses revealed a significant main effect of target character (agent, patient) on the AUXADV region ($LR\chi^2(subj) = 6.53, df = 1, p = 0.01; LR\chi^2(item) = 6.60, df = 1, p = 0.01$) in the absence of an interaction with sentence condition (active, passive) (both $LR\chi^2 s < 1$) (see Fig. 4.3). During the POSTADV time interval (see Fig. 4.4), the interaction between sentence condition (active, passive) and target character (agent, patient)

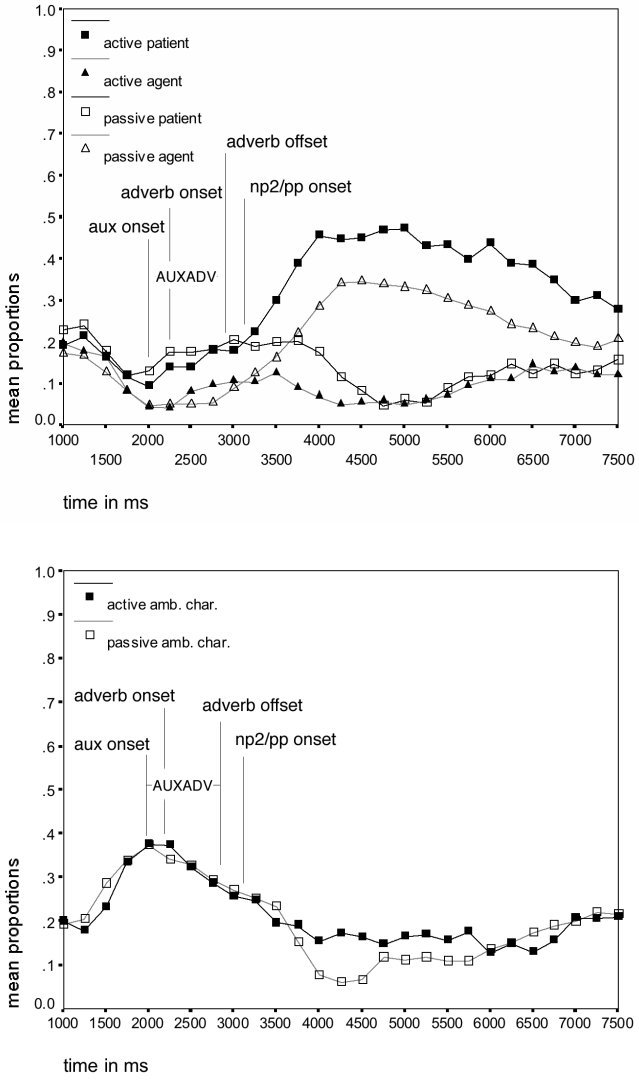


Figure 4.2: *Time course of the eye-movement data for Experiment 3 showing the mean proportion of inspections to entities from the onset of the spoken stimuli in time frames of 250 ms*

was marginal ($LR\chi^2(subj) = 3.10, df = 1, p = 0.08; LR\chi^2(item) = 3.50, df = 1, p = 0.06$). In the log-linear contrasts only the analysis by items revealed a marginal effect of more looks to the patient for active than for passive sentences (($LR\chi^2(subj) < 1; LR\chi^2(item) = 2.71, df = 1, p = 0.05$). For the other contrasts, effects were not significant (all $ps > 0.1$).

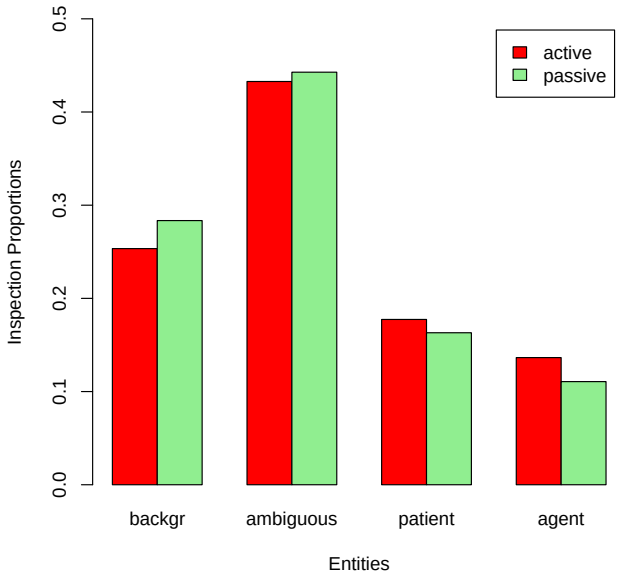


Figure 4.3: *Percentage of inspections to entities for the AUXADV region in Experiment 3*

Interestingly, when only extending the POSTADV time region by 50 ms into the second argument (hence 50 ms into the second argument for active, and 29 ms into the second argument for passive sentences), effects become significant. There was a significant interaction between target character (agent, patient) and sentence condition (active, passive) ($LR\chi^2(subj) = 4.79, df = 1, p < 0.05; LR\chi^2(item) = 4.87, df = 1, p < 0.05$). The contrasts revealed a significant effect of more looks to the patient for active than for passive sentences ($LR\chi^2(subj) = 4.37, df = 1, p < 0.05; LR\chi^2(item) = 5.25, df = 1, p < 0.05$), a non-significant result for the agent inspections by subjects ($LR\chi^2(subj) = 1.16, df = 1, p = 0.28$), and a marginal effect by items ($LR\chi^2(item) = 3.14, df = 1, p = 0.08$).

An explanation of our findings in terms of intonation cues can be excluded as all sen-

tences had active sentence intonation up to the second argument. As it is further estimated that programming of an eye-movement precedes the actual launching of a saccade by 200 ms (e.g., Matin, Shao, & Boff, 1993), we argue that fixations during the first 29 or 50 milliseconds of the second argument cannot be due to the preposition/case marked article of the second argument. Rather, it seems likely that they have been caused by the preceding temporal adverbs. The difference between the marginal log-linear contrasts for the patient inspections as compared to the non-significant result for the agent character might reflect the weakness of the passive bias versus the active bias. We assume that a difference of 20 percent between offline passive versus active completions for the passive bias was not large enough to trigger a significant effect in the contrasts. The interaction, however, provides evidence for early disambiguation.

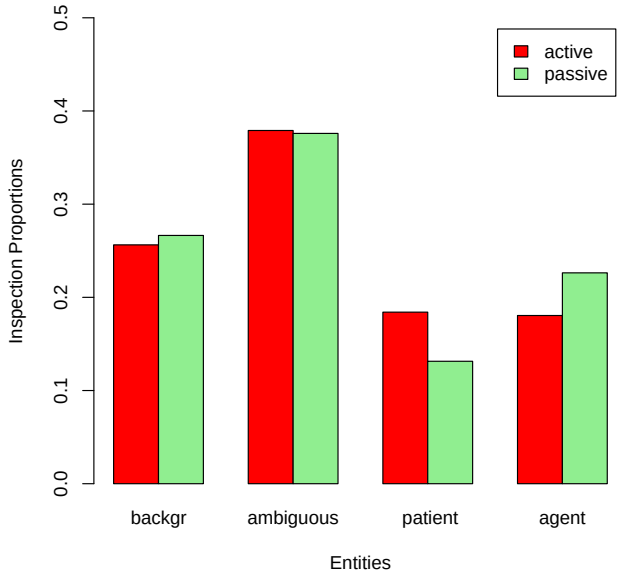


Figure 4.4: *Percentage of inspections to entities for the POSTADV region in Experiment 3*

During the SECOND ARGUMENT region, we found a significant interaction of sentence condition (active, passive) and target character (agent, patient) ($LR\chi^2(subj) = 51.52, df = 1, p < 0.0001; LR\chi^2(item) = 51.89, df = 1, p < 0.0001$). Contrasts confirmed

more inspections to the patient for active than for passive sentences ($LR\chi^2(subj) = 27.11, df = 1, p < 0.0001; LR\chi^2(item) = 27.23, df = 1, p < 0.0001$), and a significantly higher proportion of fixations on the agent for the passive than for the active condition ($LR\chi^2(subj) = 42.83, df = 1, p < 0.0001; LR\chi^2(item) = 43.17, df = 1, p < 0.0001$).

In addition, contrasts revealed significantly more looks to the ambiguous character for active than for passive sentences during this interval ($LR\chi^2(subj) = 5.71, df = 1, p < 0.05; LR\chi^2(item) = 5.71, df = 1, p < 0.05$) (see Fig. 4.5).

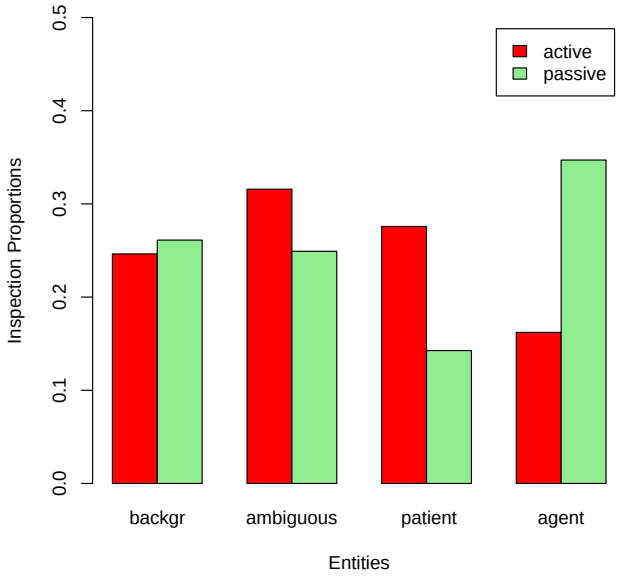


Figure 4.5: *Percentage of inspections to entities for the SECOND ARGUMENT region in Experiment 3*

While participants heard the second argument, eye-movements further confirmed the disambiguation pattern. There were more looks to the patient than agent for the active as compared to the passive condition, and more looks to the agent than patient for passive as compared to active sentences (see Fig. 4.5). This pattern is the same as the one we observed for early visual disambiguation of the SVO/OVS ambiguity in Experiments 1 and 2. As only the thematic role and not the grammatical function of the first noun

phrase is changed in disambiguation of the active/passive ambiguity, we can be certain that fixation patterns in Experiment 3 reflect thematic role assignment processes. There were, in addition, more looks to the ambiguous character for active than for passive constructions during the second argument, an effect that we did not observe for the first two experiments. We assume that it may be due to people anticipating the verb during the second argument, and hence looking at the ambiguous and agent characters' actions (as they were agent of the active and passive sentences respectively, and performing the actions).

Experiment 3 demonstrated that when the verb was sentence-final and only identified the depicted events late during the unfolding utterance, linguistic cues that preceded the verb made available the depicted events for early disambiguation. This was reflected by eye-movement patterns both during ambiguity resolution at the case-marked determiner/preposition of the second argument, and before this point as a result of the influence of temporal adverbs biasing towards the active or passive structure. This finding of early disambiguation from Experiment 3 further supports an account of our data in terms of thematic role assignment, as the depicted event scenes were used for early thematic disambiguation before the verb had referred to the depicted actions.

4.2 EXPERIMENT 4

Experiment 4 was designed to examine whether the findings for German SVO/OVS sentences (Experiments 1 and 2) and for German verb-final sentences (Experiment 3) generalize and extend to another language and sentence type.

To investigate whether findings from Experiments 1 and 2 generalized in the above way we used a sentence construction that is superficially similar to the German SVO/OVS construction. One such sentence structure is the English main verb (MV)/reduced-relative (RR) ambiguity. To illustrate this type of ambiguity, simplified examples from Experiment 4 are presented in sentences 4.2 and 4.3 (for a full version of an item sentence set see Table 4.2).

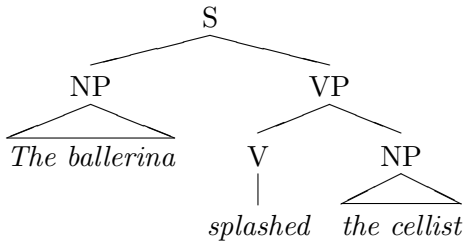
(4.2) The ballerina splashed the cellist.

(4.3) The ballerina sketched by the fencer splashed the cellist.

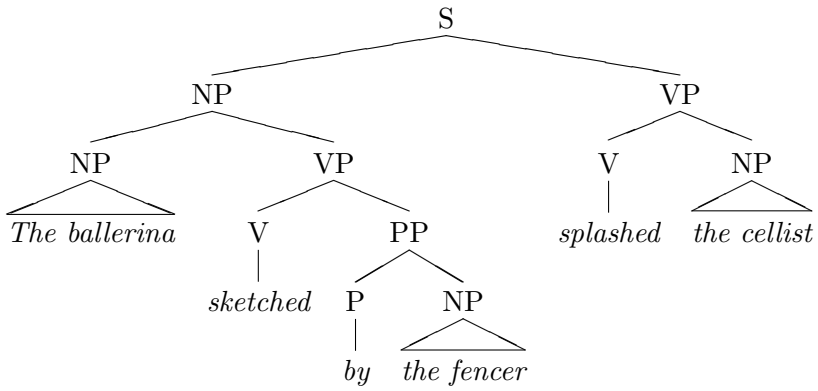
The initial NP-V fragment in these sentences (*The ballerina splashed/ sketched*) is ambiguous up to the second argument in two respects. First, just as was the case for

the SVO/OVS ambiguity, the initial noun phrase is role-ambiguous and can be either the agent or the patient of the currently processed sentence fragment. In sentence 4.2, the initial noun phrase is the agent of the main clause. In sentence 4.3, the initial noun phrase is the patient of the reduced relative clause (*sketched by the fencer*, and the agent of the main clause (*splashed the cellist*). This differs from the German SVO/OVS ambiguity, where the initial noun phrase only receives one thematic role (in the main clause). In addition to the thematic role ambiguity, the sentence fragment is structurally ambiguous. The verb phrase can either be directly attached to the S node as the main predicate of the sentence (main clause analysis, see 4.4), which is the correct analysis for sentence 4.2. Alternatively, the verb phrase can be analysed as a modifier of the initial noun phrase (reduced relative clause 4.5) which is the correct analysis for sentence 4.3.

(4.4)



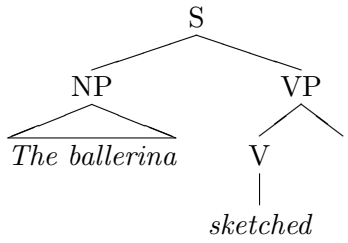
(4.5)



Provided factors such as plausibility, thematic fit or verb type do not bias towards one or the other construction during on-line comprehension (e.g., McRae et al., 1998), people preferentially adopt a main clause structure (Bever, 1970), and linguistic disambiguation only occurs late through the determiner/preposition of the second argument. This means that when people incrementally process a sentence beginning such as the one in example

4.3 (*The ballerina sketched*) and provided there are no other linguistic biases that disambiguate early, the comprehension system initially builds a main clause structure (see example tree 4.6) and then has to make important structural changes upon encountering the preposition on the second argument.

(4.6)



While prior research has investigated the use of thematic fit (McRae et al., 1998) or frequency (Trueswell, 1996) in the incremental structural disambiguation of the MV/RR ambiguity, we investigated whether direct identification of depicted events through a verb in the utterance could enable early structural disambiguation and thematic role assignment for this ambiguity.

Figure 4.6 shows an example image. The main clause sentence 4.2 describes the ballerina-splashing-cellist event (see Fig. 4.6) whereas the reduced relative clause in example 4.3 describes the fencer-sketching-ballerina event. If people can employ the verb for identification of the relevant depicted action during comprehension of the MV/RR clause, then we expect to see evidence for disambiguation in the eye-movement pattern shortly after people encountered the verb and before disambiguation in the linguistic stimuli through the determiner/preposition of the second argument (see Experiments 1 and 2). Based on findings from Experiments 1 and 2, we expect the following: For the main clause sentences (e.g., example 4.2) we should find more anticipatory eye-movements to the patient of the ballerina-splashing event (the cellist) than to the other agent (the fencer). For reduced relative clause sentences when verb was *sketched*, we expect to find more anticipatory eye-movements to the agent of the sketching action (the fencer) than to the other patient (the cellist).

The above-described MV/RR condition pair (e.g., 4.2 and 4.3) was one part of Experiment 4. In order to replicate findings from Experiment 3, we introduced a further condition pair that was designed to investigate whether the comprehension system could make rapid use of depicted events in structuring the unfolding utterance when only soft linguistic biases identified the depicted events early. To test this, we devised sentences

where the lexical verb that directly identified the events was preceded by either a future auxiliary (*will*) or by *being*, an aspect auxiliary for the passive progressive. Based on findings from Experiment 3, we expected that such fine-grained linguistic cues should, in principle, enable rapid structural disambiguation and incremental thematic role assignment shortly after the auxiliary was encountered or early during the ensuing lexical verb if the bias they establish is strong enough.

4.2.1 METHOD

PARTICIPANTS

Thirty-two participants of the University of Dundee (UK) were paid three pounds for taking part in the experiment.

MATERIALS

A new set of 24 items was constructed. The basic design was the same as in Experiment 1. Just as we had investigated the influence of depicted events on the SVO/OVS ambi-

Figure 4.6: *Example image for sentences in Table 4.2 for Experiment 4*

guity in German, here we explored their influence on the on-line disambiguation of main verb/reduced relative clause sentences. In contrast to McRae et al. (1998) there was no stereotypical or plausibility bias from the linguistic stimuli, nor was there a frequency bias for the ambiguous verb which could be used for disambiguation ($p > 0.3$) (Trueswell, 1996).

For the main verb/reduced relative comparison, the depicted actions (see Fig. 4.6) differentiated between the event described by the main clause sentence (e.g., Table 4.2, (1a)), and the event described by the reduced relative clause sentence (e.g., Table 4.2, (1b)). Whereas for the main clause condition, the ambiguous character (the ballerina, see Fig. 4.6, Table 4.2, (1a)) was the agent of a splashing-action, she was the patient of a fencer-sketching event for the reduced relative clause condition (Table 4.2, (1b)).

Recall the expectations that we outlined above: If depicted events can inform incremental thematic and structural disambiguation when they are identified by the verb, then we should observe the same pattern of eye-movements as in Experiments 1 and 2. This means that shortly after verb offset, we should find anticipatory eye-movements to the patient of the ballerina-splashing action (the cellist) for main clause sentences. For the reduced relative clause condition, the verb in the ambiguous clause referred to the fencer-sketching action. Thus, shortly after verb offset and prior to disambiguation through the determiner/preposition on the second argument, we expect anticipatory eye-movements to the agent of the sketching-action (the fencer).

Table 4.2: *Example of sentence pairs for images in Figure 4.6 for Experiment 4*

Image	Cond.		Sentences
Fig. 4.6	MV	(1a)	The ballerina splashed apparently the cellist in the white shirt.
Fig. 4.6	RR	(1b)	The ballerina sketched apparently by the fencer splashed the cellist.
Fig. 4.6	*will*	(2a)	The ballerina will apparently splash the cellist in the white shirt.
Fig. 4.6	*being*	(2a)	The ballerina being apparently sketched by the fencer splashed the cellist.

In addition, a further condition pair (*being/will* sentences, see Table 4.2) was included to test how actively depicted events can influence incremental thematic role assignment (see Experiment 3). We were interested in whether pre-verbal auxiliaries (*will/being*) could be used to identify an active future (ballerina-splashing) and a passive event (fencer-

sketching) respectively.

A noun phrase - future auxiliary sequence (e.g., *The ballerina will*, see Table 4.3 should lead people to expect an active future event performed by the referent of the initial noun phrase. For an image that offers two scene events (e.g., ballerina-splashing-cellist and fencer-sketching-ballerina), the future auxiliary should bias people to expect mention of the cellist as the patient of a future event. Encountering, in contrast, a noun phrase (*the ballerina*) followed by an auxiliary form (*being*) that indicates the passive progressive aspect of an event should raise expectation of a passive event just as was the case in Experiment 3 for present tense adverbs. In this case, people should expect mention of the agent of an event which the ballerina undergoes (the fencer).

The auxiliaries were pre-tested for the strength of their bias by means of a sample from the BNC. We randomly selected 250 occurrences each of both *will* and *being*. An analysis of their function in context revealed that *will* occurred in active sentences for 85.6 per cent of the sample sentences, in passive constructions for 14 percent and with other meanings in two percent of the samples. *Being* was followed by a past participle in 80 per cent of the occurrences, by an adjective in 20 per cent of the cases, and had other meanings in 0.4 percent of the occurrences. Thus, *will* should bias towards an active sentence structure, and *being* towards a past participle construction.

We examined whether people used such information to access depicted events for incremental thematic role assignment and structural disambiguation of the RR/MV ambiguity. If this was the case, then it would provide further evidence for the active influence of depicted events on comprehension processes. We expected anticipatory eye-movements to the patient and agent for *will* and *being* sentences respectively shortly after the auxiliaries or during the post-adverbial lexical verb (see Table 4.2). Both findings would provide evidence that the auxiliaries make depicted events available for incremental thematic role assignment since disambiguation effects due to the depicted events have so far only been observed during the post-verbal region and not during the verb itself (assuming this replicates also in the present experiment).

An item comprised two images and a set of eight sentences. Images were included in their original and a mirrored version, resulting in sixteen experimental lists. Each of these lists was presented in two different randomizations so that every participant saw the trials in an individually randomized order. Conditions were matched for length and frequency up to and including the second argument (Baayen et al., 1995). An example sentence pair for one image version (see Fig. 4.6) is given in Tables 4.2.

PROCEDURE

The procedure and analysis were the same as for Experiment 1. Tables 4.3 and 4.4 show the utterance regions that were chosen for the statistical analyses.

Table 4.3: *Analyses regions for the MV/RR condition pair in Experiment 4*

Condition		Analysis Regions		
		VERB	ADV	SECOND ARGUMENT
MV	The ballerina	splashed	apparently	the cellist
RR	The ballerina	sketched	apparently	by the fencer

Table 4.4: *Analyses regions for the will/being condition pair in Experiment 4*

Condition		Analysis Regions		
		AUXADV	VERB	SECOND ARGUMENT
will	The ballerina	will apparently	splash	the cellist
being	The ballerina	being apparently	sketched	by the fencer

For the main verb/reduced relative conditions analysis regions were the verb, the adverb, and the second argument (see Table 4.3). Since we expected no effect prior to the verb, we do not present individual analyses for the pre-verbal noun phrase region. The time curves, however, show the eye-movement pattern over the entire course of the utterance. The VERB region extends from 180 ms after verb onset (the earliest point at which verb information could affect eye-movements behaviour, (Matin et al., 1993)) to the onset of the adverb (486 ms for MV and 485 ms for the RR condition). The ADV region stretches from the onset of the adverb to the onset of the second argument (794 ms for the MV and 794 ms for the RR condition). The SECOND ARGUMENT region represents the second noun phrase for main verb, and the prepositional phrase for reduced relative clauses. The mean duration for the second noun phrase in main verb sentences was 690 ms, and this was the region used for analyses. For the prepositional phrase in reduced relative clauses the mean duration was 887 ms. There was thus a considerable length difference between the second argument region for the MV in comparison with the

RR condition. Since the second argument region for the RR condition exceeded the length of the second argument region for the MV condition, a finding such as a higher proportion of inspections to the agent in the RR condition as compared to the MV condition could be explained in terms of the longer analysis region for the RR condition. We adjusted for the inequality in length between the two conditions by subtracting the length difference (197 ms) from the offset of the prepositional phrase for each trial in the reduced relative condition. The mean duration of the prepositional phrase after this subtraction was 690 ms, and this was the region used for analyses in the reduced relative condition ('SECOND ARGUMENT').

For the *will/being* conditions, we chose the auxiliary and adverb, the verb, and the second argument as analysis regions. As the auxiliaries were very short, and the earliest effect could occur on the post-auxiliary adverb region, we collapsed the auxiliary and adverb into one region. The AUXADV region starts 180 ms after the onset of *will/being* and lasts until the onset of the verb in the ambiguous reduced relative clause (874 ms for *will* and 845 ms for *being* sentences). The lexical verb was positioned after the AUXADV region and stretched from the onset of the verb to the onset of the second argument (582 ms for *will* and 622 ms for the *being* condition) ('VERB'). The SECOND ARGUMENT region comprises the second noun phrase for the *will* and the prepositional phrase for the *being* condition. Just as for the main verb/reduced relative condition pair, the prepositional phrase exceeded the noun phrase in length. To adjust for this difference, the length difference between the noun phrase and the prepositional phrase (211 ms) was deducted from the prepositional phrase for all trials. The resulting region prepositional phrase region had 674 ms and matched the duration of the noun phrase (mean of 674 ms).

4.2.2 RESULTS AND DISCUSSION

Figures 4.7 and 4.8 presents an overview of the time course of eye-movements during presentation of the utterance. Graph 4.7a displays inspection proportions to the agent and patient, and Graph 4.7b to the ambiguous character for the main verb/reduced relative conditions. Graph 4.8a and Graph 4.8b show eye-movements to the agent/patient and to the ambiguous character respectively for the *will/being* conditions. Inspections to the background, and the distractor objects were omitted from the time course presentation. All of these are, however, included in the graphs that show the proportions of inspections during the individual analyses regions (see Figs 4.9 to 4.14). In the discussion we will only

comment on the eye-movement pattern up to and including the second argument since the remainder of the sentences was not of interest for the research questions we posed, and was not matched for length and frequency.

For clarity of presentation we split up the two condition pairs (see Tables 4.3 and 4.4) in reporting the data. The ensuing sections first describe the time-course of eye-movements and then report the inferential analyses for each condition pair. We henceforth refer to the simple MV/RR condition pair (Table 4.3) as 'Experiment 4a', and to the *being/will* condition pair (Table 4.4) as 'Experiment 4b'. After an analysis of Experiments 4a and 4b we draw relevant comparisons between these two condition pairs.

EXPERIMENT 4A: SIMPLE MV/RR CLAUSE - TIME CURVES

Prior to verb onset, when people are listening to the first noun phrase, and during the early verb time region, most inspections go to the ambiguous character, which is the referent of the first noun phrase (see Graph 4.7b). While people are hearing the verb and the subsequent adverb, eye-movements to the ambiguous character start to decline since her mention occurred some time ago. In the eye-gaze pattern to the ambiguous character during the later utterance (from 4500 ms), we observe a higher proportion of inspections to the ambiguous character for the RR in comparison with the MV condition. Since item sentences were, however, only matched for length and frequency up to and including the second argument, we cannot further interpret this pattern.

We now focus our attention on Graph 4.7a which displays the gaze-pattern to the patient (the cellist) and to the agent (the fencer) (see Fig. 4.6). The graph shows that there are more inspections to the patient than agent for both the main verb and the reduced relative conditions during the verb. This replicates findings from Experiments 1 to 3, and could just as in these experiments be due to either visual factors or linguistic expectation of - in this case - a main clause structure (e.g., Bever, 1970).

Between the onset of the adverb and the onset of the second argument, fixation patterns for the main verb condition in comparison with the reduced relative condition diverge. When participants heard the verb of a main clause that described the ballerina-splashing action, they looked more at the patient of that action (the cellist) than at the agent of the sketching event (the fencer) shortly after verb offset. In contrast, when hearing a verb that identified the action performed by the agent (fencer-sketching) in the reduced relative condition, they start looking more at the fencer than at the patient of the ballerina-

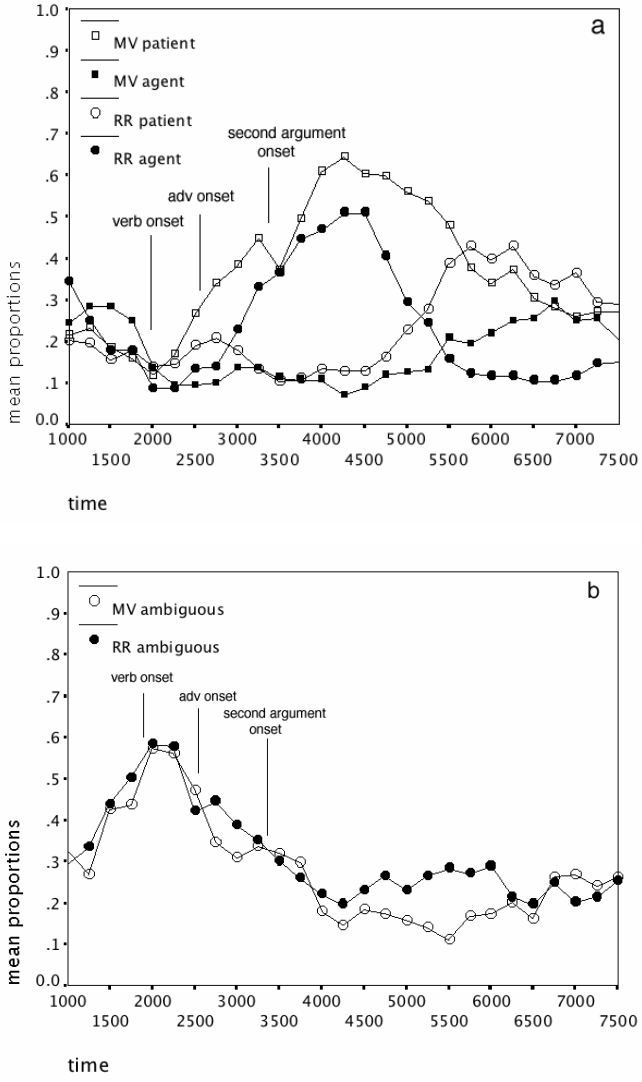


Figure 4.7: *Time course of the eye-movement data for Experiment 4 (MV/RR condition) showing the mean proportion of inspections to entities from the onset of the spoken stimuli in time frames of 250 ms*

splashing action. At the same time their inspection of the patient decreases rapidly for the reduced relative condition.

During the second argument region the clear disambiguation pattern observed during the adverb region continues. For main clause sentences, inspections to the patient and for the reduced relative condition eye-movements to the agent continue to increase. The important conclusion that emerges from the discussion of Graph 4.7a is that disambiguation is clearly visible in the time curve, early during the post-verbal adverb region and subsequently during the second argument region.

EXPERIMENT 4B: 'WILL/BEING' CONDITIONS - TIME CURVES

Graph 4.8b shows that prior to the onset of the verb, most people look at the ambiguous character. There are more eye-movements to the ambiguous character in the *being* condition than in the *will* condition starting with the onset of the auxiliary and continuing throughout the verb and second argument. This descriptive pattern was, however, not confirmed by the inferential analyses for the individual analysis regions.

Graph 4.8a shows that after the onset of the auxiliary, looks to the patient increase for the *will* condition. This could be due to a structural preference for an active sentence structure, due to the bias of the future auxiliary, or both. Crucially, in the *will/being* comparison, such a patient-preference occurs only for the active sentence structure, and is absent for the *being* condition.

This descriptive finding contrasts with results from Experiments 1 to 3, and also with findings from the main verb/reduced relative comparison (Experiment 4a). In all of these experiments, shortly after the offset of the first noun phrase, inspection of the patient increased always for both the favoured and disfavoured sentence type. The absence of such an increase when people heard *being* suggests that *being* was interpreted immediately as biasing towards a passive structure, and prevented early looks to the patient. This pattern was, however, not confirmed by an interaction in the inferential analyses (see below). Rather, the inferential analyses reveal a main effect of target character just as in Experiments 1, 2, and 4a. There was further no strong increase of looks to the agent for the *being* condition in comparison with the *will* condition prior to the onset of the verb (see Fig. 4.8a).

Importantly, however, we find evidence for disambiguation while the verb is encountered in the utterance. With verb onset, looks to the agent for the *being* condition exceed

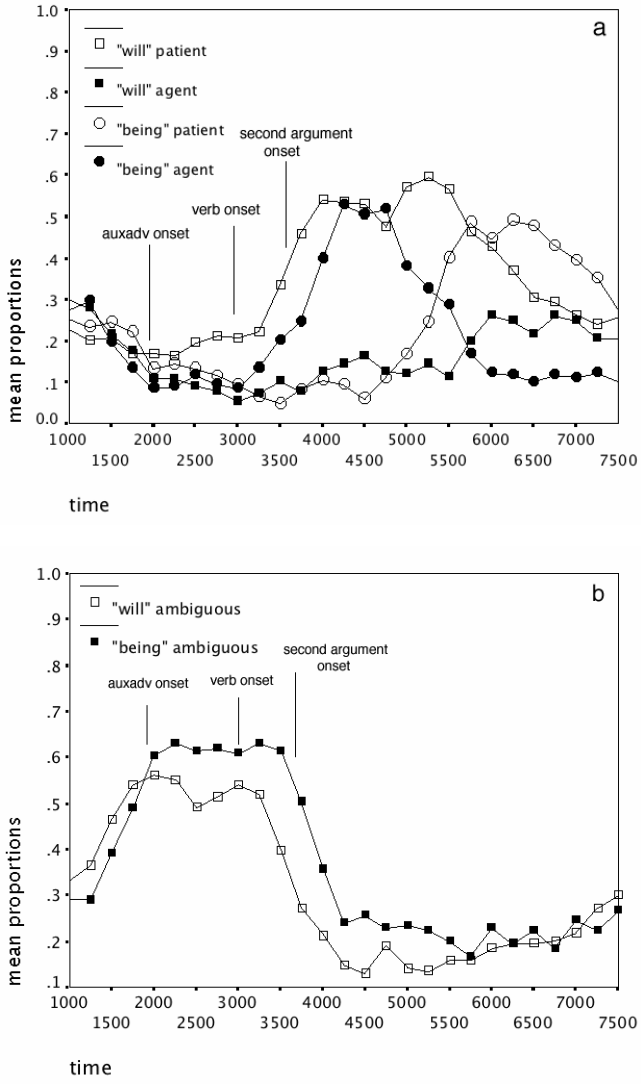


Figure 4.8: *Time course of the eye-movement data for Experiment 4 (will/being condition) showing the mean proportion of inspections to entities from the onset of the spoken stimuli in time frames of 250 ms*

looks to the agent for the *will* condition. For the *will* condition, the increased proportion of inspection to the patient in comparison with inspection of the patient for the *being* condition continues. This pattern of eye-movements provides strong evidence for rapid disambiguation. The rapid use of the depicted events for structural disambiguation and incremental thematic role assignment during the lifetime of the verb show that the linguistic cues had already made accessible the depicted events prior to verb onset. They could then be rapidly used once the first sounds of the verb were encountered. In contrast to this finding, disambiguation through verb-mediated events had only occurred after the offset of the verb for the main verb/reduced relative conditions (see Fig. 4.7a) as well as for Experiments 1 and 2.

Taken together the gaze pattern (i.e., the absence of patient anticipation for *being* sentences, and the disambiguation during the verb itself) provides convincing evidence for early disambiguation through biasing linguistic cues in the utterance that made available the relevant depicted scene events. With the onset of the verb, inspection of the patient continues to increase for the *will* condition where the verb now identifies the action performed upon the patient. For the *being* condition, eye-movements to the agent now clearly exceed looks to the patient for the same condition. This pattern of inspections continues during the second argument.

In sum, the key findings are the eye-movement patterns that reveal a rapid disambiguation effect during the post-verbal adverb region for the MV/RR comparison, and during the verb region for the *will/being* condition.

EXPERIMENT 4A: SIMPLE MV/RR CLAUSE - INFERENTIAL ANALYSES

Let us now turn to the inferential analyses. The conclusions emerging from the descriptive presentation of the data were confirmed by the inferential analyses. While the time curves show the eye-movement pattern over the course of the entire utterance, the inferential analyses focus only on the theoretically most relevant effects.

During the VERB region, there was a main effect of more inspections to the patient than agent for both sentence conditions (MV, RR) ($LR\chi^2(subj) = 11.45, df = 1, p < 0.001; LR\chi^2(item) = 11.45, df = 1, p < 0.001$) in the absence of an interaction ($ps > 0.1$) (see Figure 4.9).

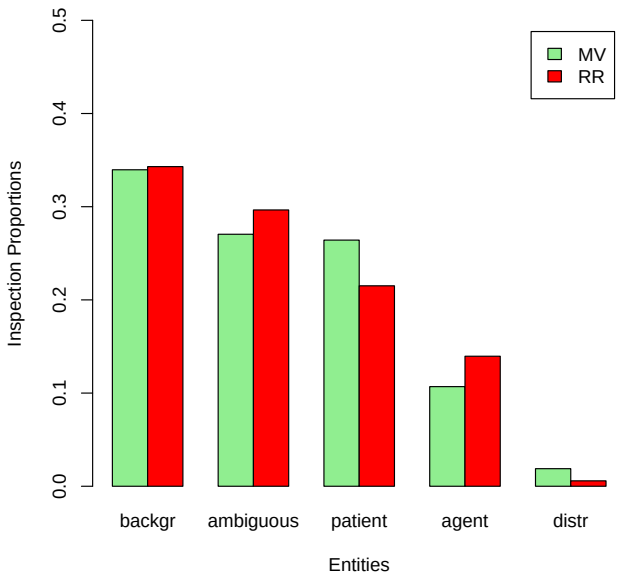


Figure 4.9: *Percentage of inspections to entities during the VERB region for the MV/RR condition in Experiment 4a*

During the ADV region (see Fig. 4.10), loglinear analyses revealed a significant interaction between target character (patient, agent) and sentence condition (MV, RR) ($LR\chi^2(subj) = 60.02, df = 1, p < 0.0001; LR\chi^2(item) = 64.00, df = 1, p < 0.0001$). The interaction was due to a higher proportion of inspections to the patient for the main verb in comparison with the reduced relative condition ($LR\chi^2(subj) = 40.40, df = 1, p < 0.0001; LR\chi^2(item) = 42.50, df = 1, p < 0.0001$), and to a greater percentage of inspection to the agent in the reduced relative compared with the main verb condition ($LR\chi^2(subj) = 35.52, df = 1, p < 0.0001; LR\chi^2(item) = 34.00, df = 1, p < 0.0001$).

For the SECOND ARGUMENT region (see Fig. 4.11), we found a significant interaction between target character (agent, patient), and sentence condition (MV, RR) ($LR\chi^2(subj) = 60.88, df = 1, p < 0.0001; LR\chi^2(item) = 59.35, df = 1, p < 0.0001$).

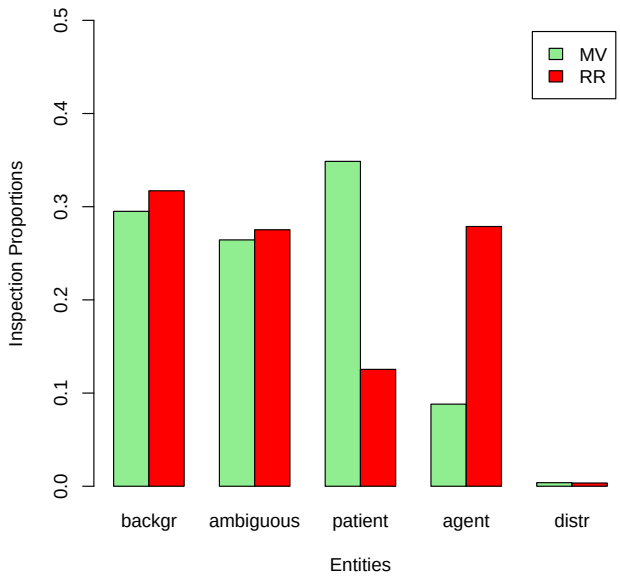


Figure 4.10: *Percentage of inspections to entities during the ADV region for the MV/RR condition in Experiment 4a*

The interaction was qualified by contrasts that showed it was due to a significantly higher proportion of patient inspections for main verb in comparison with reduced relative sentences ($LR\chi^2(subj) = 36.63, df = 1, p < 0.0001; LR\chi^2(item) = 36.74, df = 1, p < 0.0001$), and due to a higher inspection of fixations on the agent for reduced relative as opposed to main verb sentences ($LR\chi^2(subj) = 39.03, df = 1, p < 0.0001; LR\chi^2(item) = 34.79, df = 1, p < 0.0001$).

In sum, our findings from the main verb/reduced relative ambiguity (Experiment 4a) provide strong evidence for the rapid effects of depicted events on incremental thematic role assignment when the verb was second in the sentence and directly identified which depicted event was relevant for comprehension.

Early disambiguation for the main verb/reduced relative sentence occurred shortly after people had heard the verb and prior to disambiguation through the determiner/preposition on the second argument.

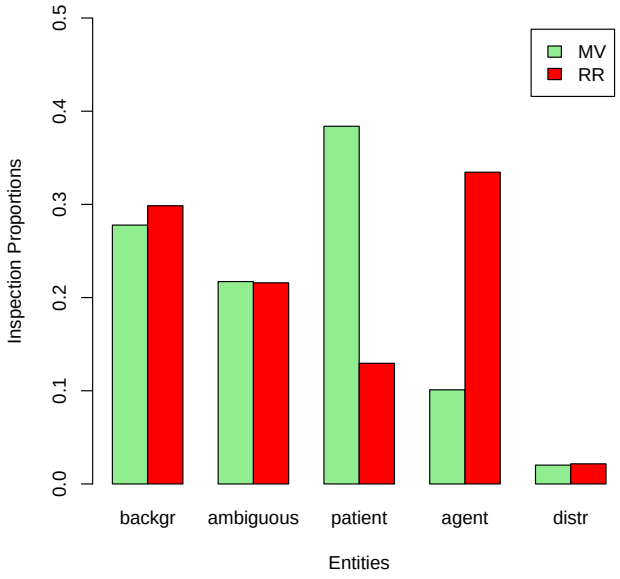


Figure 4.11: *Percentage of inspections to entities during the SECOND ARGUMENT region for the MV/RR condition in Experiment 4a*

EXPERIMENT 4B: 'WILL/BEING' CONDITIONS - INFERENTIAL ANALYSES

For the AUXADV region, there was a main effect of target character (patient, agent) $((LR\chi^2(subj) = 4.73, df = 1, p < 0.05; LR\chi^2(item) = 4.73, df = 1, p < 0.05)$ in the absence of an interaction (ps > 0.1) (see Fig. 4.12).

During the VERB region (see Fig. 4.13), log-linear analyses showed a main effect of target character (patient, agent) $(LR\chi^2(subj) = 11.50, df = 1, p < 0.001; LR\chi^2(item) = 11.50, df = 1, p < 0.001)$. Analyses further revealed a significant interaction between target character (patient, agent) and sentence condition (*will, being*) $(LR\chi^2(subj) = 30.16, df = 1, p < 0.0001; LR\chi^2(item) = 22.38, df = 1, p < 0.0001)$.

Contrasts showed that this was due to a higher proportion of inspections to the patient for the *will* as opposed to the *being* condition ($LR\chi^2(subj) = 34.59, df = 1, p < 0.0001; LR\chi^2(item) = 30.37, df = 1, p < 0.0001$). The increased proportion of inspections to the agent for the *being* in comparison with the *will* condition were not significant ($ps > 0.2$).

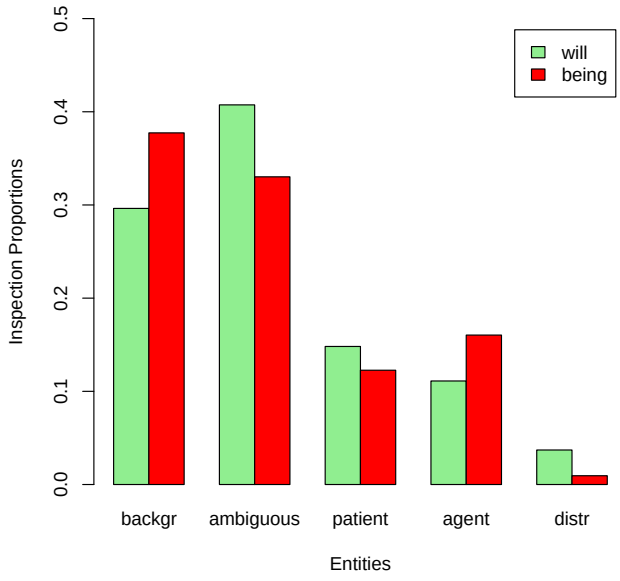


Figure 4.12: *Percentage of inspections to entities during the AUXADV region for the will/being condition in Experiment 4b*

For the SECOND ARGUMENT region (see Fig. 4.14), we found a significant interaction between target character (patient, agent) and sentence condition (*will, being*) ($LR\chi^2(subj) = 79.20, df = 1, p < 0.0001; LR\chi^2(item) = 90.96, df = 1, p < 0.0001$). The interaction was brought about by a higher proportion of inspections to the patient for the *will* in comparison with the *being* condition ($LR\chi^2(subj) = 50.38, df = 1, p < 0.0001; LR\chi^2(item) = 54.72, df = 1, p < 0.0001$), and by a higher percentage of inspections to the agent for the *being* as opposed to the *will* condition ($LR\chi^2(subj) = 59.81, df = 1, p < 0.0001; LR\chi^2(item) = 63.67, df = 1, p < 0.0001$).

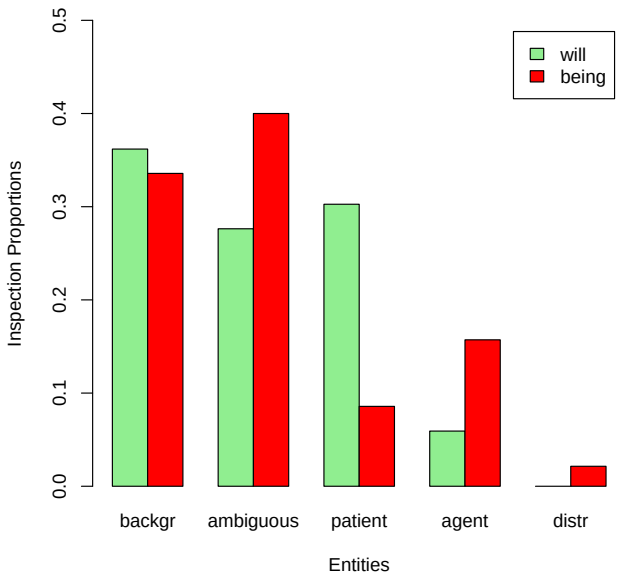


Figure 4.13: *Percentage of inspections to entities during the VERB region for the will/being condition in Experiment 4b*

Findings from Experiment 4b importantly provide evidence for the effects of soft linguistic auxiliary cues in making available the scene events for rapid disambiguation even during the lifetime of the verb. The pattern for disambiguation that we observed during the lexical verb was further confirmed during disambiguation due to the determiner/preposition on the second argument.

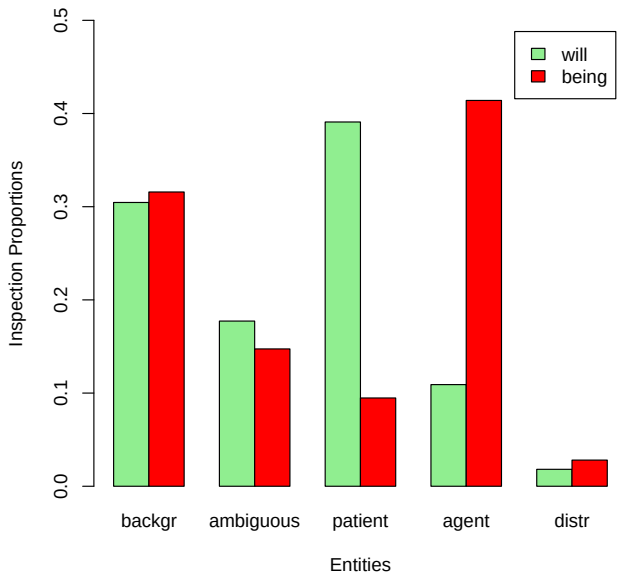


Figure 4.14: *Percentage of inspections to entities for the SECOND ARGUMENT region for the will/being condition in Experiment 4b*

COMPARING EXPERIMENTS 4A AND 4B

It seems worth pointing out one noticeable difference between Experiments 4a and 4b. While soft linguistic cues did make the scene events rapidly available, they did not have the same immediate and strong effect that the lexical verbs in Experiment 4a revealed. In Experiment 4a, disambiguation was clearly visible in the time curves during the region immediately following the verb (ADV). In contrast, for the auxiliaries, no such rapid disambiguation was found during the auxiliary and adverb region (AUXADV). While a direct comparison of the auxiliary region (Experiment 4b) with the verb region in Experiment 4a is not possible owing to substantial differences between these regions (the auxiliaries are closed class lexical items; the verbs are open class words), there is one time region that is identical between the two experimental condition pairs: The adverb region that succeeds

both the auxiliary (Experiment 4b) and verb regions (Experiment 4a) is identical for both condition pairs and can thus serve as a time window for a direct comparison.

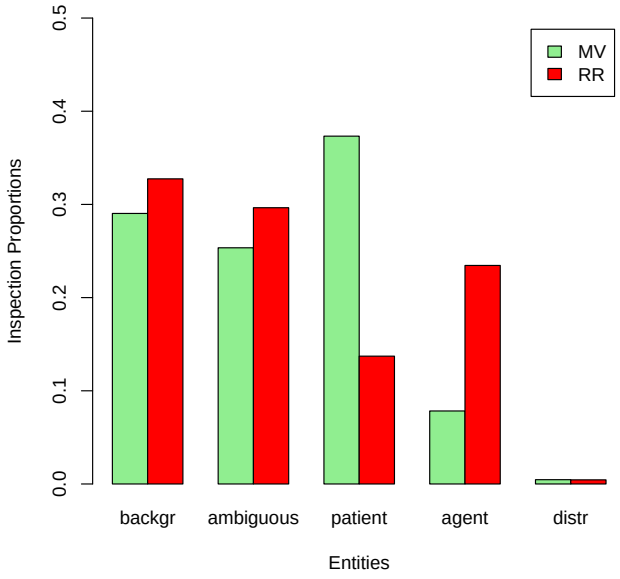


Figure 4.15: *Percentage of inspections to entities for the ADVEXACT region for the MV/RR condition in Experiment 4a*

Specifically, the region that we used for a direct comparison of findings from Experiments 4a and 4b stretched from the beginning of the adverb to the offset of the adverb in all of the conditions ('ADVEXACT'). Figures 4.15 and 4.16 present the gaze data for Experiment 4a and 4b respectively during this time region. It should be noted that despite the fact that the region which we use here for comparison differs from the ADV region used in Experiment 4a, and also from the AUXADV region in Experiment 4b, the findings that we reported for these two regions hold for the ADVEXACT region. There was a significant interaction between target character (patient, agent) and sentence condition (MV,RR) for Experiment 4a ($LR\chi^2(subj) = 39.85, df = 1, p < 0.0001, LR\chi^2(item) = 47.69, df = 1, p < 0.0001$) (see graph 4.15). In contrast, for the condition pair in Experiment 4b, we found no significant interaction (ps > 0.2) (graph 4.16).

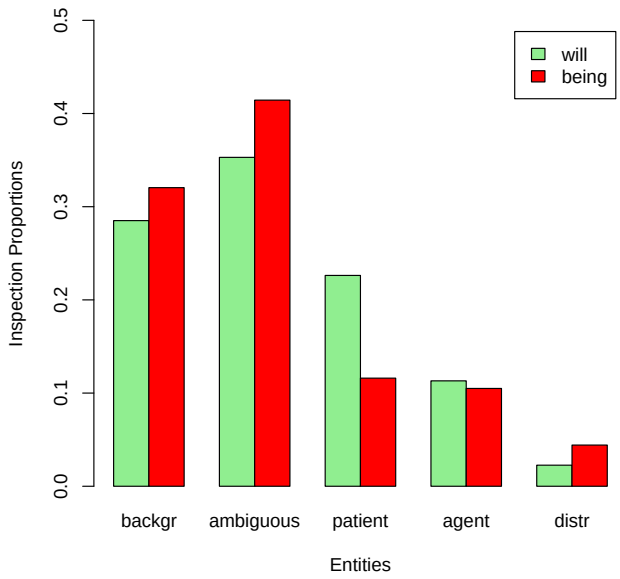


Figure 4.16: *Percentage of inspections to entities for the ADVEXACT region for the will/being condition in Experiment 4b*

In comparing these two results for the ADVEXACT time region, log-linear analyses revealed a significant three-way interaction between target character (patient, agent), sentence type (main verb, reduced relative) and disambiguating cue (verb, auxiliary) ($LR\chi^2(subj) = 3.56, df = 1, p = 0.05; LR\chi^2(item) = 5.85, df = 1, p < 0.05$). This indicates that the difference between the interaction that provided evidence for disambiguation in Experiment 4a (graph 4.15) and the absence of such a disambiguation in Experiment 4b (graph 4.16) is significant.

SUMMARY

To conclude let us summarize findings from the present chapter. Experiment 3 provided evidence for the claim that depicted agent-action-patient events play an active role in

on-line structural disambiguation and incremental thematic role assignment. The findings from this experiment demonstrate that soft linguistic cues in the utterance can make depicted events rapidly available for early structural disambiguation when the verb only identified the relevant depicted event late during the utterance due to its sentence-final position.

Experiment 4 revealed that findings from Experiments 1 to 3 generalize and extend to another language (English) and to a further type of structural ambiguity, the main verb/reduced relative ambiguity. Experiment 4a demonstrated that findings of the rapid verb-mediated influence of depicted events on incremental thematic role assignment and structural disambiguation from Experiments 1 and 2 generalize to the English main verb/reduced relative clause sentences. Experiment 4b investigated whether findings from Experiment 3 extend to the English reduced relative clause ambiguity. The soft linguistic cues (auxiliaries) that we investigated in Experiment 4b had no immediate effect on disambiguation during the time region that followed the auxiliaries. This contrasts with findings from Experiment 3. A direct comparison of the adverb region between Experiment 4a and 4b revealed further that the lexical verbs in Experiment 4a are possibly a stronger cue in making the depicted events available.

Importantly, however, in Experiment 4b the effects of the auxiliary cues became apparent *during* the verb region that followed the auxiliary and adverb. In contrast, for Experiments 1, 2, and 4a, disambiguation had always taken place *after* the verb. The disambiguation effect thus shifted one region earlier when soft cues preceded the verb in comparison to when no subtle linguistic cues preceded the lexical verb (Exps 1, 2, and 4a). This finding suggests that subtle linguistic cues such as the auxiliaries that we investigated can make depicted agent-action-patient and patient-action-agent events rapidly available for disambiguation towards either an agent-patient or a patient-agent structuring of the thematic role relations in the utterance. Taken together, findings from both Experiments 3 and 4 demonstrate that depicted events play a relatively active role in the incremental assignment of thematic roles to utterance constituents.

Chapter 5

The Relative Importance of Information

We have so far presented four experiments that demonstrated that depicted events can enable structural disambiguation and incremental thematic role assignment. In Experiments 1, 2, and 4 the scene events were mediated through verbs in the utterance, whereas in Experiments 3 and 4 we in addition investigated whether depicted events can influence comprehension when mediated through soft linguistic cues (see section 2.5 and chapters 3 and 4). For the present study, we will concentrate on verb-based mediation of depicted events. Importantly, in Experiments 1, 2, and 4, the informational source that informed incremental thematic role assignment (a depicted event) was *uniquely identified* through a verb in the utterance. Thus, we know that depicted events can influence incremental thematic role assignment. We do not know, however, how great their influence is *relative* to other informational sources that have also been found to enable incremental thematic role assignment such as stored thematic role knowledge (e.g., Kamide, Scheepers, & Altmann, 2003). We have suggested that determining their relative importance would allow us to specify in more detail how distinct informational sources such as scene information and stored thematic role knowledge combine (see section 2.5).

The experiment presented in this chapter explores the *relative importance* of depicted events and stored thematic role knowledge, two informational sources that have each been revealed as influential in incremental thematic role assignment when uniquely identified by the utterance. The present experiment pursues two goals: The first is to replicate the finding that both depicted events and stored thematic role knowledge can influence

incremental thematic role assignment when uniquely identified by the utterance. Second, within the same experiment we further explore which of these two types of information the comprehension system prefers to rely on for incremental thematic role assignment when a choice must be made between them (i.e., when the utterance identifies both a stereotypical agent and a different agent of a depicted action as relevant for comprehension).

5.1 STORED THEMATIC KNOWLEDGE VERSUS DEPICTED EVENTS

We first summarize relevant experimental evidence that has shown that both of these kinds of knowledge can be used. As detailed in section 2.1.1, Kamide, Scheepers, and Altmann (2003) have shown that unambiguous case-marking, and verb-based world-knowledge influence the anticipation of objects that are plausible thematic role fillers of the verb. Since their findings are of importance for the ensuing discussion, we briefly summarize once again the key issues of their study. In German, a case-marked article can determine the grammatical function and thematic role of the noun phrase it modifies. Both SVO and OVS orders are grammatical. Participants inspected images showing a hare, a cabbage, a fox and a distractor object while hearing sentences such as *Der Hase frisst gleich den Kohl* ('The hare (subj) eats soon the cabbage (obj)') and *Den Hasen frisst gleich der Fuchs* ('The hare (obj) eats soon the fox (subj)'). The subject and object case-marking on the article of the first noun phrase together with world knowledge about what is likely to eat what extracted at the verb allowed anticipation of the correct post-verbal referent. This was evidenced by anticipatory eye-movements to the cabbage after participants had heard 'The hare (subj) eats ...' and to the fox after having encountered 'The hare (obj) eats ...'. Hence, when the utterance is unambiguous, and linguistic/world knowledge restricts the domain of potential referents in a scene, the comprehension system may anticipate mention of scene objects.

The first two experiments of this thesis, in contrast, have considered structurally ambiguous utterances where neither case-marking nor stereotypical knowledge could assist in early disambiguation. Specifically, they examined the time course with which listeners were able to resolve an initial structural and thematic role ambiguity in German sentences. Since the linguistic input did not determine the correct syntactic analysis and thematic role assignment of the sentence, listeners had to rely on depicted events in the

scene for interpretation of the utterance. The events showed, for example, a princess washing a pirate, while a fencer painted her. The princess was thus determined as either patient or agent of an event depending on the depicted action (washing/painting respectively). Listeners heard *Die Prinzessin wäscht/malt den Pirat/der Fechter.* ('The princess (amb.) washes/paints the pirate (obj./patient)/the fencer (subj./agent)'). Once the verb had identified the relevant depicted action, anticipatory eye-movements to the appropriate other event participant (the pirate or the fencer) were observed. The anticipation of a patient and agent role-filler for initially ambiguous German subject-verb-object and object-verb-subject sentences respectively suggests rapid use of depicted events in resolving structural and thematic role ambiguity on-line (see chapter 3). This finding was extended to the English main clause (MC)/reduced relative (RR) ambiguity, and hence generalized to another language and construction (see chapter 4).

The above studies have shown that both stored thematic role knowledge (even when modulated by morpho-syntactic cues) and depicted events are rapidly applied for incremental thematic role assignment. As evidenced by anticipatory eye-movements in the scenes, thematic role assignment took place rapidly once the verb had either uniquely identified a depicted action or restricted the post-verbal domain of reference to a stereotypical entity. It is further clear, that utterance, world knowledge and the immediate visual scene interact rapidly during on-line comprehension. What remains unclear is the nature of that interaction (see section 2.5). Among other questions, we might ask: What is the relative importance, or priority of different information sources, such as linguistic, scene, and world knowledge? Does the comprehension system rely preferably on visually established event relations for incremental thematic interpretation, or on thematic relations established through stored linguistic and world knowledge?

To further investigate this empirical issue, consider an example from German. Recall that German has a rich case marking system where grammatical function is usually indicated by unambiguous case morphemes. Word order constraints are less rigid in German than in English, and both SVO and OVS order are grammatical with SVO being the preferred reading (e.g., Hemforth, 1993). In our example, people hear an OVS sentence while inspecting a scene such as the one presented in Figure 5.1a. Since there are no linguistic constraints on inspecting a scene prior to utterance onset, perceivers might notice a pilot, a wizard holding a telescope, and an entity resembling a detective who is serving some food. On hearing an OVS sentence fragment such as 'The pilot (object/patient) jinxes...' while inspecting the example scene in Figure 5.1a, a number of processes occur. When

we hear 'The pilot', object case-marking permits assignment of a patient role to the noun phrase while establishing reference to the pilot in the scene (see Kamide, Scheepers, & Altmann, 2003). Then we hear the verb, *jinxes*. Having encountered a patient and the verb, we might expect post-verbal mention of an agent. At this point, our knowledge that a wizard is a likely jinxing-agent may combine with the fact that the wizard is the only entity whose affordances match the expectations raised by the verb (e.g., a cake affords an eating action; see section 2.3). Findings by Kamide, Scheepers, and Altmann (2003) suggest that the combination of entity affordance and verb-based stereotypical knowledge allows us to anticipate the wizard as a likely-to-be-mentioned agent (Fig. 5.1a).

In contrast, when we hear 'The pilot (object/patient) serves food to...', verb-based thematic knowledge of stereotypical agents of a serving-food action (e.g., a cook), cannot enable incremental thematic role assignment, since the scene affords no such entity. However, the scene does contain a depicted food-serving event performed by the detective. While stereotypical knowledge does not allow determination of thematic role-relations in this case, the depicted scene events do. Based on findings from the first two experiments of this thesis, we would expect unmistakable resolution of the temporal uncertainty regarding the yet-to-be-mentioned agent once people have heard the verb 'serve-food-to' that identifies the depicted action (see Fig. 5.1a). It should be noted that in both of these examples the verb uniquely identified either a depicted action and its associated agent, or it restricted the post-verbal domain of reference to a stereotypical agent.

Imagine we heard instead 'The pilot (object/patient) spies-on...' In this case, the verb does not uniquely identify either only a depicted action or a stereotypical agent. Rather, the scene affords both a stereotypical agent (the detective), and an immediately depicted agent of a spying event (the wizard) as potential agents (see Fig. 5.1a). When we hear 'spies-on', lexical access makes available the meaning of the lexical item and stereotypical knowledge related to it (see Ferretti et al., 2001). After encountering the verb, word meaning, stereotypical knowledge of 'spy-on', and scene affordances are available to anticipate either a depicted spying-event and its agent (the wizard), or a stereotypical agent (the detective).

Do listeners rely more on extracting thematic role relations from stereotypical knowledge provided by the utterance ('spy-on' + WORLD KNOWLEDGE → detective), or do they rely on thematic relations afforded by the scene ('spy-on' + WIZARD-SPYING EVENT → wizard) in incremental interpretation? For the ambiguous 'spy-on' example information about who-does-what-to-whom in the depicted events conflicts with stereotypical knowl-

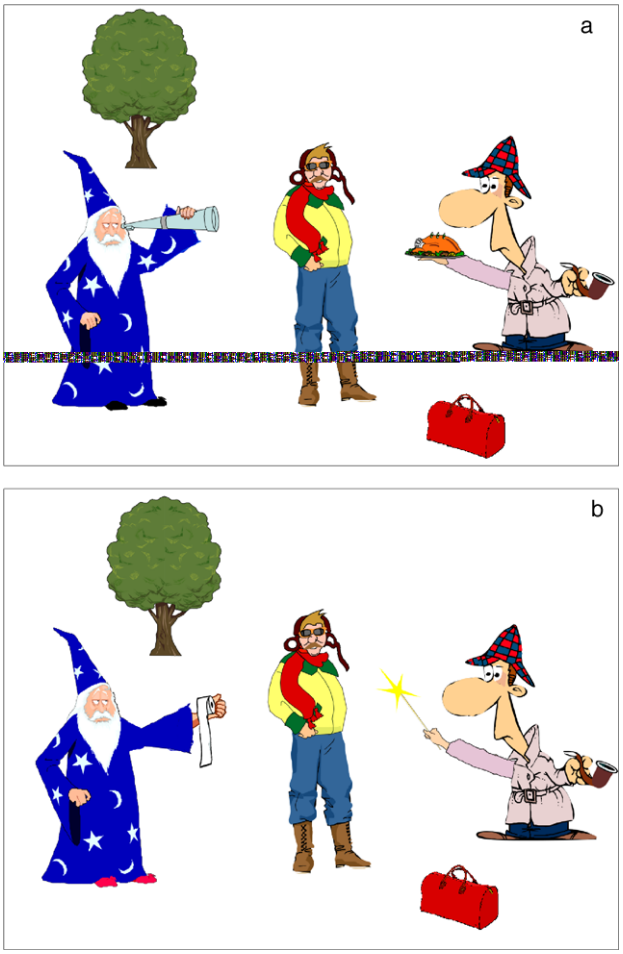


Figure 5.1: *Example image for sentences in Table 5.1*

edge of who-does-what-to-whom. The comprehension system has to choose between two available yet conflicting types of information in determining on-line thematic role assignment.

In summary, we expect the following: When the verb uniquely determines either a depicted or a stereotypical target agent only (Table 5.1, (a1) and (a2)), verb-derived knowledge of stereotypical role-fillers and depicted events should each allow anticipation

of the appropriate target agent. This would replicate - within a single study - findings by Kamide, Scheepers, and Altmann (2003) and findings that have been presented in this thesis (Experiments 1 and 2). When the verb is 'serves-food-to' (a1), we expect a higher percentage of anticipatory looks to the only depicted food-serving agent (the detective) than to the respective other agent in the scene (the wizard) before the second noun phrase. Conversely, when the verb is 'jinxes' (a2), more looks should occur to the stereotypical agent (the wizard) than to the other agent in the scene (the detective) before the subject noun phrase.

In the interesting case of competition, when the verb ('spy-on') allows more than one potential scene entity as target agent, no early interaction is expected since the utterance is identical for conditions (b1) and (b2) up to the onset of the second noun phrase (see Table 5.1). There are several possibilities of what to expect in this case. While we associate theoretical positions in the literature with various possible experiment outcomes below, it should be noted that to the best of our knowledge no existing theory/account makes explicit prediction about the relative importance of two distinct informational sources such as depicted events and stored thematic role knowledge.

In section 2.5, we suggested that a line of research such as the one pursued by Altmann and Kamide (1999) and Kamide, Scheepers, and Altmann (2003) would appear to be best associated with a greater importance of stored thematic role knowledge in comprehension. If the comprehension system relies on stored thematic role knowledge in preference to depicted events, then this should be revealed in a higher percentage of inspections to the stereotypical spying-agent (the detective) than to the depicted spying-agent (the wizard) for sentences (b1) and (b2) shortly after the verb (see Fig. 5.1 and Table 5.1). Crucially, these looks should occur before people hear the disambiguating second noun, and hence reveal on-line expectations of thematic role interpretation.

Alternatively, if we assume scene information is "just another " constraint in a model such as the one used by Tanenhaus et al. (2000) or a framework such as Jackendoff (2002), we would expect no clear preference provided both information types are comparable in strength (this would be expected if they can each inform comprehension, a point tested by conditions (a1) and (a2), see Table 5.1). While it is not clear which predictions the procedurally underspecified Jackendovian account makes, the model by Tanenhaus et al. (2000) predicts comprehension difficulty in this case and hence we should not find a higher proportion of anticipatory eye-movements to any individual scene object in comparison with another. Competition should continue until the second noun phrase disambiguates

towards the respective target type (depicted, stereotypical).

A third possibility is that we observe a preferred use of thematic relations proffered by the scene in incremental thematic interpretation. If this was true, we expect to observe a higher proportion of anticipatory eye-movements to the depicted agent than to the stereotypical agent for both conditions (b1) and (b2).

5.2 EXPERIMENT 5

PARTICIPANTS

Twenty-four German native speakers with normal or corrected-to-normal vision received 5 euro for participation.

5.2.1 METHOD

MATERIALS, DESIGN, AND PROCEDURE

We created 48 images using commercially available graphic programs. An image showed two agents (e.g., a wizard, and a detective) that each performed an action upon a patient (e.g., the pilot, see Fig. 5.1a). The depicted events provided information about thematic role-relations (e.g., wizard-spying-on-pilot and detective-serving-food-to-pilot, see Fig. 5.1a). In addition, each agent provided stereotypical thematic role knowledge (e.g., a detective is a stereotypical agent of a spying action, and a wizard a stereotypical agent of a jinxing action). We refer to the agent of a depicted event as the 'depicted agent' and to an entity that is identified as agent through verb-based stored thematic knowledge as the 'stereotypical agent'.

There were two versions of an image (Fig. 5.1a and 5.1b). This ensured that each verb identified once the agent of a depicted event (e.g., 'jinx', detective-jinxing Fig. 5.1b), and once a stereotypical agent ('jinx', wizard-jinxing Fig. 5.1a). It further ensured that each of the target agents (wizard, detective) was once identified as a potential agent through verb-mediated depicted actions as a depicted agent (wizard-bandaging, Fig. 5.1b) and through stereotypical verb-based knowledge (e.g., a wizard is a stereotypical agent of a jinxing action, Fig. 5.1a). As a result of this design, plausible/depicted biases were counter balanced.

We recorded 4 sentences for each image (Table 5.1, (a1), (a2), (b1), (b2)). Manipulation of the verb created four conditions, crossing the factors *target type* (depicted,

Table 5.1: *Example of sentence pairs for image (a) in Figure 5.1*

Image	Condition		Sentences
5.1:a	Unique identification & Depicted. target	(a1)	Den Piloten verköstigt gleich der Detektiv. The pilot (PAT.) serves-food-to soon the detective. 'The detective will soon serve food to the pilot.'
5.1:a	Unique identification & Stereotypical target	(a2)	Den Piloten verzaubert gleich der Zauberer. The pilot (PAT.) jinxes soon the wizard. 'The wizard will soon jinx the pilot.'
5.1:a	Ambiguous identification & Depicted target	(b1)	Den Piloten bespitzelt gleich der Zauberer. The pilot (PAT.) spies-on soon the wizard. 'The wizard will soon spy on the pilot.'
5.1:a	Ambiguous identification & Stereotypical Target	(b2)	Den Piloten bespitzelt gleich der Detektiv. The pilot (PAT.) spies-on soon the detective. 'The detective will soon spy on the pilot.'

stereotypical) with *identification* (unique, ambiguous). 'Identification' refers to whether the verb uniquely identifies a target agent based on either stored thematic knowledge or depicted events ((a1) and (a2)), or whether it identifies both available agents (a stereotypical agent and an agent of a depicted action) as relevant ((b1) and (b2)) ('ambiguous identification'). 'Target type' refers to which target agent type (depicted , stereotypical) the second noun phrase identifies as the appropriate agent. In conditions (a1) and (b1) the target type is depicted and in conditions (a2) and (b2) the target type is stereotypical. Note the difference between 'target type' and 'identification': 'Identification' characterizes which informational source is identified as relevant by the *verb*, whereas 'target type' refers to the target agent type that is identified by the *second noun phrase*. When we refer to the target agent identified by the second noun phrase we will use the expression 'target type'. When we refer, in contrast, to the scene entities, we will use the expressions 'target agent' (depicted agent and stereotypical agent).

For the unique identification conditions (see Table 5.1, sentences (a1) and (a2)), the verb permitted either a depicted (a1) or a stereotypical agent (a2) only. For the ambiguous identification conditions (see Table 5.1, sentences (b1) and (b2)) the verb *bespitzeln* ('spy-on') allowed two scene entities as relevant agents (Fig. 5.1a).

Sentences were unambiguous OVS sentences. They always started with an object case-marked noun phrase referring to a patient role-filler (Fig. 5.1a, the pilot). While the OVS structure is dispreferred, this was a constant across conditions. Conditions were matched for length (number of spoken syllables), and frequency of lemmas as much as possible (Baayen et al., 1995). For the image in Fig. 5.1a, the sentences in Table 5.1 were recorded. A set of 24 items was created. Each consisted of 8 spoken sentences and 2 images (Table 5.1 and Fig. 5.1a represent one half of such an item set).

In addition to the 24 items, there were 48 filler items. The filler items were balanced for the number of stereotypical and depicted actions and agents. They further contained twelve filler trials that showed a depicted stereotypical action. This was done to minimize the possibility that participants pay less attention to their stored linguistic and world knowledge than to the immediate (implausible) depicted events in on-line comprehension. Such strategy could be induced if participants never encountered a plausible action that was also depicted. Including twelve filler items where a plausible action was depicted should help to avoid this problem. Experimental items were separated from one another by at least one filler item.

An SMI Eye-Link head-mounted eye-tracker monitored participants' eye-movements in the depicted scenes while participants were listening to sentences that described part of the scenes. Each participant heard only one of the eight sentences of an item, and the order of appearance of items was randomized individually for every participant. The preview time for images was 1500 ms. Prior to the experiment, participants were told to listen to the sentences and to inspect the images, and to try to understand both sentences and depicted scenes. There was no other task. The entire experiment lasted approximately 25 min.

ANALYSIS REGIONS

The time regions for the inferential analyses were the verb ('VERB'), the post-verbal adverb ('ADV'), and the second noun phrase ('NP2'). The VERB region stretched from verb onset to the onset of the adverb (mean duration of 1003 ms for (a1), of 1008 ms for

(a2), of 989 ms for condition (b1), and of 999 ms for (b2),; see Table 5.1 for the labelling of conditions). The ADV region went from the onset of the adverb to the onset of the second noun phrase (mean duration of 1004 ms for (a1), 1001 ms for (a2), 1013 ms for (b1), and 1002 ms for (b2)). The NP2 region extended from the onset of the second noun phrase to its offset (mean duration of 1056 ms for (a1), 1042 ms for (a2), of 1062 ms for (b1), 1064 ms for (b2)).

5.2.2 RESULTS AND DISCUSSION

We report the mean proportions of inspections to the target agents over the course of the utterance (Figs 5.2 and 5.3), and present descriptive and inferential analyses of the proportion of inspections for individual time regions (Figs 5.4 to 5.9).

Figures 5.2 and 5.3 present the time course of the eye-movements, and display the mean proportion of inspections to entities in time frames of 250 ms. Figure 5.2 displays inspections to the scene entities (depicted agent, stereotypical agent) in the *unique identification* condition (Table 5.1, (a1) and (a2)). Figure 5.3 displays inspections to the depicted and to the stereotypical agent in the *ambiguous identification* conditions (Table 5.1, (b1) and (b2)). Inspections to the background and the distractors were left out for clarity of presentation, as were inspections to the patient. All of these are, however, displayed in the graphs of the individual time regions (see Figs 5.4 to 5.9).

Let us first briefly recall the unique identification condition: In this condition, the verb ('jinx') either only restricts the post-verbal domain to a stereotypical agent (e.g., the wizard, see Table 5.1 and Fig. 5.1a) or to the agent of a depicted action (e.g., the detective who is serving food). Based on findings by Kamide, Scheepers, and Altmann (2003) and findings from chapter 3 of this thesis, we expected to find evidence for early disambiguation in the eye-gaze patterns shortly after people have heard the verb. Indeed, for the unique identification conditions ((a1), (a2), see Table 5.1, Fig. 5.1a), when the verb singled out either a depicted (a1) or a stereotypical agent (a2), the time curves show clear *early* disambiguation during the life-time of the adverb using either depicted or stereotypical information about who-does-what-to-whom (see Fig. 5.2).

The inspection patterns in Figure 5.2 show that shortly after adverb onset inspections to the depicted agent increase when the verb identified a depicted target ('serve-food-to' - detective-serving-food, (a1)), while at the same time fixations to the stereotypical agent for the depicted target condition are relatively low. In contrast, when the verb restricted the

post-verbal domain of reference to a stereotypical agent ('jinx' – wizard, (a2)), inspections to the stereotypical agent clearly exceeded inspections to the other agent (the detective) during the adverb region. While the eye-gaze pattern during the adverb renders (in the time curves) the impression that overall there were more looks to the depicted agent in the depicted target condition than to the stereotypical agent in the stereotypical target condition, this impression was importantly *not* confirmed by the inferential analyses. It should be noted that the inferential analyses provide a more precise picture of the eyegaze data since they (but not the time curves) are based on trial-by-trial computations of the precise time regions (see chapter 3 for further details).

This gaze pattern continues throughout the second noun phrase, a fact which provides corroboratory evidence for the successful influence of both stored thematic relations and depicted event relations in incremental thematic role assignment. Findings in the unique identification conditions thus clearly replicated the eye-gaze patterns from Experiments

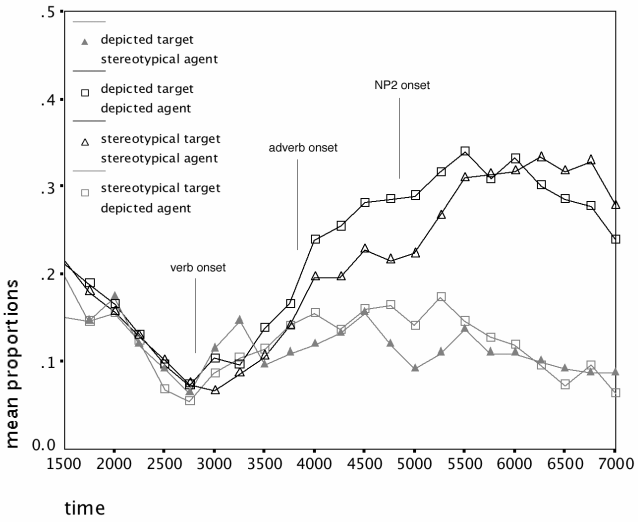


Figure 5.2: *Time course of the eye-movement data for Experiment 5 (unique identification conditions) showing the mean proportion of inspections to target agents from the onset of the spoken stimuli in time frames of 250 m*

1 and 2 of the present thesis and from Kamide, Scheepers, and Altmann (2003). They demonstrate that - when uniquely identified - both stereotypical knowledge and depicted information about who-does-what-to-whom enable early incremental thematic role assignment.

Recall now the ambiguous identification conditions: In contrast to the unique identification conditions, the verb in the ambiguous identification conditions did not uniquely identify either a depicted or a stereotypical agent. Rather, it was compatible with both a stereotypical agent and the agent of a depicted action. Indeed, the stimuli for (b1) and (b2) were identical prior to the second noun phrase (see Table 5.1). We thus expected no early disambiguation towards either a depicted or a stereotypical agent during the adverb time region for the ambiguous identification condition ((b1), (b2), see Table 5.1).

Rather, we expected a main effect with the interesting point being the directionality of the effect. When people hear 'spy-on', do they rely in preference on the depicted information about who-does-what-to-whom and anticipate a depicted but non-stereotypical agent (the wizard), or do they rely on stereotypical thematic role knowledge and anticipate a stereotypical agent but that is not depicted as performing the verb-action (the detective)?

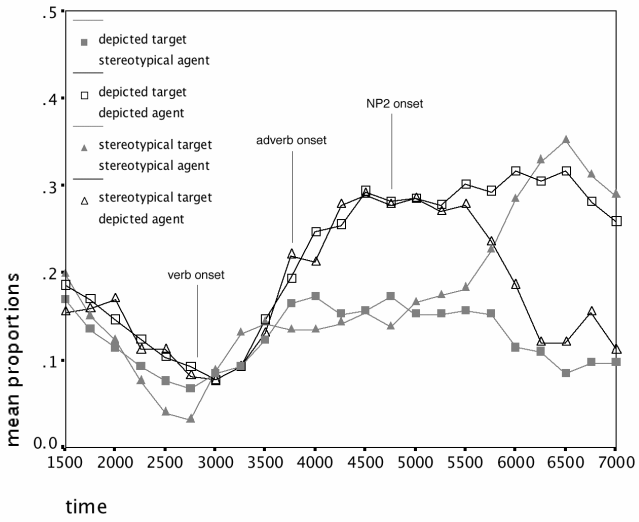


Figure 5.3: *Time course of the eye-movement data for Experiment 5 (ambiguous identification conditions) showing the mean proportion of inspections to the target agents from the onset of the spoken stimuli in time frames of 250 m*

We observed more anticipatory fixations to the agent of the depicted spying-event (the wizard), than to the stereotypical agent (the detective) for sentences (b1) and (b2) as shown in Figure 5.3. These looks occurred after people had heard 'The pilot (object/patient) spies-on...' and crucially before they heard the respective second noun, which then disambiguated towards the depicted (b1) or stereotypical (b2) target type. This means that the comprehension system had to revise the interpretation of the utterance for the stereotypical target condition (b2) during the second noun phrase. Indeed, the eye-movement patterns provide evidence for revision of the initial depicted-event preference. Late during the second noun phrase region, inspections to the stereotypical agent for the stereotypical target condition (b2) increase, while at the same time looks to the depicted agent for this condition decrease. After this disambiguation the pattern of inspections is the same as for the unique identification conditions (see Figs 5.2 and 5.3).

The core finding is the preference of comprehension processes to rely on depicted actions and their associated thematic relations for *early* thematic role assignment in the condition in which the verb identified both thematic relations established by stored knowledge and by the scene events as relevant (Fig. 5.3, ambiguous identification). Clearly, this finding is only meaningful in comparison with the early disambiguation towards either a depicted or a stereotypical agent in the unique identification condition when the verb uniquely identified depicted or stereotypical thematic relations as relevant for comprehension ((a1) and (a2), Table 5.1 and Fig. 5.2).

UNIQUE IDENTIFICATION CONDITIONS

The inferential analyses confirmed the observations made on the basis of the time course of eye-movements. We first present the inferential analyses for the unique identification condition (Figs 5.4 to 5.6). During the VERB region, there were no significant effects that involved the analysis variables target agent or target type (all $ps > 0.1$) (see Fig. 5.4).

Figure 5.5 shows the proportion of inspections to scene entities during the ADV region. For the unique identification condition, there was a significant interaction between target agent (depicted agent, stereotypical agent) and target type (depicted target, stereotypical target) ($LR\chi^2(subj) = 18.53, df = 1, p < 0.0001; LR\chi^2(item) = 27.30, df = 1, p < 0.0001$). There was no main effect of target type (depicted target, stereotypical target) or target agent for the target type conditions (all $ps > 0.2$).

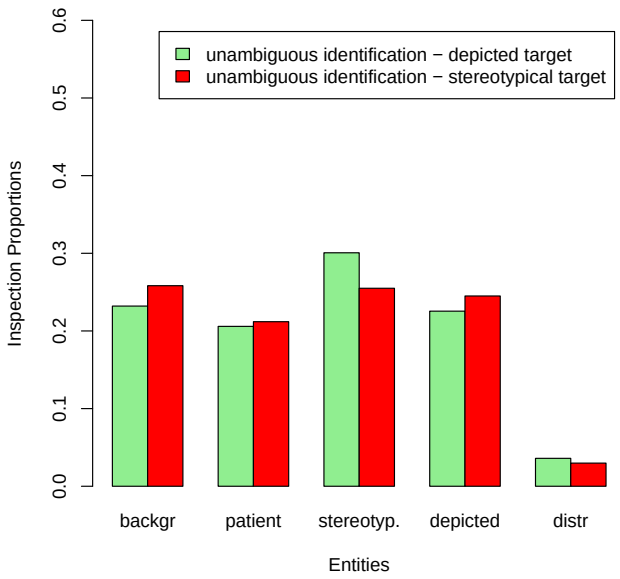


Figure 5.4: *Percentage of inspections to entities during the VERB region for the unique identification condition in Experiment 5*

Log-linear contrast revealed that the interaction was due to a significantly higher percentage of inspections to the depicted agent than to the stereotypical agent for the depicted target condition ((a1), Table 5.1) ($LR\chi^2(subj) = 16.01, df = 1, p < 0.001; LR\chi^2(item) = 16.01, df = 1, p < 0.001$), and due to a higher percentage of inspections to the stereotypical agent in comparison with the depicted agent for the stereotypical target condition ($LR\chi^2(subj) = 12.81, df = 1, p < 0.001; LR\chi^2(item) = 12.81, df = 1, p < 0.001$).

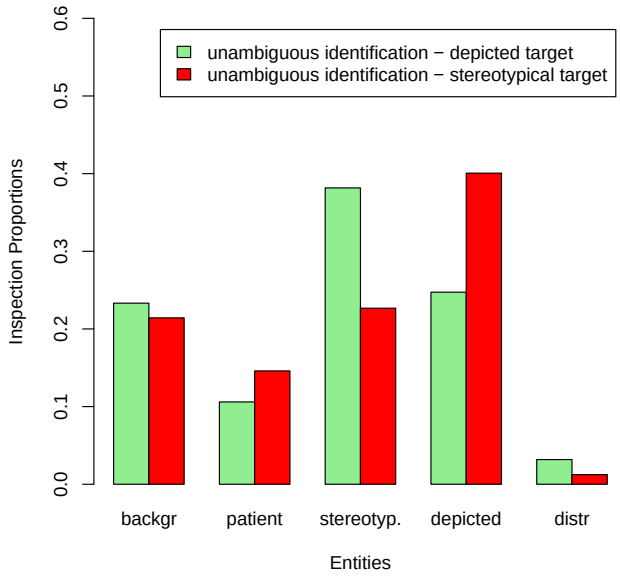


Figure 5.5: *Percentage of inspections to entities during the ADV region for the unique identification condition in Experiment 5*

For the NP2 region, analyses revealed a significant interaction between target agent (depicted agent, stereotypical agent) and target type (depicted, stereotypical) ($LR\chi^2(subj) = 47.55, df = 1, p < 0.0001; LR\chi^2(item) = 52.41, df = 1, p < 0.0001$) (see Fig. 5.6). Contrasts showed that this was due to a higher proportion of inspections to the depicted agent than to the stereotypical agent for the depicted target type ($LR\chi^2(subj) = 4.82, df = 1, p < 0.05; p > 0.1$ by items), and to a higher proportion of inspections to the stereotypical than to the depicted agent for the stereotypical target type sentences ($LR\chi^2(subj) = 851.72, df = 1, p < 0.0001; LR\chi^2(item) = 851.72, df = 1, p < 0.0001$).

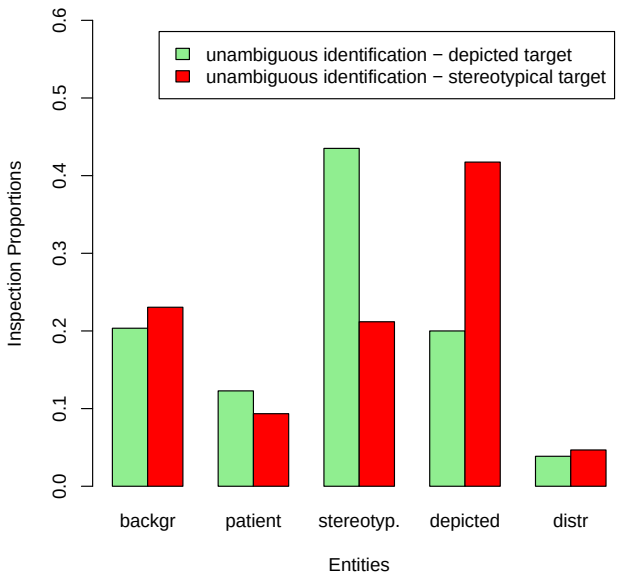


Figure 5.6: *Percentage of inspections to entities during the NP2 region for the unique identification condition in Experiment 5*

AMBIGUOUS IDENTIFICATION CONDITIONS

For the VERB region, there was no significant main effect of target agent (depicted agent vs. stereotypical agent) or target type (depicted, stereotypical) ($LR\chi^2s < 1$), nor was there a significant interaction of these two factors ($LR\chi^2s < 1$) (see Fig. 5.7).

Figure 5.8 shows the proportions of inspections to entities during the ADV region for the ambiguous identification condition. We found a significant main effect of target agent (depicted agent, stereotypical agent) ($LR\chi^2(subj) = 31.66, df = 1, p < 0.0001; LR\chi^2(item) = 31.66, df = 1, p < 0.0001$), in the absence of an interaction of target agent and identification ($LR\chi^2s < 1$).

The difference between the main effect for the ambiguous identification condition ((b1) and (b2)), and the significant interaction for the unique identification condition ((a1)

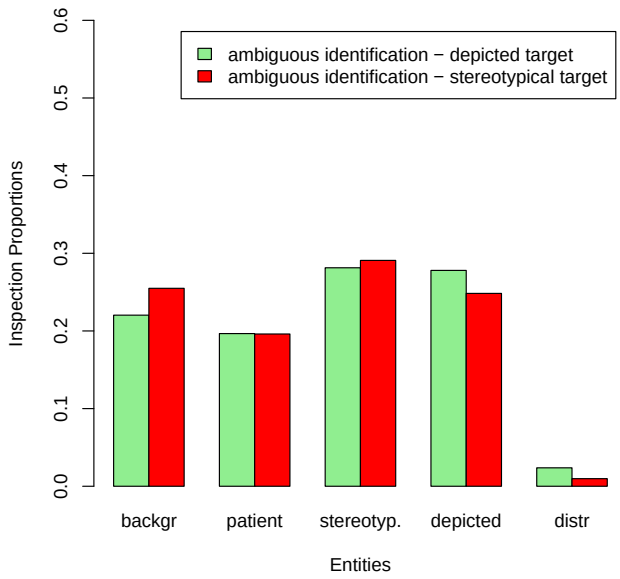


Figure 5.7: *Percentage of inspections to entities during the VERB region for the ambiguous identification condition in Experiment 5*

and (a2), see Table 5.1) was significant for the ADV region. Analyses revealed a three-way interaction between target agent (depicted agent, stereotypical agent), identification (unique, ambiguous) and target type (depicted target, stereotypical target) ($LR\chi^2(subj) = 9.20, df = 1, p < 0.01; LR\chi^2(item) = 11.72, df = 1, p < 0.001$).

For the NP2 region in the ambiguous identification condition (Fig. 5.9), log-linear analyses revealed that the interaction between target agent and identification was not significant ($p > 0.1$). However, when extending the NP2 region until the end of the utterance, the interaction turned significant. For this later NP2 region (from NP2 onset to the end of the utterance), there was an interaction between target agent (depicted agent, stereotypical agent) and identification (depicted, stereotypical) ($LR\chi^2(subj) = 62.74, df = 1, p < 0.0001; LR\chi^2(item) = 62.50, df = 1, p < 0.0001$). Contrasts revealed a significantly higher proportion of inspections to the depicted than to the stereotypical agent for the depicted

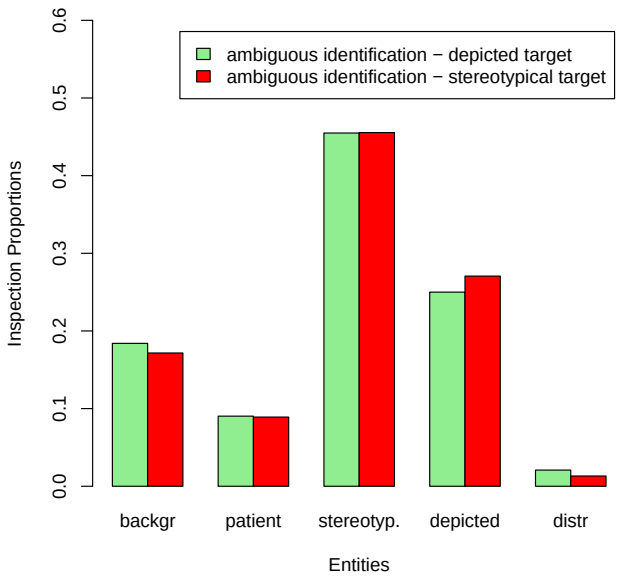


Figure 5.8: *Percentage of inspections to entities during the ADV region for the ambiguous identification condition in Experiment 5*

target condition ($LR\chi^2(subj) = 20.68, df = 1, p < 0.0001; LR\chi^2(item) = 20.68, df = 1, p < 0.0001$), and a significantly higher percentage of inspections to the stereotypical in comparison with the depicted agent for the stereotypical target type condition ($LR\chi^2(subj) = 11.27, df = 1, p < 0.001; LR\chi^2(item) = 11.27, df = 1, p < 0.001$).

SUMMARY

Within a single experiment, the observed pattern of inspections has shown that both stereotypical thematic role knowledge and depicted events each allow rapid thematic role interpretation of an unfolding utterance. In this respect, our findings have replicated the findings from Kamide, Scheepers, and Altmann (2003) and the findings presented in chapter 3 of the present thesis. When the verb uniquely identified the agent of a stereotyp-

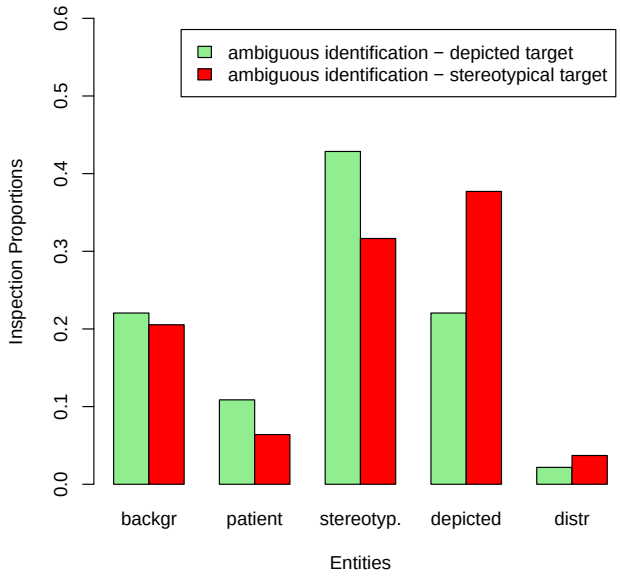


Figure 5.9: *Percentage of inspections to entities during the NP2 region for the ambiguous identification condition in Experiment 5*

ical (non-depicted) action, the comprehension system rapidly availed itself of stereotypical thematic role knowledge in interpreting the utterance. When the verb rather uniquely identified another agent of a depicted action, verb-mediated depicted events in the scene also lead to rapid thematic interpretation of the utterance. Importantly, the inferential analyses revealed that while there was a clear interaction during the ADV time region for the unique identification conditions (see Fig. 5.5), there was no main effect of target agent or target type. The absence of a main effect provides strong support for the view that both informational sources were used equally well, i.e., that there was no bias towards one of the two available informational sources in the unique identification conditions. While we found clear early disambiguation in the unique identification conditions, the ambiguous identification conditions demonstrate no such early disambiguation. In the face of competition, when both thematic role relations established by stored knowledge and thematic

role relations established by depicted events were identified as relevant, people have a clear preference for relying on thematic relations established by depicted events.

Taken together, our findings provide strong evidence that the comprehension system relies in preference on thematic role relations established by depicted events rather than on thematic relations established by stored stereotypical knowledge in interpreting an unfolding utterance that relates to the immediate scene. The generality of our findings for other comprehension situations is an open matter (see section 6.2.4 for more extended discussion).

Chapter 6

Towards an Account of On-Line Comprehension in Visual Scenes

Numerous studies have investigated language comprehension in situations when language is not about the immediate scene. In contrast, there has been comparatively little investigation of comprehension when language relates to the immediate scene, and when the scene is relevant for comprehension. Based on the suggestion that a complete account of on-line sentence comprehension must ultimately be able to describe comprehension in both of these situations, the present thesis has set out to explore the little-investigated influence of depicted events on on-line sentence comprehension.

The findings that we have presented in chapters 3 to 5 reveal the great importance of depicted events for structural disambiguation and incremental thematic role assignment. Our investigation was motivated by limited but convincing evidence making clear that visual scene information can influence the incremental interpretation of an utterance. In particular, Tanenhaus et al. (1995) have demonstrated that a visual referential context influences the incremental structuring of an utterance.

On the basis of their insights and those of others (e.g., Bergen et al., 2003; Zelinsky & Murphy, 2000) we identified two research questions the examination of which would enable us to more fully characterize the influence of scene information on comprehension and the interplay between comprehension processes and attention in visual scenes.

One question was whether findings from Tanenhaus et al. (1995) generalize to other types of scene information such as depicted actions and events. Can depicted events *rapidly* influence sentence comprehension processes such as structural disambiguation and incremental thematic role assignment on-line? What is the time-course of their influence?

A second issue that we examined was the precise nature of the interplay between ongoing visual perception and comprehension processes. Is this interplay of a highly coordinated nature (see sections 1.3 and 2.5) or more loosely coupled? More precisely, is the influence of depicted events constrained by identification of scene events through words in the utterance? Furthermore, what is the relative importance of depicted events in comparison with stored thematic role knowledge? Determining the importance of depicted events relative to stored thematic role knowledge represents the first step towards a more fully specified account of on-line sentence comprehension in situations where the immediate environment is relevant for comprehension.

In what follows, we first summarize and discuss the key findings of our experiments. We then position them within existing experimental evidence on on-line sentence comprehension in visual scenes and outline theoretical assumptions that follow from our findings (sections 6.2.1, 6.2.2, and 6.2.3). In the course of this discussion, we explore the extent to which theories that prior research has drawn on to describe the influence of scene information (e.g., the Referential Theory of Sentence Processing), or current thematic role theories can account for our findings. Section 6.2.3 then investigates in more detail the nature of the interplay between visual perception and comprehension. In the second part, we integrate the insights that we gained from our experiments and those of others into a first approximation account of on-line sentence comprehension in visual scenes that complements existing theories on comprehension in situations where the scene is irrelevant.

6.1 SUMMARY OF EXPERIMENTAL RESULTS

Experiments 1 and 2 (chapter 3) revealed the rapid influence of depicted events on structural disambiguation and incremental thematic role assignment. Specifically, we examined the time-course with which depicted events (e.g., princess-washing-pirate, fencer-painting-princess, see Fig. 3.1) enabled structural disambiguation and incremental thematic role assignment of initially structurally ambiguous German subject-verb-object (SVO)/ object-verb-subject (OVS) sentences. An example item sentence pair is presented below for convenience (sentences 6.1 and 6.2). Once the verb had identified the relevant depicted action,

anticipatory eye-movements in the event scenes provided evidence for expectations of a patient (the pirate) and agent role filler (the fencer) for initially ambiguous SVO (6.1) and OVS (6.2) sentences respectively. Experiment 2 replicated the early visual disambiguation effect in the presence of distractor objects and investigated more fully the potential influence of intonation.

(6.1) Die Prinzessin wäscht offensichtlich den Pirat (obj).

'The princess (amb./subj) washes apparently the pirate (obj)'.

(6.2) Die Prinzessin malt offensichtlich der Fechter (subj).

'The princess (amb./obj) paints apparently the fencer (subj)'.

The rapid verb-mediated influence of the depicted actions on on-line sentence comprehension importantly revealed a tight coupling of visual and linguistic processing. We suggest the verb initiated a visual search for an appropriate action in the scene. The action in turn established thematic role relations that became available for comprehension when the verb directed attention to the action. The early appearance of the action verb in the utterance and the fact that it uniquely identified one of the two actions ensured that the depicted events exerted a highly rapid and strong effect on incremental interpretation.

Experiments 3 and 4 (chapter 4) investigated the time-course of disambiguation through depicted events when only soft linguistic cues (rather than the verb) could be used to identify depicted events for early incremental thematic role assignment. They further extended and generalized findings from the first two experiments to two other sentence constructions and another language (English). Experiment 3 examined comprehension of structurally ambiguous German verb-final active/passive sentences (see examples 6.3 and 6.4).

(6.3) Die Prinzessin wird sogleich den Pirat waschen.

'The princess (amb./agent) will soon the pirate (patient) wash'.

(6.4) Die Prinzessin wird soeben von dem Fechter gemalt.

'The princess (amb./patient) is currently by the fencer (agent) painted'.

Findings from Experiment 3 demonstrated that even when the main verb was sentence-final and did not establish early reference to the depicted events, linguistic cues that appeared prior to the verb still enabled disambiguation. This was reflected by eye-movement

patterns both during ambiguity resolution at the case-marked determiner/preposition of the second argument, and before this point as a result of the influence of temporal adverbs biasing towards the active or passive event.

Crucially, this shows that the rapid influence of depicted agent-action-patient events on on-line utterance comprehension does not depend upon reference by the main verb. Even when the verb did not make the depicted actions available for *early* disambiguation and incremental role assignment, soft adverbial cues were sufficient to make the relevant scene information accessible. The finding of early disambiguation from Experiment 3 further supports an account of our data in terms of thematic role assignment, since the depicted event scenes were used for early thematic role assignment before the verb had referred to the depicted actions.

The active influence of depicted events on incremental thematic role assignment (see Experiments 1 to 3) importantly generalizes to another language and sentence construction as revealed by findings from Experiment 4. Results from Experiment 4a demonstrated the early verb-mediated influence of depicted events for on-line disambiguation of the English main verb/reduced relative ambiguity (see sentences 6.5 and 6.6), thus extending and generalizing findings from Experiments 1 and 2. We observed a higher percentage of anticipatory eye-movements to the available patient (the cellist) in comparison with the agent (the fencer) for main verb sentences, and an increased proportion of looks to the agent (the fencer) in comparison with the patient (the cellist) for reduced relative clause sentences shortly after the verb (*splashed/sketched*) had been encountered. For this ambiguity, the first noun phrase was ambiguous as to its grammatical role. The observed anticipatory eye-movements to the available depicted patient (the cellist) or agent (the fencer) shortly after the verb (*splashed/sketched*) had been encountered thus reflected early resolution of thematic role ambiguity.

(6.5) The ballerina (amb.) splashed apparently the cellist (patient) in the white shirt.

(6.6) The ballerina (amb.) sketched apparently by the fencer (agent) splashed the cellist.

Experiment 4b further explored whether findings from Experiment 3 also generalize to another language and construction. Here, we investigated the influence of auxiliaries that biased towards either an active future main clause or a passive progressive reduced relative construction (see sentences 6.7 and 6.8). Analyses of the post-adverbial verb region for

the *will/being* conditions (6.7 and 6.8) revealed clear disambiguation. Finding evidence for disambiguation on the verb region itself demonstrates that the auxiliaries did make the depicted events available for comprehension since without such pre-verbal auxiliary cues disambiguation occurred only during the time region *after* the verb (see Experiments 1, 2, and Experiment 4a). In Experiment 4, we presented furthermore evidence that suggests that verb-mediated depicted events had a stronger influence on incremental thematic role assignment and structural disambiguation than purely functional auxiliaries. This was revealed by a direct comparison of the post-verbal adverb region (*apparently*) for the main verb/reduced relative (6.5 and 6.6) and *will/being* (6.7 and 6.8) conditions. While there was evidence for disambiguation during the adverb time region when the *verb* preceded the adverb (6.5 and 6.6), no such evidence was found when the auxiliaries preceded the adverb region (6.7 and 6.8). Importantly, however, the finding of early disambiguation during the verb itself after preceding auxiliary cues provides strong evidence for the claim that such soft linguistic cues rapidly make available depicted events for comprehension.

(6.7) The ballerina (amb.) will apparently splash the cellist (patient) in the white shirt.

(6.8) The ballerina (amb.) being apparently sketched by the fencer (agent) splashed the cellist.

Taken together, findings from Experiments 3 and 4 provide support for the claim that depicted events play an active role in incremental thematic role assignment even when listeners had only subtle linguistic cues for early identification of the events.

Experiment 5 (chapter 5) further investigated the importance of depicted events by directly comparing their influence on incremental thematic role assignment to that of stored thematic role knowledge. It has been found that these two informational sources each affect incremental thematic role assignment on-line. Prior research by Kamide, Scheepers, and Altmann (2003) demonstrated the rapid influence of stored thematic role knowledge on incremental thematic role assignment. The experimental findings presented in Experiments 1 to 4 in the present thesis indicate, in contrast, great importance of depicted event scenes for processes of incremental thematic role assignment.

Within a single study we examined two research questions. The first aim was to see whether we could replicate that stored knowledge of thematic role relations (that were not depicted) and immediately depicted events (that were non-stereotypical) each allow rapid thematic role interpretation. In this case, the utterance *uniquely* identified

either a stereotypical agent or an agent of a depicted event. An example scene showed a wizard, a pilot and a detective who is serving food. When people had heard 'The pilot (object/patient) serves food to...' (6.9), perception of the detective who was depicted as serving food to the pilot enabled early incremental thematic interpretation of the detective as the agent of the food-serving action during the adverb region (no other agent in the scene was a plausible agent for a food-serving action, or depicted as serving food). In contrast, after people had heard the beginning of sentence 6.10 (up to and including the verb) we found evidence on the post-verbal adverb region that the wizard was identified as the most plausible agent on the basis of stored thematic role knowledge (no other agent in the scene was a plausible agent for a jinxing action or depicted as performing such an action). This was revealed by a higher proportion of anticipatory eye-movements to the stereotypical agent (wizard) than to the other agent when the verb was 'jinx'. In contrast, when the verb was 'serve-food-to', we found a higher percentage of inspections to the depicted agent (detective) than to the other agent.

(6.9) Den Piloten verköstigt gleich der Detektiv.

The pilot (pat.) serves-food-to soon the detective (dep. agent)

(6.10) Den Piloten verzaubert gleich der Zauberer.

The pilot (pat.) jinxes soon the wizard (stereotyp. agent)

(6.11) Den Piloten bespitzelt gleich der Zauberer.

The pilot (pat.) spies-on soon the wizard (dep. agent)

(6.12) Den Piloten bespitzelt gleich der Detektiv.

The pilot (pat.) spies-on soon the detective. (stereotyp. agent)

We contrasted this condition with a situation in which the utterance did *not determine uniquely* whether the comprehension system should rely on stored thematic role knowledge (identifying a stereotypical agent) or on depicted events (identifying an alternative, depicted agent) for thematic interpretation of the utterance (see sentences 6.11 and 6.12). Scenes depicted, for instance, a pilot, a wizard involved in a spying action, and a detective. Crucially, up to the second noun phrase, sentences were identical: 'The pilot (object/patient) spies-on...'. In this case, both stereotypical thematic role knowledge and scene events provide relevant information about thematic relations. The detective

is a stereotypical agent of a spying action (but is not depicted as performing such an action). The wizard, in contrast, is depicted as involved in a spying action but is not a stereotypical agent for a spying action. In this case, we observed a strong preference of the comprehension system to rapidly rely on depicted events over stored thematic knowledge for processes of incremental thematic role assignment. Evidence for this came from a higher proportion of anticipatory eye-movements to the depicted agent (the wizard) in comparison with the stereotypical agent (the detective) for both sentences (6.11 and 6.12) shortly after people had heard the verb. We discuss and qualify these findings further in sections 6.2.3 and 6.2.4.

To sum up, the experiments that we carried out brought the following new empirical insights:

(a) Depicted events can rapidly influence structural disambiguation and incremental thematic role assignment in German and in English when uniquely identified by a verb (Experiments 1, 2, and 4a)

(b) Soft linguistic cues can also make depicted events rapidly available for early incremental thematic role assignment and structural disambiguation (Experiments 3 and 4b)

(c) The comprehension system relies on depicted events in preference to stored thematic relations in incremental thematic interpretation when the verb identifies both informational sources as relevant (Experiment 5)

6.2 THEORETICAL ASSUMPTIONS

Having summarized the key empirical insights of our experiments, we now relate them to prior research. Sections 6.2.1, 6.2.2, and 6.2.3 discuss in more detail the influence of depicted events on structural disambiguation and on incremental thematic role assignment and outline key theoretical assumptions that follow from our findings.

6.2.1 EXPLICITLY DEPICTED RELATIONS: BEYOND REFERENTIAL CONTEXT

In situating our findings with respect to prior research, one important question is the extent to which our findings add to what has previously been known. As outlined in section 2.2, despite convincing evidence for the effects of scene information on comprehension prior to

our investigation, existing findings were limited to the insight that scenes which establish a referential context can affect on-line comprehension. To explore the factors that were responsible for the influence of scene information, we analysed prior studies that had found a rapid incremental influence of scene information on comprehension. Based on this analysis we came to the conclusion that one factor that enables scene information to inform comprehension processes such as structural disambiguation and semantic interpretation is its ability to establish relations between scene entities (see section 2.3).

Specifically, our analysis provided support for the claim that scene information influenced on-line comprehension because perceivers established important relations between entities. One such relation between entities is referential contrast between objects. We suggested that contrastive comparison of two same-category objects that differ in one property can enable people to establish unique reference to an entity (Sedivy et al., 1999; Spivey et al., 2002; Tanenhaus et al., 1995). It should be note that this is our interpretation of the findings by Tanenhaus et al. (1995) which we discussed in more detail in section 2.3. Perceived contrast between the properties of two objects is one example of how people can establish (non-linguistic) relations between otherwise similar entities through comparison and interpretation of the scene. Recall that we proposed that a further way in which the scene can establish relations between entities is through entity affordances (e.g., a carrousel affords a riding action) (e.g., Kamide, Altmann, & Haywood, 2003). Such contrastive and affordance relations arise through comparison and interpretation by a perceiver, rather than being overtly realized in the scene, and hence represent an *implicit* relationship between entities.

In contrast, in the experiments that we presented in this thesis it was *explicitly* depicted event relations between entities that informed language comprehension processes such as the structuring of an utterance and incremental thematic role assignment. Our findings thus provide support for the view that agent-action-patient events are a further type of relation between entities that influences comprehension. Here, the perceived relations are explicitly realized visually by the depicted actions. Our results demonstrate that such explicit visual event relations influence interpretation of the linguistic input, and hence substantially extend the findings by Kamide, Altmann, and Haywood (2003), Sedivy et al. (1999) and Tanenhaus et al. (1995). In building upon their and our own work, we arrive at the view that perception of non-linguistic information – such as contrast, affordances, actions, or events - that establishes relevant relations between entities, can affect how linguistic input is interpreted. Emphasizing the importance of relations between

entities makes it possible to account for the effects of depicted event relations as well as visual referential contrast and affordance relations on the incremental interpretation of an utterance.

While an explanation in terms of relations between entities accounts for both our own findings and those of prior research on the influence of scene information, we must emphasize that the type of relational information proffered by our scenes differs in important respects from the type of object relations provided by scenes in prior research (e.g., Kamide, Scheepers, & Altmann, 2003; Sedivy et al., 1999; Tanenhaus et al., 1995). Scenes in their experiments essentially provide a *referential* context since they only contain entities (objects, animals, and events) with no explicitly depicted relations between them. In contrast to scenes that were used in their studies, scenes in our experiments provided propositional information in addition to referential information. We suggest that as a result the context in our scenes has a more "active" part in determining the interpretation and structuring of an utterance.

The main conclusion that we draw from the above discussion is the following: As a consequence, theories of on-line sentence comprehension must explicitly include an account of how the immediate scene context provides both referential and propositional information such as events. Previous studies (e.g., Sedivy et al., 1999; Tanenhaus et al., 1995) have drawn upon the Referential Theory of Sentence Processing (Altmann & Steedman, 1988; Crain & Steedman, 1985) as a framework for describing comprehension processes in visual contexts. Such a theory provides a suitable framework for the description of referential processes in sentence comprehension when reference is limited to the ontological category of entities (see section 2.2.2). Our findings, however, suggest that on-line comprehension mechanisms exploit a wider range of ontological categories than just entities, including actions and events. In order to describe how depicted actions and events, and other relational scene information are exploited for on-line linguistic disambiguation, we need to establish reference to them through, for example, verbs and sentences. Theories of on-line sentence processing therefore require a suitably rich array of ontological categories and corresponding referential expressions, as well as mental representations for both.

6.2.2 THEMATIC ROLE ASSIGNMENT THROUGH SCENE EVENTS

By highlighting the notion of 'relations between entities', we succeeded in providing an account both for the effects of depicted events and of implicit scene relations on on-

line sentence comprehension. We further made clear that an account of on-line sentence comprehension requires a model of how the visual environment can establish event relations of who-does-what-to-whom in addition to providing referential information.

While the Referential Theory does not provide such a model of the visual environment, recall that the thematic role theories reviewed in section 2.1.2 do provide an inventory of conceptual representations including thematic agent/patient roles and representations for events. Since they provide such a rich inventory as is required for describing our findings, we investigate in the following whether thematic role theories are a more appropriate basis in accounting for our findings than the Referential Theory of Sentence Processing (see section 6.2.1). It should be noted, however, that while for our findings thematic role theories may provide a more appropriate account than the Referential Theory, they alone cannot straightforwardly explain the findings by Tanenhaus et al. (1995), which still require a model of referential context.

In the following, we consider whether thematic role theories that account for findings on thematic role assignment by Altmann and Kamide (1999) and Kamide, Scheepers, and Altmann (2003) can also account for our findings of incremental thematic role assignment through depicted events. To investigate the extent to which current thematic role theories can describe our findings we specifically contrast the origin and type of information that enabled incremental thematic role assignment in our studies with studies by Kamide, Scheepers, and Altmann (2003). This is a first step in evaluating whether existing theories of thematic roles can account for our findings, since in describing the on-line influence of depicted events they will have to represent both the origin (the scene) and the type of information (depicted events).

Experimental research on the influence of *stored knowledge* on comprehension has demonstrated that stored thematic role knowledge that is associated with specific events can influence incremental thematic role assignment. Prior research on incremental thematic role assignment has, for instance, been carried out by Kamide, Scheepers, and Altmann (2003) (recall section 2.1.1 for more thorough discussion). Their experiments show that anticipatory eye-movements to role fillers in the scene were influenced by morpho-syntactic case-marking and experience-based knowledge about the arguments a verb typically requires that could be extracted at the verb (e.g., Ferretti et al., 2001). Furthermore, investigation of the English active/passive ambiguity by Kamide, Scheepers, and Altmann (2003) has shown that *a combination of voice information, the verbs' semantic constraints, and real-world knowledge* disambiguates an initially role-ambiguous noun phrase (e.g., *The*

hare will eat.../The hare will be eaten...) (Kamide, Scheepers, & Altmann, 2003, p. 51). Crucially, the cue for thematic role assignment in their studies was stored linguistic and world knowledge provided by the linguistic context alone, while scenes provided only a relevant referential context (for a discussion of the possibility that a referential context can provide stored thematic role knowledge see section 2.3).

In contrast, for Experiments 1 and 2, as well as for the main verb/reduced relative study presented in this thesis (Experiment 4), neither the linguistic context *alone* nor lexically-specific stored thematic role knowledge associated with the verb enabled early disambiguation. Rather, once the verb had been processed, it was purely situation-specific *depicted* role information in the event scenes, and verb-mediated identification of the appropriate event, that provided the information about a character's role necessary for early disambiguation. Further, such thematic role information provided by depicted events even enabled early incremental thematic role assignment when not the verb (a strong cue), but only soft linguistic cues could be used to identify the relevant event early in the utterance (Experiments 3 and 4b). This suggests a highly active influence of depicted events on incremental thematic role assignment. Experiment 5 moreover provided evidence that depicted thematic role-relations informed incremental thematic role interpretation in preference to stored thematic relations when the verb identified both a stereotypical agent and an agent of a depicted event as relevant for comprehension. Our findings thus demonstrate the great importance of depicted events in a scene for the incremental assignment of thematic roles (for further qualification of these findings see section 6.2.4).

The discussion of thematic role theories in section 2.1.2 made clear that a theory of thematic role assignment *in comprehension* should prioritize different theoretical assumptions in comparison with a theory that aims to describe thematic roles from a *linguistic* (e.g., Dowty, 1991) or *formal semantics* viewpoint (e.g., Davidson, 1967, 1986b). This is largely owing to the different goals that these theories pursue. The purpose of a theory of incremental thematic role assignment in comprehension is to account for how thematic roles are assigned (e.g., the kinds of informational sources, processes, and mechanisms that enable thematic role assignment). As a consequence, such a theory should be developed on the basis of experimental evidence on incremental thematic role assignment (see the discussion in section 2.1.2). We drew attention to the fact that among the theories that we reviewed in section 2.1.2, the account of thematic role assignment by McRae et al. (1997), which is broadly compatible with theories of thematic roles such as Fillmore (1985), fulfills the requirement of having been developed on the basis of experimental findings. We

pointed out that their account of thematic role assignment appears to be well suited to accounting for findings by Kamide, Scheepers, and Altmann (2003).

Since our findings present new experimental evidence concerning the kinds of information that can influence incremental thematic role assignment, let us re-consider theories of thematic roles and thematic role assignment. Thematic relations that are provided by the immediate environment have received little consideration in theories of thematic roles such as Dowty (1991), Fillmore (1985), and McRae et al. (1997). These approaches describe thematic roles in isolation from information that is present in visual scenes and provide no distinct mental representations for actions and events that derive from an immediate visual scene. Our discussion is based on the assumption that we might (at least sometimes) require distinct mental representations for thematic relations derived from the immediate scene as opposed to thematic knowledge from our knowledge base (see Jackendoff, 1983, 2002). We suggest that this is important in particular when the scene contains information that is not part of our linguistic and world knowledge, and that we have to acquire (e.g., in language acquisition when small children acquire new concepts for entities and events that they do not yet know). As a result of their tacit assumption that thematic roles can be sufficiently characterized in terms of experience-based *knowledge* and linguistic conceptual representations (Fillmore, 1985; McRae et al., 1997) or abstract linguistic generalizations (e.g., Dowty, 1991), we suggest that these theories cannot in their present state fully account for our findings.

It should be noted, however, that some semantic theories provide a model of a situation or world. In formal semantic theories such as Davidson (1986a, 1986b) or related approaches (e.g., Parsons, 1990), language is explicitly related to a model of the world. Information about entities and events in the scene might thus be available from the model. Moreover, this type of theory provides an inventory of mental representations for ontological categories such as actions and events (see Davidson, 1986b) (for critical discussion see Jackendoff (1983, 2002)). However, these theories only explicitly relate a sentence to a model of the world through a referential relation that evaluates a sentence as true or false with respect to a given situation model (see section 2.1.2). The immediate scene has thus a relatively "passive" role, and only provides a referential context against which sentences can be evaluated. We suggest that the scene in this type of theory cannot contribute in a more active way towards deriving a full semantic interpretation of a sentence. As a result, such a theory cannot serve as a basis for describing our findings of the *active* influence of depicted events in determining the structuring and interpretation of a sentence.

There are further implications for theories of thematic role assignment. From the perspective of *language comprehension* all of the above theories lack a consideration of language and of the immediate scene as two types of information that are *processed incrementally* by *distinct cognitive systems* such as the comprehension and visual perception systems. To account for the incremental nature with which depicted events contributed to structural disambiguation and incremental thematic role assignment in our experiments, we require a framework that explicitly includes other cognitive systems such as visual perception. Furthermore, an account of on-line language comprehension requires processes (and ultimately mechanisms) to describe precisely how thematic role assignment and structural disambiguation take place.

6.2.3 COORDINATED INTERPLAY & RELATIVE PRIORITY OF SCENE EVENTS

To develop a more detailed sketch of the processes that take place in incrementally assigning thematic roles through depicted events, the present section discusses the interplay between incremental utterance comprehension and scene perception. In Section 2.5, we identified two dimensions along which we might specify the nature of the interplay between visual perception and utterance comprehension. First, we asked whether the interplay between these distinct cognitive processes was temporally coordinated. Second, we set out to explore the relative importance of stored thematic role knowledge and depicted scene events. Based on findings from Experiments 1 to 5 of the present thesis, we now specify in more detail the nature of comprehension in visual scenes along the above two dimensions. In the course of this discussion, we provide a first sketch a of procedural account for our findings which we will refine in section 6.3.2. Recall to this end that we have associated existing lines of research in psycholinguistics with different positions on how comprehension proceeds in visual scenes (see section 2.5). We review these positions with respect to the findings that we have presented. We emphasize once more, however, that while it is possible to associate existing lines of research with positions on the nature of comprehension in visual scenes, none of the reviewed studies/accounts have directly investigated, for instance, the relative importance of depicted events. They furthermore do not provide an explicit and specified account of how comprehension in visual scenes proceeds.

Research carried out by Altmann and colleagues (e.g., Altmann & Kamide, 1999; Kamide, Scheepers, & Altmann, 2003) would appear to be best associated with a great importance of stored thematic role knowledge in comprehension. In section 2.5 we asso-

ciated research by Tanenhaus et al. (1995) with a great importance of the influence of scene information on comprehension, highlighting that the influence of scene information on the structuring of the utterance did not appear to be closely linked to incremental utterance comprehension in their studies. As an alternative to this position, we sketched an approach that views scene information as playing an important role, however that assumes its influence may be closely temporally coordinated with utterance comprehension. Finally, a further account that we identified were constraint-based interactionist models (e.g., Tanenhaus et al., 2000). This type of model would appear to expect no preference for either depicted events or stored thematic role knowledge. Rather it would predict comprehension difficulty and high competition between these two informational sources provided they are comparable in strength. Notice moreover that it is not clear what a Jackendovian framework would predict since it does not provide a specified account of on-line language comprehension at all.

Let us first consider the temporal coordination of the interplay between comprehension and attention in visual scenes. An analysis of previous studies on utterance comprehension has suggested that there is a close time-lock between the comprehension of a *word* in the utterance and inspection of the relevant object(s) (see section 2.2.1). Tanenhaus and colleagues observed a close time-lock between when a word referred to an object (e.g., *apple*, *towel*), and when people inspected that object (i.e., shortly after the word had referred to it) (see also Cooper, 1974). We suggested that while the eye-gaze pattern in their studies reveal such a close time-lock at word level and for referential processing, it is less obvious that gaze-patterns reveal a close time-lock between utterance comprehension, scene perception, and the *precise time* when the comprehension system decided between the alternative ways of *structuring* the utterance.

To illustrate this argumentation, re-consider findings by Tanenhaus et al. (1995) (see sections 1.1 and 2.2). Tanenhaus et al. (1995) demonstrated that the *type of visual referential context* biases people to either adopt a destination (one-referent context) or a location interpretation (two-referent context) of a temporarily structurally ambiguous phrase. We specifically suggest that in their studies, upon inspecting the two-referent scene, people became aware that there were two same-category objects (two apples) in the scene. As a result of having established this type of referential context, the parsimonious way of structuring an initially ambiguous utterance such as *Put the apple on the towel ...* was to attach the prepositional phrase as a modifier to the initial noun phrase. Importantly, the referential context that informed the decision of how to structure the utterance was

presumably already established *prior* to utterance onset. This account of the findings by Tanenhaus et al. (1995) receives support from the observation that the pattern of eye-gaze between the two types of contexts differ already at the very start of the utterance when people had heard no more than *Put the apple*. While eye-movements at this point in time oscillated between the two apples in the two-referent condition, listeners in the one-referent condition immediately inspected the only available apple.

For our investigation, in contrast, eye-gaze patterns clearly reveal the precise point in time when the scene events influenced comprehension. Consider first findings from Experiments 1, 2, and 4a when the verb uniquely identified the action for early structural disambiguation and incremental thematic role assignment. What differed between conditions (e.g., SVO and OVS) in our study was not the type of scene but rather the meaning of the verb in the utterance. The action that the verb identified was accordingly a different one for the preferred (e.g., SVO) in comparison with the disfavoured (e.g., OVS) structural analysis of a sentence. Eye-movement patterns in Experiments 1, 2, and 4a did not differ between the SVO and OVS condition prior to the offset of the verb. It was only when people had heard the verb that the *relevant* of the two available actions was identified and that the thematic relations associated with the action became available. The thematic role relations then in turn rapidly constrained interpretation of the utterance. Findings from Experiments 1, 2, and 4 are thus compatible with an account that in particular prioritizes the important function of referential expressions such as verbs in making depicted actions available for comprehension.

However, if the interplay between comprehension and attention in scenes were indeed based upon direct referential processing alone, then we would expect that reference from verbs to the depicted actions is a necessary pre-requisite for the influence of depicted events on on-line comprehension processes. Findings from Experiments 3 and 4 show, crucially, that this is not the case. They demonstrate that depicted events enabled on-line thematic role assignment and structural disambiguation before/while people encountered the lexical verb. In this case the utterance contained soft cues (temporal adverbs/aspectual auxiliaries) that listeners could use for identifying which of the two available scene events was relevant for comprehension of the utterance. Since these cues do not directly refer to an action, the finding that they can make the depicted events accessible must be explained in other ways than through direct verb-action reference. We assume that these soft linguistic cues are associated with general event structures, and that this knowledge (rather than verb meaning) identifies the relevant depicted event.

In the studies by Tanenhaus and colleagues such utterance-mediated identification of relevant scene information does not appear to determine the time when the choice about the structuring of the utterance could be made (see discussion above). While their findings reveal a close time-lock between the mention of a *word* and the time when listeners establish reference to relevant scene objects, they do not reveal a *closely time-locked* reciprocal influence of the perceived scene in determining the *structuring* of the utterance. Our findings thus importantly extend findings by Tanenhaus et al. (1995) in revealing a closely time-locked, mutually-constraining interplay between scene perception and the incremental structuring of an utterance. The close coordination of the time when structural disambiguation sets in with the time when the verb identifies the action and its associated thematic relations allows us to exclude a procedural account of our findings in which the timing of the influence of scene events is independent of the time when the utterance identifies them (e.g., Tanenhaus et al., 1995) (see also the discussion in section 2.5).

Let us now in more detail consider the extent to which our findings are compatible with the other lines of research that we briefly recapitulated in the introductory paragraphs to this section (for a more detailed discussion see section 2.5). Findings from Experiments 1 to 4 seem to be broadly compatible with all of these lines of research/accounts. They essentially reflect a rapid incremental interaction of diverse constraints, a finding that would appear to broadly fit with interactionist accounts, the research pursued by Altmann and Kamide (1999), and the studies by Tanenhaus et al. (1995). The important insight that our results add to insights from these lines of research is the observation that there is - at least for depicted events - a close temporal coordination between the time when scene information is identified and when it actively informs comprehension.

In contrast to findings from Experiments 1 to 4, results from Experiment 5 (i.e., the preferred reliance on depicted events as opposed to stored thematic role knowledge for incremental thematic interpretation) does not appear to be straightforwardly accounted for by existing research. The relative priority of depicted events during comprehension cannot, for instance, be fully explained by interactionist models such as the one proposed by Tanenhaus et al. (2000). As detailed above, this type of model would expect processing difficulty if constraints that are comparable in strength compete (see section 2.5). If constraints that are similar in strength compete, the model would predict that none of the competing constraints can be *rapidly* applied in comprehension. Recall that our findings showed that the competing constraints - stored thematic role knowledge and depicted events - were comparable in strength (see the inferential analyses and discussion in

chapter 5). Importantly, analyses did not reveal any difference in the strength of these two informational sources when they were uniquely identified by the utterance (i.e., there was no main effect of looks to the depicted in comparison with the stereotypical agent during early disambiguation in the unique identification condition). Given constraints that are comparable in strength, constraint-based models would appear to predict that we should not observe a substantially higher proportion of inspections to any individual scene object (assuming these looks indicate that we have successfully applied an informational source in comprehension). Thus, if comprehension in visual scenes proceeded in the way suggested by constraint-based models, then we should not have observed the rapid preferred use of depicted events that was revealed by the anticipatory eye-movements to the depicted in comparison with the stereotypical agent. As a result, we argue that this type of model is not an appropriate model in accounting for our findings.

Let us briefly re-capitulate. So far, we have discussed two of the four accounts/lines of research that we reviewed further above (see section 2.5 and the beginning of the present section). The first was an account emphasizing the importance of scene information, albeit independent of the guiding force of the utterance. The second was a constraint-based interactionist account. A third account that we introduced in chapter 2 was an approach that prioritized linguistic and world knowledge in constructing mental representations of the unfolding utterance. While broadly compatible with findings from Experiments 1 to 4, it is not clear whether this account can explain findings from Experiment 5. We re-consider and further explore this issue when developing a more specific sketch of on-line sentence comprehension in visual scenes in section 6.3.2.

What the above analysis makes clear is that while Experiments 1 to 4 can be explained by a range of frameworks and lines of research, findings from Experiment 5 are potentially not explained by existing accounts of on-line sentence comprehension in psycholinguistic research. (see section 6.3.2). We propose that in accounting for findings from Experiment 5 the first step might be to identify the factor(s) that caused the preferred reliance of the comprehension system on depicted events in comparison with stored thematic role knowledge. The origin of the preferential reliance on scene events might, for instance, derive from developmental and/or evolutionary comprehension strategies. The rapid impact of explicitly depicted thematic relations between entities identifies the comprehension system to be highly adapted towards acquiring new information from its environment rather than always relying on linguistic and world knowledge.

Such a proposal is compatible with theories of language acquisition such as Gleitman (1990). In her account of language acquisition, Gleitman argues that a child can extract event structure from the world around it. When the child perceives an event, the structural information it extracts from it can determine how the child interprets a sentence that describes the event in question. The interpretation of a sentence can in turn direct the child's attention within the visual environment. The fact that the child can draw on two informational sources (sentence and scene) enables it to infer information that it has not yet acquired from what it already knows. Assume, for instance, the child perceives a girl throwing a ball, and hears the sentence *The girl is throwing the ball*. Assume further, the child knows the word *girl*. Even if it does not know what *throwing* means, perception of the event can enable it to deduce the meaning of the unknown verb it just heard. It can further identify *girl* as the thing performing the throwing-action.

Such a developmental account offers one explanation for why the comprehension system relies in preference on depicted events over stored thematic role knowledge. Indeed, the observed priority of depicted events may have developed in the course of language acquisition. There is furthermore clearly an important interplay between the function of language in guiding attention in the immediate scene during language acquisition and the influence of scene information on language understanding. The developmental account that Gleitman proposes motivates thus not only the relative priority of depicted events, but also the close temporal coordination between utterance comprehension and the use of scene information during comprehension that we observed in our experiments. We suggest this interplay of utterance comprehension and scene perception during language *acquisition* may fundamentally shape the architecture and mechanisms of language *comprehension*.

6.2.4 CAVEATS

The various experimental situations investigating the interplay between visual processing and comprehension that we have discussed share one fundamental aspect: the utterances are about the immediate visual scene. In language acquisition, when parents talk to their children, language is likely often about the immediate scene which children typically explore. In these situations, it serves a specific function, namely making the immediate scene accessible for the child, and identifying objects that it perceives (see, e.g., Richards & Goldfarb, 1986; Roy & Pentland, 2002, for related research in modelling).

During adult language comprehension, however, this is not true to the same extent. Language often sub-serves other functions or tasks and is only used to refer to entities in the immediate scene for part of the communication we engage in. Much of our day is spent talking, reading, or writing about things that are not immediately present. Examples that come to mind are, for instance, the expression of abstract ideas or thoughts, or narration of past events. As a result of this observation, we deem it necessary to qualify our findings regarding the comprehension situation in visual scenes.

We acknowledge that in situations where the utterance does not directly refer to the immediate visual environment, depicted events will almost certainly not have the importance that is suggested by findings from Experiment 5 in the present thesis since the scene is irrelevant. It is the immediate presence and relevance of both utterance and scene for comprehension which enables the rapid interplay between these two informational sources. We do expect, however, that in situations where the utterance is about the immediate environment, our findings of the priority of depicted events over stored thematic role knowledge in thematic interpretation will hold true. We moreover propose that while language is often not about the immediate scene in adult life, we have spent a substantial part of our lives acquiring language. We suggest that this period may indeed have shaped both our cognitive architecture (i.e., providing for rapid interaction between cognitive systems such as language and vision), and comprehension mechanisms (e.g., we rapidly avail ourselves of information from the immediate scene when the utterance identifies it).

6.3 AN ACCOUNT OF ON-LINE SENTENCE COMPREHENSION

The preceding discussion has outlined that neither the Referential Theory of Sentence Processing (see section 6.2.1), nor the formal theories of thematic role and thematic role assignment that we reviewed (section 6.2.2) can fully account for our findings. Furthermore, most current psycholinguistic theories of on-line language comprehension provide only a limited account of the findings that we presented in chapters 3 to 5. This is owing to their assumption that the comprehension system can be investigated and characterized in isolation from other perceptual systems and from the immediate environment. Neither restricted theories (e.g., Crocker, 1996; Forster, 1979; Frazier & Fodor, 1979; Frazier, 1987; Kimball, 1973; Mitchell et al., 1995) nor unrestricted psycholinguistic accounts of on-line sentence comprehension (e.g., MacDonald et al., 1994; McRae et al., 1998; Tanenhaus et al., 2000; Trueswell et al., 1993, 1994; Trueswell, 1996) have explicitly included the imme-

diate visual environment as an informational source in their account of comprehension (see chapter 1). Admittedly, these accounts have been developed with the goal of describing on-line language comprehension in situations when language is not about the immediate environment. In such situations (e.g., reading of isolated sentences), the immediate environment will likely play a less important role since it is irrelevant for comprehension.

We suggest, however, that a complete account (and eventually theory) of on-line sentence comprehension should be able to account for on-line language comprehension both when language is and when it is not about the immediate scene (on the assumption that the same cognitive mechanisms are engaged in both situations). We complement existing research on comprehension in situations where the scene is irrelevant with an account of how comprehension proceeds when the scene is relevant for comprehension of the utterance (see section 6.2.3). The great role that the interplay between perception of the immediate scene and comprehension plays during language acquisition motivates the preferred reliance of the comprehension system on depicted events.

At the same time, drawing on language acquisition to motivate this preference still leaves open the possibility that - in adult life - the immediate scene is not always relevant for comprehension and that language may have other functions in addition to relating to the immediate scene. We thus view our research as adding to current psycholinguistic theories that have been developed on the basis of experimental findings from comprehension during reading. With this goal in mind, we briefly summarize the theoretical assumptions that we have identified as necessary in accounting for language comprehension both when language is and when it is not relevant to the immediate scene (see sections 6.2.1, 6.2.2, and 6.2.3). Such an account requires:

(a) A model of the immediate visual environment that may contain ontological categories such as entities, actions, and events

(b) A framework that explicitly relates the comprehension system to the visual perceptual system

(c) A linguistic framework that allows us to characterize stored linguistic and world knowledge

(d) An inventory of mental representations for both stored linguistic/world knowledge and scene-derived ontological categories such as entities, actions, and events

(e) A set of processes (and ultimately mechanisms) that describe the coordinated interplay between visual perception and comprehension

One framework that fulfills many of these requirements ((a) to (d)) is Jackendoff (2002). While we suggested that the Jackendovian framework is not best associated with findings of the preferred reliance on depicted events, it is not entirely incompatible with such a finding either. This is due to the fact that our findings reveal insights about the *procedural* nature of comprehension. The Jackendovian framework, however, is underspecified with respect to comprehension processes and mechanisms and might be combined with a procedural account (see section 6.3.2). While not obviously "predicting" findings from Experiment 5, his framework has many advantages (see also, e.g., Garrod & Pickering, 2003, for further discussion of the advantages of his framework concerning other recent accounts of language processing).

One important characteristic of his framework for our research is that Jackendoff has begun to explicitly position the comprehension system in relation to the visual perceptual system. As a linguistic framework his theory permits us further to specify how phonological information is mapped onto syntax, and syntax onto semantics. Jackendoff's semantic theory straightforwardly includes actions, events, and other ontological categories under the heading of reference. This follows from his assumption that the information conveyed by language is about the *perceptual* world (or *projected* world). He defines the perceptual world *not as absolute reality, but as the 'reality' constructed by our perceptual systems in response to whatever is 'really out there'* (2002, p. 308). He shows through both linguistic and non-linguistic arguments that reference in language must include a wide range of ontological categories among them places, actions, and events, and that *the class of referring expressions includes not only NPs but also Ss, VPs, PPs, and AdvPs* (1983, p. 41). On the premise that a semantic theory should explain otherwise arbitrary generalizations about the syntax and the lexicon, he maintains further that the array of mental representations *must contain a rich range of ontological categories* that correspond to different types of projected entities (1983, p. 56).

The notion of the perceptual world situates his approach further in proximity to an embodied view of the relation between language comprehension and visual environments (e.g., Bergen et al., 2003) (but see, e.g., Spivey & Gonzalez-Marquez, 2003, for critical

discussion of the Jackendovian framework concerning research in cognitive science). In comparison with existing embodied approaches to language, Jackendoff's framework has, however, the added advantage of offering an explicit and formal specification of the architecture of the language system and of linguistic knowledge, enabling us to derive testable predictions in our future research. We discuss his framework in more detail in section 6.3.1. Section 6.3.2 will then provide a detailed procedural sketch of on-line language comprehension in visual scenes based on our findings, and further discuss issues of combining our procedural account with the Jackendovian architecture.

6.3.1 ARCHITECTURE

For the reasons outlined above we adopt Jackendoff's architecture (1990, 2002) as a formal framework for describing the processing mechanisms that underlie on-line comprehension (see Fig. 6.1). Jackendoff (2002, p. 198) proposes a parallel constraint-based processing architecture, consisting of three independent but related levels of linguistic structure: phonological, syntactic, and semantic/conceptual structure (see also the discussion in chapter 1). For each of these levels, there are constraints on processes internal to a level (*integrative*), and on processes between the levels (*interface*). An integrative and an interface processor respectively accomplish internal processes and linking between the levels.

The individual levels in Jackendoff's architecture are modular in the sense of being domain-specific (i.e., their representational vocabulary is specialized), but unlike Fodorian modularity (Fodor, 1983), linguistic structures are linked among themselves and to other cognitive sub-systems (see Fig. 6.1). The version of modularity advocated in Jackendoff hence permits incremental communication between ongoing processes of the phonological, syntactic, and conceptual systems. Importantly, it also allows incremental information exchange between conceptual structure and perception or action via interface processors (2002, pp. 220f.). Jackendoff's framework provides an interface at which conceptual structure and the upper end of the visual system (termed 'spatial structure', see Fig. 1.2) can be linked, and thus provides a means for describing how comprehension proceeds when these different types of information are available and interact.

The processing architecture Jackendoff proposes is *logically non-directional* (Jackendoff, 2002, p. 198), in principal permitting building of mental representations starting from vision or semantics and passing it on to syntax and phonology, or starting from phonology, transmitting it to syntax, semantics, and visual perception. The findings of our experi-

ments are broadly compatible with this characteristic of his architecture. Both visual and linguistic information enabled utterance interpretation while visual scenes remained the same in many of the experiments (e.g., Experiments 1 and 3): Utterance-derived mental representations of noun phrases, verbs, and soft linguistic cues served to identify scene entities and actions. Scene-derived mental representations of depicted events, in turn, enabled structural disambiguation and incremental thematic interpretation of the utterance. Taken together, these findings support such a non-directional account of the interaction between visual and linguistic information in on-line sentence processing as is proposed by Jackendoff (2002) (see also Jackendoff, 2003).

Despite being logically non-directional, the parallel architecture put forward in Jackendoff (2002) does impose certain constraints on the sequential order with which the individual levels of linguistic structure and other cognitive systems can interact. Jackendoff (2002, p. 202f) sketches an example of utterance comprehension in visual scenes that illustrates the kinds of constraints his architecture imposes: He suggests that - when we listen to an unfolding utterance - visual search singles out the correct interpretation of what we

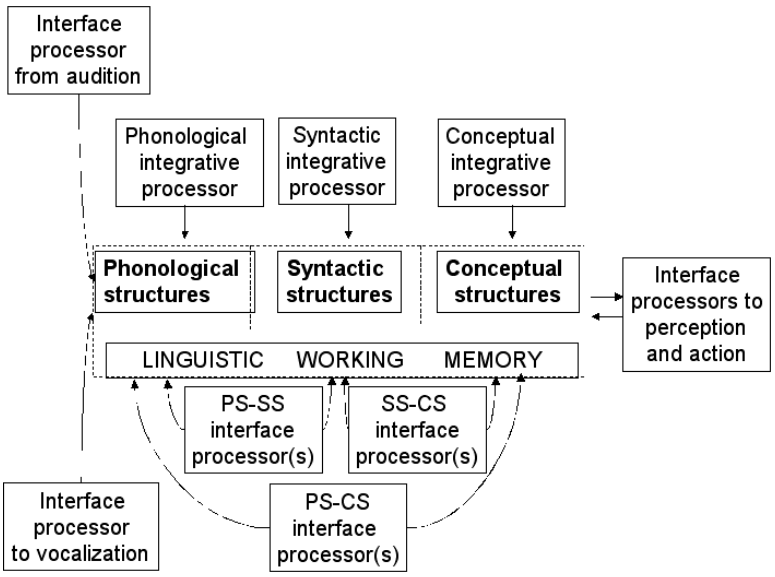


Figure 6.1: *Schematic sketch of the Jackendovian architecture (Jackendoff, 2002, p. 199)*

hear. The constraint on the influence of scene information on comprehension in his framework is that visual contexts cannot affect syntactic structure directly, but only through communication that is passed from the visual system via the visual-conceptual and the conceptual-syntactic interfaces to syntax (see Fig. 6.1). Inhibition of the incorrect alternatives is hence propagated to semantic and syntactic structure via the interfaces. While our findings are compatible with the non-directionality of Jackendoff's architecture, it is not clear whether they are compatible with or support the kinds of constraints that Jackendoff proposes for the interaction of distinct structures in his architecture. For present purposes, we adopt the Jackendovian framework, however, without definitely committing to the constraints it imposes on the sequential order with which mental representations from distinct cognitive systems can interact. We discuss this issue in more detail in section 6.3.2.

Findings from Experiments 1 to 4 in this thesis are *broadly* compatible with a Jackendovian framework. There are, however, some caveats. First, in the above-mentioned processing example from Jackendoff (2002), visual search singles out the correct interpretation of an utterance. The scene is not described as playing an active role, but rather as "filtering" interpretations proposed by the language system. It is not entirely clear whether the Jackendovian framework would accomodate cases where the scene plays a more important role, and actively proposes a thematic interpretation of the utterance through depicted events as was the case in our experiments. Second, a Jackendovian framework alone does not permit us to provide a detailed sketch of on-line comprehension in accounting for our findings since it does not specify a processing account at all. We thus develop a procedural account in the ensuing section. Finally, while findings from Experiments 1 to 4 are broadly compatible with the Jackendovian framework, we discuss in section 6.3.2 whether a Jackendovian framework can also, in principle, be combined with the procedural account required by findings from Experiment 5.

6.3.2 PROCEDURAL ACCOUNT

Jackendoff's theory provides essentially an architecture and theory of linguistic competence along with an ontological classification for various things/relations in the visual environment (e.g., objects, locations, actions and events) and relevant mental representations. His theory focuses thus primarily on architectural and representational issues (e.g., Jackendoff, 1983, 1997, 2002).

What Jackendoff does not provide is a detailed account of on-line incremental sentence comprehension. Nor does he specify the nature of the interplay between the visual perceptual system and comprehension. In section 6.2.3 we provided a first approximation sketch of the nature of this interplay. This sketch is, however, still missing a description of precisely how mental representations are constructed compositionally (see our postulate in section 1.2). Nonetheless, the preliminary sketch and discussion of section 6.2.3 allow us to draw some important conclusions concerning comprehension in visual scenes: Neither an account of comprehension that views the influence of scene information as independent from utterance comprehension, nor a constraint-based interactionist model succeed in accounting for our combined findings. Rather, we proposed that we can best explain our findings by emphasizing that utterance comprehension, attention in scenes, and the influence of scene information (depicted events) on comprehension are closely temporally coordinated, and that this mutual interplay leads to the observed priority of depicted events in comprehension.

While such an account appears to be an appropriate description our findings, and to also provide a plausible description of the nature of comprehension in visual scenes in general, we have so far omitted discussion of an alternative account that is supported by findings from numerous studies on comprehension in visual scenes (e.g., Altmann & Kamide, 1999; Chambers et al., 2002; Hanna et al., 2003; Kaiser & Trueswell, in press; Kamide, Altmann, & Haywood, 2003; Kamide, Scheepers, & Altmann, 2003; Runner et al., 2003; Sedivy et al., 1999; Sussman & Sedivy, 2003). As outlined in chapter 1, most studies in visual scenes have demonstrated how compositional interpretation of the *linguistic input* can influence interpretation of the utterance and attention in a scene. Attention in scenes in these studies is viewed as merely reflecting comprehension processes as opposed to actively contributing to comprehension. This line of research appears to be compatible with the Jackendovian theory, espousing a similar view of the role of scenes in comprehension as is sketched in Jackendoff (see the discussion above). Since the overwhelming majority of current psycholinguistic research is compatible with this account, we think it is necessary that we investigate in more detail whether such an account (i.e., that emphasizes the guiding force of the utterance and of stored linguistic/world knowledge during comprehension) might describe our findings. We henceforth refer to this account as 'Account 1'. We contrast this position with the one that we proposed in section 6.2.3 (i.e., a mutually constraining coordinated interplay between comprehension and scene perception; henceforth 'Account 2').

Despite not providing an account of the processes and mechanisms of on-line language comprehension, Jackendoff does specify some performance-related constraints on the order in which information can be passed on between distinct sub-systems in his architecture (see section 6.3.1). We thus also explore whether the account of processing that appears to best describe our findings can be integrated within the Jackendovian framework in agreement with the above-discussed performance-related constraints.

COMPOSITIONAL INTERPRETATION

In what follows, we sketch in detail how mental representations might be constructed under Accounts 1 and 2 respectively during utterance comprehension. Based on studies that examined language comprehension in the absence of visual information as well as on the basis of experiments that investigated comprehension in visual scenes, we assume that the structure and interpretation of a sentence is established incrementally as the words in an utterance are read/heard (e.g., Frazier & Fodor, 1979; Frazier, 1987; Tanenhaus et al., 1995; Trueswell et al., 1993, 1994).

In our description, we focus on the construction and integration of conceptual and spatial structures (the "upper end" of the visual perception system, see Fig. 1.2). Note that we do assume that spatial structure can represent and propose concepts such as thematic roles, an assumption that is not necessarily as clearly present in Jackendoff (2002). The notation we use in sketching the mental representations is adapted from (Jackendoff, 2002). 'CS' ('conceptual structures') refers to the upper end of the language system and 'SS' ('spatial structures') to the upper end of the visual system (see Fig. 1.2). We index grammatical function and thematic roles with subscripts in conceptual structure (e.g., 'subj/agent'), and thematic roles with subscripts in spatial structure (e.g., 'agent'). Corresponding conceptual and spatial structures are indicated via co-indexing (e.g., 1, 2). Let us now consider an example sentence from Experiment 1:

(6.13) 'The princess (patient) paints the fencer (agent).' (OVS)

Recall the visual scenes from Experiment 1. The scene for the above example sentence depicted a princess-washing-pirate event and a fencer-painting-princess event. We assume that people have recognized the available objects and actions during the preview time. Then they encounter the first word 'the' (*die*). Identification retrieves the meaning of the determiner. At this point gender marking on the determiner of the first noun phrase in

combination with the available scene entities narrows down the domain of reference in the scene (see, e.g., Dahan, Swingley, Tanenhaus, & Magnuson, 2000). Only the female scene entity matches the input. The comprehension system reconciles the mental representations derived from the utterance with scene-derived concepts. When people encounter 'princess', the expectations derived from the determiner and scene context are confirmed. The mental representation for 'the princess' can be integrated with the mental representation of the determiner and of the princess character in the scene. We assume that either during processing of the first noun phrase, or shortly afterwards once the verb has appeared, the initial noun phrase is assigned the grammatical function of a subject (German SVO preference). We suggest that at this point there are at least two possibilities in which structure-building might continue in Account 1 as compared to Account 2.

ACCOUNT 1: THE PRIORITY OF STORED LINGUISTIC AND WORLD KNOWLEDGE

Consider the first account that appears to closely fit the assumptions that are being made in research by Altmann and Kamide (1999) and Kamide, Scheepers, and Altmann (2003). We continue describing comprehension of the example sentence in 6.13 where we left off above (after the first noun phrase). We assume based on findings by Altmann and Kamide (1999), Kamide, Scheepers, and Altmann (2003), and Kamide, Altmann, and Haywood (2003) that once people hear the verb, stored knowledge about the verb's arguments and likely thematic role fillers becomes available. Once available, this knowledge is used to integrate the verb with the existing left context and to assign the grammatical function of (in this case) a subject and thematic role of an agent to the initial noun phrase (example 6.13). Furthermore, results reported in the above studies by Altmann and colleagues indicate that listeners compositionally build mental representations of a post-verbal argument and thematic patient role slot (indicated as 'X' in CS). In this case, we suggest that the conceptual and spatial structures have the following form:

- CS: [PRINCESS$_{subj/agent1}$ PAINT X$_{obj/patient}$]
- SS: [PRINCESS$_1$]

On the basis of the above conceptual structures people attempt to find a postverbal patient role filler in the scene. They find a patient of an action performed by the princess (the pirate). The search for a patient of a princess-painting action, however, fails when people try to reconcile conceptual and spatial structure in the attempt to match the

verb they have just heard (*paint*) with the action that the princess performs in the scene (washing). Since the princess-washing action does not match the verb in the input, the corresponding parts of conceptual and spatial structure cannot be reconciled.

- CS: [PRINCESS$_{subj/agent1}$ PAINT X$_{obj/patient}$]
- SS: [PRINCESS$_1$ WASH PIRATE]

As a result of the failed search, the parser is prompted to provide an alternative conceptual structure. It should be noted that our description at this point may not necessarily reflect the assumptions made by Altmann and colleagues. In the procedural account that we sketch here, the parser rather than the scene proposes interpretations. This approach is compatible with an account in which the scene merely reflects the interpretations proposed by the comprehension system, and can "filter" bad interpretations that are proposed on the basis of linguistic and world knowledge. Here we assume that the parser proposes an OVS structure as an alternative to the failed SVO structure. Based on the representation of an OVS construction, a new search is initiated, this time for a filler of a post-verbal object and agent slot (again indicated as 'X' with a subscript).

- CS: [PRINCESS$_{obj/patient1}$ PAINT X$_{subj/agent}$]
- SS: [PRINCESS$_1$]

A search of the scene identifies the fencer as the agent of a princess-painting action. The resulting conceptual and spatial structures now match:

- CS: [PRINCESS$_{obj/patient1}$ PAINT$_2$ FENCER$_{subj/agent3}$]
- SS: [PRINCESS$_1$ PAINT$_2$ FENCER $_3$]

ACCOUNT 2: COORDINATED INTERPLAY ACCOUNT

We think that based on the above discussion of how the building of conceptual and spatial structure proceeds for findings from Experiment 1 (see example sentence 6.13), Account 1 is compatible with our findings. Consider, however, that there is an alternative account that can also account for findings from the first experiment. In contrast to the priority-of-stereotypical-knowledge account, an alternative possibility ('coordinated interplay' account) would instead suggest the following order of processes in the structure building for

the example sentence 6.13: People have processed the noun phrase 'the princess' and established reference to the princess entity in the scene, as well as assigned the grammatical function of a subject to the initial noun phrase (German SVO preference). However, upon encountering the verb, they do not immediately attach it to the left context and compositionally build conceptual structure on the basis of stored linguistic and world knowledge as suggested in Account 1. Rather, after lexical access and recovery of the meaning of the currently processed lexical item, attention in the scene might first rapidly be directed to the relevant part of the scene. Having found an object that matches the retrieved word meaning, we check that conceptual structure and spatial structure match.

- CS: [PRINCESS$_{subj1}$ PAINT$_2$]
- SS: [PRINCESS$_1$ PAINT$_2$]

While inspecting the painting action people might notice that there is an entity that is doing the painting action, and that this entity looks like a fencer.

- CS: [PRINCESS$_{subj1}$ PAINT$_2$]
- SS: [PRINCESS$_{patient1}$ PAINT$_2$ FENCER$_{agent3}$]

The scene information that has been acquired while inspecting the relevant scene part(s) is then directly used to build conceptual structure. The comprehension system uses the thematic role information derived from the immediately depicted events to assign grammatical function and thematic role to the linguistic input. For the time being we assume that the assignment of thematic role also percolates to syntax and results in assignment of the corresponding grammatical function to the sentential arguments. We do not, however, definitely commit to this view. Further research is required to decide wether thematic role information from the scene only influences the assignment of thematic roles in the linguistic input, or rather also the assignment of grammatical function.

- CS: [PRINCESS$_{obj/patient1}$ PAINT$_2$ FENCER$_{subj/agent3}$]
- SS: [PRINCESS$_{patient1}$ PAINT$_2$ FENCER $_{agent3}$]

DECIDING BETWEEN ACCOUNTS 1 AND 2

We suggest that either of the above two descriptions (Accounts 1 and 2) can account for findings from Experiments 1 to 4. To help determine which of the two accounts fits the

combined findings that we have presented in Experiments 1 to 5, consider now a further processing example from Experiment 5:

(6.14) 'The pilot (pat.) spies-on soon the wizard (depicted agent).'

We focus only on the theoretically most relevant aspects of how processing proceeds for this example. Recall that the scene depicted a pilot who was inactive, a wizard (the depicted agent) who is shown as spying on the pilot (the patient), and a detective (the stereotypical agent) of a spying action. We first consider Account 1. People process the initial noun phrase and assign object-case to it based on unambiguous case marking. Once people have perceived the verb, they access its meaning and stored linguistic/world knowledge associated with it. On the basis of the retrieved linguistic and word knowledge, they integrate the verb with the left context, build an object-verb structure and furthermore project a post-verbal argument and thematic role slot of a spying action of which the first noun phrase is the grammatical object and thematic patient. Stored knowledge associated with the verb enables people to derive expectations of a stereotypical agent.

- CS: [PILOT$_{obj/patient1}$ SPY-ON$_2$ X$_{subj/agent}$]
- SS: [PILOT$_1$]

If comprehension processes are driven by the construction of mental representations from stored linguistic and world knowledge, we would expect that people in preference anticipate a role filler for a stereotypical agent of a spying-action (the detective) in the scene. Clearly, this is not what we find in our data (see chapter 5). Rather, people issued more anticipatory eye-movements to the agent of a depicted spying-action. In contrast to Account 1, the coordinated interplay account (Account 2) correctly predicts this finding. Under Account 2, once people have heard the verb, they rapidly identify the relevant spying-action in the scene. When inspecting the depicted spying-action they are able to identify the wizard as the entity performing the action:

- CS: [PILOT$_{obj/patient1}$ SPY-ON$_2$]
- SS: [PILOT$_{patient1}$ SPY-ON$_2$ WIZARD$_{agent3}$]

The spatial structure built from depicted thematic role relations in the scene is then used to inform the building of conceptual structure:

- CS: [PILOT$_{obj/patient1}$ SPY-ON$_2$ WIZARD$_{subj/agent3}$]
- SS: [PILOT$_{patient1}$ SPY-ON$_2$ WIZARD$_{agent3}$]

In sum, while findings from Experiment 1 do not allow us to decide between the two alternative accounts that we suggsted (Accounts 1 and 2), findings from Experiments 5 appear to fit better with the coordinated interplay account, enabling rapid use of the scene. The coordinated interplay account receives further support from data by Huettig and Altmann (2004). They investigated the influence of similarity of object shape on lexical access. In their experiments, people were presented with utterances such as the following:

(6.15) In the beginning, the zookeeper was worried, but then he looked at the snake and realised that it was harmless.

An example scene showed a snake, and several unrelated distractor objects. Huettig and Altmann (2004) found that participants looked more often to the snake at the onset of *snake* than to any of the distractor objects. In a comparison condition, people inspected a scene that contained no snake, but only an object that was visually similar to a snake (a cable), and further several unrelated distractor objects. When they heard the above utterance about a snake, they inspected the cable more often after having heard *snake* than any of the distractor objects. We view this finding as providing support for our coordinated interplay account. The fact that in the above example (6.15), people looked at the cable when hearing *snake* suggests that there is a strong preference of searching the scene on the basis of knowledge that is retrieved during utterance comprehension. This provides support for our claim that - as soon as people have accessed the meaning of a word - rather than basing the building of conceptual structure on stored linguistic and world knowledge *alone*, they search the scene for something that fits the meaning of the currently processed lexical item, and then build conceptual structure on the basis of the perceived scene information. The mechanism that we identify here can be characterized as follows:

(6.16) Upon accessing the meaning of a word and linguistic/world knowledge associated with it, search the scene and find the scene part that best matches the retrieved linguistic/world knowledge. Then use the acquired scene information to build conceptual structure

Postulating such a mechanism raises some interesting issues. First, we consider how such a mechanism might integrate with the architecture of the language system that Jackendoff (2002) proposes. Despite the fact that the Jackendovian theory makes no clear predictions for on-line comprehension, we have argued that its architectural framework does impose certain constraints on the relative order in which comprehension processes take place. In particular, it would appear that in his framework, the meaning of and knowledge associated with a newly incoming word is first integrated with the already existing conceptual representation (i.e., as detailed in Account 1). The scene is only searched after this process has taken place. If this was indeed the case, then it is not clear how well his framework combines with the procedural account (Account 2) that we proposed. While we think that more research is required, we draw attention to the possibility that it may become necessary to revise the Jackendovian architecture for our purposes if future research should confirm the present findings. For the time being, we think, however, that the overall framework that Jackendoff provides is the best framework that we can adopt owing to its other advantages (e.g., providing a specified linguistic theory, a rich inventory of mental representations, and an explicit interface between the language and visual perception systems).

Further, we draw attention to the status of the mechanism that we proposed for comprehension in general. We assume that the rapid coordinated interplay that we observed only determines comprehension processes when the scene is relevant for comprehension of the utterance. This means that the preferred reliance on depicted events, and the important role of the coordinated interplay during comprehension are presumably not caused by an automatic mechanism of comprehension in general that always triggers a search of the immediate environment upon hearing a word. Rather, we suggest that people can strategically invoke such a mechanism depending on the requirements of the comprehension situation.

Taken together this reveals a view of language comprehension that emphasizes the on-line adaptation of comprehension mechanisms to the requirements of the comprehension situation (see also, e.g., Anderson, 1990; Chater, Crocker, & Pickering, 1998). Such an adaptive character of comprehension mechanisms may have developed over a long time. Indeed, as outlined in section 6.2.3, it may have its roots in the paramount role that the rapid coordination of comprehension and scene perception plays during language acquisition. The mechanism that we identified (see 6.16) presumably dominates comprehension during language acquisition. However, with the acquisition of abstract concepts during

later childhood, and owing to the fact that the immediate scene is only sometimes relevant for comprehension during later adult life, this mechanism has become strategic: quickly available if needed, but not interfering with comprehension when the scene is irrelevant for recovering the meaning of a sentence.

Crucially, despite the fact that language is often about abstract ideas in adult life, the undiminished importance of the mechanism that we identified becomes clear in situations when language is related to the immediate scene. In this situation, the mechanism that we sketched in 6.16 is responsible for the key findings presented in this thesis. The first fundamental insight is that depicted events rapidly influence structural disambiguation and incremental thematic role assignment when identified by cues in the utterance. Further, our findings show that there is a tight coordination between comprehension processes and concurrent perceptual processing of the scene: The utterance identifies relevant events in the scene. Closely temporally coordinated with this identification, depicted events inform spoken language comprehension. Finally, we have demonstrated that the comprehension system relies on verb-mediated depicted events in preference to stereotypical thematic role knowledge for incremental thematic interpretation when the utterance identifies both of these types of knowledge as relevant.

CHAPTER 7

CONCLUSIONS

The present thesis has investigated comprehension in situations in which language was related to the immediate scene. While current psycholinguistic theories have mainly been developed on the basis of insights from comprehension during reading of isolated sentences, we provide a description of how utterances are processed that relate to the immediate environment. The present research thus adds to existing theories of comprehension, making the first step towards a theory that can account for comprehension both when language is and when it is not about the immediate scene.

We have suggested that, under the influence of Fodorian modularity, theoretical as well as experimental psycholinguistic research on on-line sentence comprehension has predominantly been concerned with comprehension during reading and with the effects of stored linguistic and world knowledge on comprehension. This was due to the assumption that core comprehension processes such as the structuring of an utterance could not be influenced on-line by ongoing non-linguistic processes such as visual perception. As a result, relatively little is known about the influence of scene perception on comprehension. One important study, however, by Tanenhaus et al. (1995) has demonstrated the rapid incremental influence of a visual referential context on the on-line structuring of an utterance.

To extend our knowledge on how comprehension proceeds when language is related to the scene, we investigated two key issues. First, we explored which types of information in scenes have the potential to affect on-line comprehension processes such as structural disambiguation and incremental thematic role assignment. Second, we examined the nature of the interplay between visual perception and on-line utterance comprehension.

Taken together, our findings show that perceived event relations between entities in the immediate visual context allow the rapid recovery of mental representations such as thematic roles in the on-line comprehension of spoken sentences, and actively influence incremental resolution of initial structural and role ambiguity in the linguistic input. The influence of scene information on comprehension was determined by a *closely temporally coordinated* interplay between visual perception and comprehension, in which both scene-driven and utterance-driven processes can take the lead. Consider now in more detail findings from the individual experiments that we conducted.

In Experiments 1 and 2, anticipatory eye-movements in the event scenes provided evidence for expectations of a patient and agent role filler for initially ambiguous subject-verb-object and object-verb-subject sentences respectively once the verb had uniquely identified one of two depicted actions. Experiment 2 replicated the early visual disambiguation effect in the presence of distractor objects and investigating more fully the possible influence of intonation. The rapid verb-mediated accessing of the depicted actions and the use of thematic role relations associated with these actions for incremental thematic role assignment suggests a tight coupling of visual and linguistic processing. An utterance-driven visual search of the scene cooperates in tight coordination with the scene-driven influence of event structures on incremental interpretation of the sentence.

Experiment 3 demonstrated that when the verb was sentence-final and did not establish reference to the depicted events early, linguistic cues that preceded the verb identified the relevant depicted event for early disambiguation. Disambiguation was reflected by eye-movement patterns both during ambiguity resolution at the case-marked determiner/preposition of the second argument, and before this point as a result of the influence of temporal adverbs biasing towards the active or passive structure. This finding of early disambiguation further supports an account of our data in terms of thematic role assignment, since the depicted event scenes were used for early thematic disambiguation before the verb had referred to the depicted actions.

Findings from Experiment 4 extended the insights gained from Experiments 1 to 3 to a further sentence construction (main verb/reduced relative clause sentences) and moreover established the generality of our findings for another language (English). Our results showed that verbs as well as other soft linguistic cues could be used to identify depicted actions and their associated thematic relations during comprehension of initially structurally ambiguous English sentences. When the verb identified the depicted action, anticipatory eye-movements to a patient and agent role filler for main clause and reduced relative

clause sentences respectively provided evidence for rapid structural disambiguation and incremental thematic role assignment during the post-verbal time region. When soft linguistic cues preceded the lexical verb in the ambiguous clause, we found evidence for such disambiguation already during the lifetime of the verb, and thus earlier than in cases where no soft linguistic cues preceded the verb (Experiments 1, 2, and 4a).

Experiments 1 to 4 provide strong support for the view that depicted events play an important role in incremental thematic role assignment and structural disambiguation, and further that this is not limited to a specific construction, word order, or language. We suggested importantly that the effects of scene information on comprehension are probably not the result of one informational source and process (i.e., scene perception) alone. Rather, we propose that this effect might be the outcome of a coordinated interplay between distinct cognitive processes such as visual perception and comprehension processes. Studies on comprehension during reading have examined the temporal coordination between when a word in the utterance makes important information available, and when that information influences comprehension processes. It has been demonstrated that there is a close temporal coordination between when a word is perceived and when its meaning and associated stored linguistic/world knowledge affect comprehension. When language referred to the immediate scene, a similarly close temporal coordination was found between comprehension of a word and attention to objects in the scene. In contrast, little attention has been devoted to the coordination between utterance comprehension, ensuing attention in the scene, and the time when *scene information* influences the *structuring* of an utterance. Experiments 1 to 4 made clear that there was a very close temporal coordination between when a cue in the utterance made a depicted event available, and the time when eye-movements revealed that structural disambiguation of the utterance took place.

To investigate the nature of this interplay in more detail, Experiment 5 examined the relative importance of distinct informational sources such as stored thematic role knowledge and depicted events. We compared the use of these two types of information for comprehension in two different situations. In one situation the verb uniquely identified either the agent of a stereotypical (non-depicted) action indicated by the verb or the agent of a depicted (non-stereotypical) action as relevant for comprehension. In this condition we expected to replicate previous findings that have shown that each of these two informational sources can influence incremental thematic interpretation. In a critical second condition, we manipulated the verb so that it identified both an agent of a stereotypical

action and another agent of a depicted action in the scene. The comprehension system was thus free to rely on stored thematic knowledge and choose the agent of the stereotypical event for thematic interpretation. Alternatively it could rely on perceptual processing of the depicted event, or neither of the two.

Within a single study the observed eye-gaze pattern show that both stereotypical knowledge and depicted scene events each allow rapid thematic role interpretation of an unfolding utterance. However, when the verb did not uniquely identify either a depicted or a stereotypical agent people had a clear preference for relying on thematic relations available from the depicted action performed by an agent in the scene. Such a finding can be motivated by insights from language acquisition. During acquisition, children explore their environment and acquire new information from entities and events in the immediate scene. When they hear an adult describe that a girl is kicking a ball, and they do not know the verb *kick*, perception of the event may allow them to acquire the conceptual meaning of the kicking-action, and to further infer that the entity doing the kicking is called *girl*. Crucially, such perceptual exploration of the immediate scene for the acquisition of new concepts is utterance-mediated. The sentence directs attention to entities or events that then in turn can inform language acquisition. The important role of the immediate environment for language *acquisition* may well have shaped the architecture and mechanisms of language *comprehension*.

This interplay between scene perception and comprehension presumably only determines comprehension in situations in which language is about the immediate scene. This is clearly only sometimes the case. In adult life, for instance, language often has other functions such as discussing political views, abstract ideas or recent events. This means that the observed preference for the use of depicted events is probably not automatic but a context-sensitive mechanism that we can rapidly use whenever it is needed but that does not interfere with comprehension when the immediate scene is clearly irrelevant for comprehension.

Importantly, while language in adult life is often about abstract ideas rather than related to the immediate scene, the findings from our experiments reveal the undiminished importance of scene information for comprehension when language does relate to the immediate scene. First, our findings show that depicted events can rapidly influence incremental structural disambiguation and thematic role assignment if they are uniquely identified by the utterance. Further, processing of scene and utterance are tightly related. Findings from our studies revealed that the effects of depicted scene events on the struc-

turing and interpretation of an utterance are closely temporally coordinated with their identification through cues in the utterance. Finally, in a situation where the scene is clearly relevant for spoken language comprehension, the comprehension system relies in preference on depicted events rather than on stored thematic role knowledge for incremental thematic interpretation. Taken together, these findings reveal that the important role of the coordinated interplay between visual perception and comprehension in language acquisition has fundamentally shaped the mechanisms of language comprehension.

Chapter 8

German Summary

Die vorliegende Doktorarbeit hat Sprachverstehen in Situationen erforscht, in denen sich Sprache auf die unmittelbare Umgebung bezieht. Während die meisten psycholinguistischen Theorien anhand von Befunden aus Studien entwickelt wurden, die das Leseverständnis untersuchten, ermöglichen unsere Forschungsergebnisse es, Sprachverstehen zu beschreiben, wenn sich die zu verstehenden Sätze auf die unmittelbare Umgebung beziehen. Unsere Forschung bereichert somit bestehende Theorien, die aus der Untersuchung des Leseverständnisses hervorgingen, und macht einen ersten Schritt in Richtung einer psycholinguistischen Theorie, die sowohl für das Sprachverstehen während des Lesens, als auch für das Verstehen gesprochener Sprache die sich auf die unmittelbare Umgebung bezieht, Gültigkeit besitzt.

Wir haben dargelegt, dass theoretische als auch experimentelle psycholinguistische Forschung über on-line Sprachverstehen sich in erster Linie mit Sprachverstehen während des Lesens befasste, und die Effekte von linguistischem und Weltwissen auf das Sprachverstehen untersuchte. Diese einseitige Ausrichtung der Forschung kam durch die Annahme zustande, dass zentrale Verstehensprozesse wie zum Beispiel die Strukturierung einer Äußerung nicht on-line durch gleichzeitig ablaufende Prozesse wie zum Beispiel visuelle Perzeption beeinflusst werden können. Infolgedessen ist nur verhältnismäßig wenig über den Einfluss visueller Wahrnehmung auf Sprachverstehen bekannt. Eine wichtige Studie, durchgeführt von Tanenhaus et al. (1995), hat allerdings gezeigt, dass ein visueller referentieller Kontext einen äußerst raschen Einfluss auf die inkrementelle Strukturierung einer Äußerung haben kann.

Zur Erweiterung unseres Wissens bezüglich der Sprachverstehensprozesse, die ablaufen wenn wir Sprache verstehen, die sich auf die unmittelbare Umgebung bezieht, haben wir zwei grundlegende Forschungsfragen untersucht. Erstens haben wir erforscht, welche Arten von Information in visuellen Szenen das Potential haben, on-line Sprachverarbeitungsprozesse wie zum Beispiel die Strukturierung einer Äußerung oder die inkrementelle Zuweisung thematischer Rollen zu beeinflussen. Zweitens haben wir genauer ermittelt, welcher Art das Zusammenspiel zwischen visueller Perzeption und dem inkrementellen Verstehen einer Äußerung ist.

Unsere Befunde zeigen, dass wahrgenommene Handlungsbezüge und Ereignisse zwischen Personen im unmittelbaren visuellen Umfeld es ermöglichen, rasch mentale Repräsentationen thematischer Rollen zu bilden. Solche mentalen Strukturen thematischer Relationen ermöglichten sodann aktiv die inkrementelle Zuweisung von thematischen Rollen und die Strukturierung einer anfänglich strukturell ambigen Äußerung. Der Einfluss von Bildinformation auf die Sprachverarbeitung wurde durch ein zeitlich genau koordiniertes Zusammenspiel zwischen visueller Perzeption und Sprachverstehen bestimmt, wobei sowohl die sich entfaltende Äußerung als auch die wahrgenommene Szene die Führungsrolle in diesem Zusammenspiel übernehmen kann.

In den Experimenten 1 und 2 enthüllten antizipatorische Augenbewegungen in Agens-Aktion-Patiens-Handlungsszenen die Erwartung eines Patiens- und Agens-Rollen-Füller für initial ambige SVO- und OVS-Sätze respektive, sobald das Verb eine der beiden abgebildeten Handlungen eindeutig identifiziert hatte. Diese Augenbewegungen traten kurz nach dem Verb, und ehe Kasusmarkierung auf der zweiten Nominalphrase die Satzstruktur als SVO- oder OVS-Struktur disambiguierte, auf. Experiment 2 replizierte diese frühe visuelle Disambiguierung unter der Hinzunahme von Distraktoren-Objekten in den Szenen und untersuchte dabei zusätzlich genauer den möglichen Einfluss von Intonation. Der rasche, auf der Identifikation durch das Verb basierende Zugriff auf die abgebildeten Handlungen und der schnelle Einfluss der ihnen angegliederten thematischen Relationen deutet auf ein enges Zusammenspiel von visueller und sprachlicher Verarbeitung hin. Der soeben vernommene Satz lenkt die visuelle Aufmerksamkeit in der Szene. Dieser Prozess bestimmt dann in engem Zusammenspiel mit dem Einfluss von Ereignisstrukturen aus der visuellen Szene die inkrementelle Interpretation der Äußerung.

Experiment 3 hat gezeigt, dass auch, wenn das Verb am Satzende erscheint und somit erst spät im Satz die zur Verfügung stehenden abgebildeten Handlungen der Sprachverarbeitung zugänglich machen kann, dennoch andere sprachliche, dem Verb vorausgehende

Hinweise, wie zum Beispiel Adverbien oder Hilfsverben, das jeweils relevante abgebildete Ereignis für die strukturelle Disambiguierung zugänglich machen können. Die Disambiguierung war aus den Augenbewegungsmustern sowohl während der Kasusmarkierung/ Präposition der zweiten Nominalphrase, als auch früher, während der Verarbeitung temporaler Adverbien, die der Nominalphrase und dem Verb vorausgingen und die die Bildung einer Aktiv- und Passivstruktur zur Folge hatten, ersichtlich. Der Fund einer frühen Disambiguierung in Experiment 3 bestätigt unsere Interpretation der beobachteten Augenbewegungen als einen Indikator für thematische Rollenvergabe, da die abgebildeten Ereignisse für eine frühe thematische Disambiguierung verwendet wurden, ehe das Verb auf sie Bezug genommen hatte.

Die Befunde des vierten Experiments bestätigten und erweiterten die Erkenntnisse, die wir aus den ersten drei Experimenten gewonnen haben. Experiment 4 übertrug die bestehenden Befunde auf eine weitere Sprache (Englisch), und dehnte sie auf eine weitere Satzkonstruktion (die Relativsatz-Ambiguität im Englischen) aus. Wie schon in den Experimenten 1 und 2 konnten abgebildete Handlungen, nachdem sie durch ein Verb im Satz identifiziert worden waren, entscheidend zur strukturellen Disambiguierung und thematischen Rollenvergabe beitragen. Weiterhin konnten wir zeigen, dass auch in der strukturell ambigen englischen Relativsatzkonstruktion schwache linguistische Hinweise die Ereignisse in den Szenen zugänglich machen können, und in der Lage sind, die Disambiguierung und thematische Rollenvergabe zu beeinflussen. Wenn das Verb die abgebildeten Handlungen identifizierte, fanden wir - wie auch in den Experimenten 1 und 2 - kurz nach dem Verb und vor der zweiten Nominalphrase rasche Augenbewegungen zum Patiens und Agens in Abhängigkeit von der Satzstruktur (Hauptsatz/reduzierter Relativsatz). Wenn hingegen schwache sprachliche Hinweise vorhanden waren, um die abgebildeten Handlungen für die Spachverarbeitung verfügbar zu machen, zeigten die Augenbewegung bereits strukturelle Disambiguierung während des Verbs an, und somit früher als in den Experimenten, in welchen dem Verb keine schwachen sprachlichen Hinweise vorangingen (Experiment 1, 2, und 4a).

Die Experimente 1 bis 4 haben somit der Behauptung, dass dargestellte Ereignisse eine wichtige Rolle bei der inkrementellen thematischen Rollenvergabe und der strukturellen Disambiguierung spielen, starken Rückhalt gegeben. Ferner ist diese Erkenntnis nicht auf eine bestimmte Satzkonstruktion, Wortstellung, oder Sprache beschränkt. Entscheidend ist hierbei auch, dass die von uns beobachteten Effekte vermutlich nicht auf den alleinigen Einfluss einer Informationsquelle und eines Prozesses (z.B., visuelle Perzeption) zurück-

zuführen sind. Vielmehr ist es wahrscheinlich, dass unseren Ergebnissen ein koordiniertes Zusammenspiel von Sprachverstehen und visueller Perzeption zugrunde liegt. Studien zum Sprachverstehen während des Lesens zeigen, dass es einen engen zeitlichen Zusammenhang zwischen dem Zeitpunkt, zu dem die Bedeutung eines Wortes sowie assoziiertes Wissen zur Verfügung stehen, und dem Zeitpunkt, zu dem diese Informationen das Sprachverstehen beeinflussen, gibt. Im Gegensatz dazu wurde dem Zusammenspiel zwischen Sprachverständnis, Aufmerksamkeit in der visuellen Szene, und dem Zeitpunkt zu welchem ein *Szenenkontext* die *Strukturierung* einer Äußerung beeinflussen kann, nur wenig Aufmerksamkeit gewidmet. Die ersten vier Experimente dieser Doktorarbeit zeigten, dass zwischen dem Zeitpunkt, zu welchem ein Hinweis in der sprachlichen Äußerung ein dargestelltes Ereignis der Sprachverarbeitung zugänglich macht, und dem Zeitpunkt zu welchem Augenbewegungen deutlich machen, dass sprachliche Disambiguierung der Äußerung stattfindet, ein sehr enger zeitlicher Zusammenhang besteht.

Um den Charakter diese Zusammenspiels noch genauer zu erforschen hat Experiment 5 die relative Gewichtung verschiedenartiger Informationsquellen wie beispielsweise stereotypem thematischem Rollenwissen und dargestellten Ereignissen in einer Szene untersucht. Wir haben den Einfluss dieser beiden Informationsarten in zwei verschiedenen Situationen näher betrachtet: Alle verwendeten Sätze hatten eine eindeutige OVS Struktur. Nachdem die erste Nominalphrase einen Patiens identifiziert hatte, identifizierte das Verb in einer Situation entweder nur den Agens einer stereotypen, aber nicht abgebildeten Handlung oder nur den Agens einer abgebildeten, aber nicht stereotypen Handlung. In dieser Situation erwarteten wir bestehende Ergebnisse zu replizieren, die gezeigt haben, dass jede dieser beiden Informationsarten Sprachverarbeitung und inkrementelle Rollenzuweisung beeinflussen kann. In der entscheidenden zweiten Situation hingegen identifizierte das Verb den relevanten Agens nicht eindeutig, sondern identifizierte sowohl den Agens einer stereotypen als auch den weiteren Agens einer abgebildeten Handlung als relevant für das Verstehen der Äußerung. Das Sprachverarbeitungssystem konnte sich somit bei der thematischen Rollenzuweisung entweder auf stereotypes thematisches Wissen verlassen und sich für den stereotypen Agens entscheiden. Als Alternative konnte es auf die Perzeption der Szenenereignisse zurückgreifen, oder aber auf keine der beiden genannten Informationsarten.

Die beobachteten Augenbewegungen zeigten, dass sowohl stereotypes thematisches Rollenwissen als auch abgebildete Ereignisse rasche thematische Rollenvergabe in einer sich entfaltenden sprachlichen Äußerung ermöglichten, wenn das Verb die jeweilige Informa-

tionsquelle eindeutig identifizierte. Wenn hingegen das Verb nicht eindeutig entweder einen stereotypen Agens oder den Agens einer abgebildeten Handlung identifizierte, verließ sich das Sprachsystem vorzugsweise auf dargestellte Ereignisse in der thematischen Rollenvergabe.

Solch ein Befund kann unter Zuhilfenahme von Erkenntnissen der Spracherwerbsforschung erklärt werden. Während des Spracherwerbs erforschen Kinder ihre Umgebung und eignen sich höchstwahrscheinlich Wissen über die Dinge und Ereignisse in ihrer unmittelbaren Umgebung an. Wenn Kinder zum Beispiel einen Erwachsenen sprechen hören, der gerade ein Ereignis beschreibt, in dem ein Mädchen einen Ball schießt, und sie das Verb *schießen* nicht kennen, so kann die Wahrnehmung des Ereignisses es ihnen ermöglichen, die konzeptuelle Bedeutung der Schieß-Handlung zu begreifen. Ferner ermöglicht die Wahrnehmung der Schieß-Handlung zu schlußfolgern, dass das Objekt, welches den Schuss ausführt, *Mädchen* genannt wird. Diese Ausführungen machen deutlich, dass das Sprachsystem darauf ausgerichtet ist neue Information aus der Umgebung zu erwerben, statt sich nur auf erworbenes Wissen zu verlassen. Die Wahrnehmung der unmittelbaren Umgebung ist dabei entscheidend durch sprachliche Äußerungen, die sich auf die Umgebung beziehen, beeinflusst. Ein Satz vermag Aufmerksamkeit auf Dinge und Ereignisse in der Umgebung zu lenken. Information über Dinge und Ereignisse im visuellen Kontext steht dann zur Verfügung um das Sprachverstehen inkrementell zu beeinflussen.

Wir vermuten, dass das Zusammenspiel von visueller Wahrnehmung der Umgebung und inkrementeller Interpretation einer Äußerung das Sprachverstehen nur in Situationen beeinflusst, in denen der visuelle Kontext für das Sprachverständnis von Bedeutung ist (also zum Beispiel wenn sich eine Äußerung auf Dinge und Ereignisse in der unmittelbaren Umgebung bezieht). Die Präferenz für unmittelbar abgebildete im Vergleich mit stereotypen aber nicht abgebildeten Ereignissen beim Sprachverstehen hat also vermutlich nur in Situationen Gültigkeit hat, in denen der visuelle Kontext für das Sprachverstehen relevant ist. Dies bedeutet, dass die von uns beobachtete Präferenz wahrscheinlich nicht automatischer Natur ist, sondern eher den Charakter eines kontext-sensitiven Mechanismus hat, der je nach Bedarf zur Verfügung steht, der aber nicht in den Sprachverstehensprozess eingreift wenn der Kontext für das Verstehen irrelevant ist. Vor allem im Erwachsenenleben ist dies meist nicht der Fall. Sprache hat größtenteils andere Funktionen und bezieht sich häufig auf abstrakte Ideen: Dinge die nicht unmittelbar präsent sind, oder Erzählungen vergangener Ereignisse.

Eine Beschreibung unseres Befunds als kontext-sensitivem Mechanismus erscheint uns, angesichts der Tatsache, dass sich Sprache im Erwachsenenleben nur manchmal auf die unmittelbare Umgebung bezieht, plausibel. Unsere Ergebnisse belegen jedoch den unverminderten Einfluss der unmittelbaren Szene in Situationen in denen sich Sprache auf die Umgebung bezieht: Erstens zeigen unsere Befunde, dass abgebildete Ereignisse rasch strukturelle Disambiguierung und inkrementelle thematische Rollenvergabe beeinflussen. Zweitens, demonstrieren sie eine enge zeitliche Koordination zwischen Sprachverstehen und der Perzeption des visuellen Umfelds. Ganz entscheidend machen sie deutlich, dass das Sprachsystem sich, wenn der visuelle Kontext relevant ist, eher auf den visuellen Kontext als auf das erworbene Wissen verlässt. Basierend auf diesen Ergebnissen sind wir davon überzeugt, dass die wichtige Rolle, die das koordinierte Zusammenspiel von visueller Perzeption und Sprachverstehen im Sprach*erwerb* einnimmt, die Architektur des Sprachsystems und die Mechanismen des Sprach*verstehens* längerfristig entscheidend beeinflusst hat.

APPENDIX A

ITEM SENTENCES

Below are listed the item sentences for Experiments 1 to 5. Sentences that are marked with the Roman numeral (1) were presented with Image 1 of an item set. Sentences marked with (2) were presented with Image version 2 of an item set (see Appendix B for the item images).

A.1 SENTENCES FOR EXPERIMENTS 1 AND 2

Item 1

1. (a) Die Prinzessin wäscht offensichtlich den Pirat.
   (b) Die Prinzessin malt offensichtlich der Fechter.
2. (a) Die Prinzessin malt offensichtlich den Fechter.
   (b) Die Prinzessin wäscht offensichtlich der Pirat.

Item 2

1. (a) Die Amazone erdolcht gerade den Mechaniker.
   (b) Die Amazone besprüht gerade der Fußballspieler.
2. (a) Die Amazone besprüht gerade den Fußballspieler.
   (b) Die Amazone erdolcht gerade der Mechaniker.

Item 3

1. (a) Die Krankenschwester schubst in diesem Moment den Sportler.
   (b) Die Krankenschwester fönt in diesem Moment der Priester.
2. (a) Die Krankenschwester fönt in diesem Moment den Priester.
   (b) Die Krankenschwester schubst in diesem Moment der Sportler.

Item 4

1. (a) Die Journalistin fesselt in diesem Moment den Matrosen.
   (b) Die Journalistin füttert in diesem Moment der Oberarzt.
2. (a) Die Journalistin füttert in diesem Moment den Oberarzt.
   (b) Die Journalistin fesselt in diesem Moment der Matrose.

Item 5

1. (a) Die Bauarbeiterin attackiert offensichtlich den Cellist.
   (b) Die Bauarbeiterin interviewt offensichtlich der Golfer.
2. (a) Die Bauarbeiterin interviewt offensichtlich den Golfer.
   (b) Die Bauarbeiterin attackiert offensichtlich der Cellist.

Item 6

1. (a) Die Teufelin beschenkt in diesem Moment den Clown.
   (b) Die Teufelin skizziert in diesem Moment der Koch.
2. (a) Die Teufelin skizziert in diesem Moment den Koch.
   (b) Die Teufelin beschenkt in diesem Moment der Clown.

Item 7

1. (a) Die Putzfrau bewirft soeben den Kellner.
   (b) Die Putzfrau ohrfeigt soeben der Ritter.
2. (a) Die Putzfrau ohrfeigt soeben den Ritter.
   (b) Die Putzfrau bewirft soeben der Kellner.

Item 8

1. (a) Die Schlittschuhläuferin schrubbt mal eben den Detektiv.
   (b) Die Schlittschuhläuferin stupst mal eben der Zauberer.
2. (a) Die Schlittschuhläuferin stupst mal eben den Zauberer.
   (b) Die Schlittschuhläuferin schrubbt mal eben der Detektiv.

Item 9

1. (a) Das Dienstmädchen parfümiert in diesem Moment den Henker.
   (b) Das Dienstmädchen bandagiert in diesem Moment der Trommler.
2. (a) Das Dienstmädchen bandagiert in diesem Moment den Trommler.
   (b) Das Dienstmädchen parfümiert in diesem Moment der Henker.

Item 10

1. (a) Die Tennisspielerin boxt hier den Sträfling .
   (b) Die Tennisspielerin kämmt hier der Flötist.
2. (a) Die Tennisspielerin kämmt hier den Flötist.
   (b) Die Tennisspielerin boxt hier der Sträfling.

Item 11

1. (a) Die Meerjungfrau krönt gerade den Student.
   (b) Die Meerjungfrau zupft gerade der Soldat.
2. (a) Die Meerjungfrau zupft gerade den Soldat.
   (b) Die Meerjungfrau krönt gerade der Student.

Item 12

1. (a) Die Nonne impft gerade den Schülerlotsen.
   (b) Die Nonne zwickt gerade der Klarinettist.
2. (a) Die Nonne zwickt gerade den Klarinettist.
   (b) Die Nonne impft gerade der Schülerlotse.

Item 13

1. (a) Die Oma kratzt soeben den Bogenschützen.
   (b) Die Oma filmt soeben der Saxophonist.
2. (a) Die Oma filmt soeben den Saxphonist.
   (b) Die Oma kratzt soeben der Bogenschütze.

Item 14

1. (a) Die Fee bürstet hier den Gangster.
   (b) Die Fee bespritzt hier der Tourist.
2. (a) Die Fee bespritzt hier den Tourist.
   (b) Die Fee bürstet hier der Gangster.

Item 15

1. (a) Die Joggerin verhext mal eben den Doktor.
   (b) Die Joggerin frottiert mal eben der König.
2. (a) Die Joggerin frottiert mal eben den König.
   (b) Die Joggerin verhext mal eben der Doktor.

Item 16

1. (a) Die Cheerleaderin verprügelt offensichtlich den Pagen.
   (b) Die Cheerleaderin vergiftet offensichtlich der Angler.
2. (a) Die Cheerleaderin vergiftet offensichtlich den Angler.
   (b) Die Cheerleaderin verprügelt offensichtlich der Page.

Item 17

1. (a) Die Braut verhaut gerade den Pfadfinder.
   (b) Die Braut verbrüht gerade der Postbote.
2. (a) Die Braut verbrüht gerade den Postboten.
   (b) Die Braut verhaut gerade der Pfadfinder.

Item 18

1. (a) Die Stewardess pudert soeben den Leichtathlet.
   (b) Die Stewardess rempelt soeben der Wanderer.
2. (a) Die Stewardess rempelt soeben den Wanderer.
   (b) Die Stewardess pudert soeben der Leichtathlet.

Item 19

1. (a) Die Hexe bestrahlt hier den Zeitungsverkäufer.
   (b) Die Hexe bestiehlt hier der Strassenkehrer.
2. (a) Die Hexe bestiehlt hier den Strassenkehrer.
   (b) Die Hexe bestrahlt hier der Zeitungsverkäufer.

Item 20

1. (a) Die Japanerin beschmiert mal eben den Kameramann.
   (b) Die Japanerin bekränzt mal eben der Ordnunghüter.
2. (a) Die Japanerin bekränzt mal eben den Ordnungshüter.
   (b) Die Japanerin beschmiert mal eben der Kameramann.

Item 21

1. (a) Die Rollstuhlfahrerin kostümiert hier den Schiedsrichter.
   (b) Die Rollstuhlfahrerin besoldet hier der Chinese.
2. (a) Die Rollstuhlfahrerin besoldet hier den Chinesen.
   (b) Die Rollstuhlfahrerin kostümiert hier der Schiedsrichter.

Item 22

1. (a) Die Geschäftsfrau umgürtet mal eben den Klempner.
   (b) Die Geschäftsfrau verköstigt mal eben der Imker.
2. (a) Die Geschäftsfrau verköstigt mal eben den Imker.
   (b) Die Geschäftsfrau umgürtet mal eben der Klempner.

Item 23

1. (a) Die Badenixe maskiert soeben den Skifahrer.
   (b) Die Badenixe entlohnt soeben der Musketier.
2. (a) Die Badenixe entlohnt soeben den Musketier.
   (b) Die Badenixe maskiert soeben der Skifahrer.

Item 24

1. (a) Die Aerobic-Trainerin verwarnt offensichtlich den Astronaut.
   (b) Die Aerobic-Trainerin bekocht offensichtlich der Handwerker.
2. (a) Die Aerobic-Trainerin bekocht offensichtlich den Handwerker.
   (b) Die Aerobic-Trainerin verwarnt offensichtlich der Astronaut.

A.2 SENTENCES FOR EXPERIMENT 3

Item 1

1. (a) Die Prinzessin wird sogleich den Pirat waschen.
   (b) Die Prinzessin wird soeben von dem Fechter gemalt.
2. (a) Die Prinzessin wird sogleich den Fechter malen.
   (b) Die Prinzessin wird soeben von dem Pirat gewaschen.

Item 2

1. (a) Die Amazone wird sofort den Mechaniker erdolchen.
   (b) Die Amazone wird derzeit von dem Fußballspieler besprüht.
2. (a) Die Amazone wird sofort den Fußballspieler besprühen.
   (b) Die Amazone wird derzeit von dem Mechaniker erdolcht.

Item 3

1. (a) Die Krankenschwester wird unverzüglich den Sportler schubsen.
   (b) Die Krankenschwester wird im Augenblick von dem Priester gefönt.

2. (a) Die Krankenschwester wird unverzüglich den Priester fönen.
   (b) Die Krankenschwester wird im Augenblick von dem Sportler geschubst.

Item 4

1. (a) Die Journalistin wird umgehend den Matrosen fesseln.
   (b) Die Journalistin wird momentan von dem Oberarzt gefüttert.
2. (a) Die Journalistin wird umgehend den Oberarzt füttern.
   (b) Die Journalistin wird momentan von dem Matrosen gefesselt.

Item 5

1. (a) Die Bauarbeiterin wird unverzüglich den Cellist attackieren.
   (b) Die Bauarbeiterin wird im Augenblick von dem Golfer interviewt.
2. (a) Die Bauarbeiterin wird unverzüglich den Golfer interviewen.
   (b) Die Bauarbeiterin wird im Augenblick von dem Cellist attackiert.

Item 6

1. (a) Die Teufelin wird unverzüglich den Clown beschenken.
   (b) Die Teufelin wird im Augenblick von dem Koch skizziert.
2. (a) Die Teufelin wird unverzüglich den Koch skizzieren.
   (b) Die Teufelin wird im Augenblick von dem Clown beschenkt.

Item 7

1. (a) Die Putzfrau wird umgehend den Kellner bewerfen.
   (b) Die Putzfrau wird momentan von dem Ritter geohrfeigt.
2. (a) Die Putzfrau wird umgehend den Ritter ohrfeigen.
   (b) Die Putzfrau wird momentan von dem Kellner beworfen.

Item 8

1. (a) Die Schlittschuhläuferin wird unverzüglich den Detektiv schrubben.
   (b) Die Schlittschuhläuferin wird im Augenblick von dem Zauberer gestupst.

2. (a) Die Schlittschuhläuferin wird unverzüglich den Zauberer stupsen.
   (b) Die Schlittschuhläuferin wird im Augenblick von dem Detektiv geschrubbt.

Item 9

1. (a) Das Dienstmädchen wird sofort den Henker parfümieren.
   (b) Das Dienstmädchen wird derzeit von dem Trommler bandagiert.
2. (a) Das Dienstmädchen wird sofort den Trommler bandagieren.
   (b) Das Dienstmädchen wird derzeit von dem Henker parfümiert.

Item 10

1. (a) Die Tennisspielerin wird umgehend den Sträfling boxen.
   (b) Die Tennisspielerin wird momentan von dem Flötist gekämmt.
2. (a) Die Tennisspielerin wird umgehend den Flötist kämmen.
   (b) Die Tennisspielerin wird momentan von dem Sträfling geboxt.

Item 11

1. (a) Die Meerjungfrau wird sofort den Student krönen.
   (b) Die Meerjungfrau wird derzeit von dem Soldat gezupft.
2. (a) Die Meerjungfrau wird sofort den Soldat zupfen.
   (b) Die Meerjungfrau wird derzeit von dem Student gekrönt.

Item 12

1. (a) Die Nonne wird umgehend den Schülerlotsen impfen.
   (b) Die Nonne wird momentan von dem Klarinettist gezwickt.
2. (a) Die Nonne wird umgehend den Klarinettist zwicken.
   (b) Die Nonne wird momentan von dem Schülerlotsen geimpft.

Item 13

1. (a) Die Oma wird unverzüglich den Bogenschützen kratzen.
   (b) Die Oma wird im Augenblick von dem Saxophonist gefilmt.

2. (a) Die Oma wird unverzüglich den Saxophonist filmen.
   (b) Die Oma wird im Augenblick von dem Bogenschützen gekratzt.

Item 14

1. (a) Die Fee wird sogleich den Gangster bürsten.
   (b) Die Fee wird soeben von dem Tourist bespritzt.
2. (a) Die Fee wird sogleich den Tourist bespritzen.
   (b) Die Fee wird soeben von dem Gangster gebürstet.

Item 15

1. (a) Die Joggerin wird sogleich den Doktor verhexen.
   (b) Die Joggerin wird soeben von dem König frottiert.
2. (a) Die Joggerin wird sogleich den König frottieren.
   (b) Die Joggerin wird soeben von dem Doktor verhext.

Item 16

1. (a) Die Cheerleaderin wird umgehend den Pagen verprügeln.
   (b) Die Cheerleaderin wird momentan von dem Angler vergiftet.
2. (a) Die Cheerleaderin wird umgehend den Angler vergiften.
   (b) Die Cheerleaderin wird momentan von dem Pagen verprügelt.

Item 17

1. (a) Die Braut wird sofort den Pfadfinder verhauen.
   (b) Die Braut wird derzeit von dem Postboten verbrüht.
2. (a) Die Braut wird sofort den Postboten verbrühen.
   (b) Die Braut wird derzeit von dem Pfadfinder verhauen.

Item 18

1. (a) Die Stewardess wird umgehend den Leichtathlet pudern.
   (b) Die Stewardess wird momentan von dem Wanderer gerempelt.

2. (a) Die Stewardess wird umgehend den Wanderer rempeln.
   (b) Die Stewardess wird momentan von dem Leichtathlet gepudert.

Item 19

1. (a) Die Hexe wird sofort den Zeitungsverkäufer bestrahlen.
   (b) Die Hexe wird derzeit von dem Straßenkehrer bestohlen.
2. (a) Die Hexe wird sofort den Straßenkehrer bestehlen.
   (b) Die Hexe wird derzeit von dem Zeitungsverkäufer bestrahlt.

Item 20

1. (a) Die Japanerin wird sogleich den Kameramann beschmieren.
   (b) Die Japanerin wird soeben von dem Ordnungshüter bekränzt.
2. (a) Die Japanerin wird sogleich den Ordnungshüter bekränzen.
   (b) Die Japanerin wird soeben von dem Kameramann beschmiert.

Item 21

1. (a) Die Rollstuhlfahrerin wird sogleich den Schiedsrichter kostümieren.
   (b) Die Rollstuhlfahrerin wird soeben von dem Chinesen besoldet.
2. (a) Die Rollstuhlfahrerin wird sogleich den Chinesen besolden.
   (b) Die Rollstuhlfahrerin wird soeben von dem Schiedsrichter kostümiert.

Item 22

1. (a) Die Geschäftsfrau wird sofort den Klempner umgürten.
   (b) Die Geschäftsfrau wird derzeit von dem Imker verköstigt.
2. (a) Die Geschäftsfrau wird sofort den Imker verköstigen.
   (b) Die Geschäftsfrau wird derzeit von dem Klempner umgürtet.

Item 23

1. (a) Die Badenixe wird unverzüglich den Skifahrer maskieren.
   (b) Die Badenixe wird im Augenblick von dem Musketier entloht.

2. (a) Die Badenixe wird unverzüglich den Musketier entlohnen.
   (b) Die Badenixe wird im Augenblick von dem Skifahrer maskiert.

Item 24

1. (a) Die Aerobic-Trainerin wird sogleich den Astronaut verwarnen.
   (b) Die Aerobic-Trainerin wird soeben von dem Handwerker bekocht.
2. (a) Die Aerobic-Trainerin wird sogleich den Handwerker bekochen.
   (b) Die Aerobic-Trainerin wird soeben von dem Astronaut verwarnt.

A.3 SENTENCES FOR EXPERIMENT 4

Item 1

1. (a) The ballerina splashed apparently the cellist in the white shirt.
   (b) The ballerina sketched apparently by the fencer splashed the cellist.
   (c) The ballerina being apparently sketched by the fencer splashed the cellist.
   (d) The ballerina will apparently splash the cellist.
2. (a) The ballerina sketched apparently the fencer in the white suit.
   (b) The ballerina splashed apparently by the cellist sketched the fencer.
   (c) The ballerina being apparently splashed by the cellist sketched the fencer.
   (d) The ballerina will apparently sketch the fencer.

Item 2

1. (a) The Amazon stabbed apparently the mechanic in the grey overall.
   (b) The Amazon sprayed apparently by the footballer stabbed the mechanic.
   (c) The Amazon being apparently sprayed by the footballer stabbed the mechanic.
   (d) The Amazon will apparently stab the mechanic.
2. (a) The Amazon sprayed apparently the footballer in the white sweatshirt.
   (b) The Amazon stabbed apparently by the mechanic sprayed the footballer.

(c) The Amazon being apparently stabbed by the mechanic sprayed the footballer.

(d) The Amazon will apparently spray the footballer.

Item 3

1. (a) The journalist pushed seemingly the sportsman in the pink shirt.

   (b) The journalist fed seemingly by the dentist pushed the sportsman.

   (c) The journalist being seemingly fed by the dentist pushed the sportsman.

   (d) The journalist will seemingly push the sportsman.

2. (a) The journalist fed seemingly the dentist in the white gown.

   (b) The journalist pushed seemingly by the sportsman fed the dentist.

   (c) The journalist being seemingly pushed by the sportsman fed the dentist.

   (d) The journalist will seemingly feed the dentist.

Item 4

1. (a) The nurse arrested apparently the sailor in the white uniform.

   (b) The nurse interviewed apparently by the vicar arrested the sailor.

   (c) The nurse being apparently interviewed by the vicar arrested the sailor.

   (d) The nurse will apparently arrest the sailor.

2. (a) The nurse interviewed apparently the vicar in the black suit.

   (b) The nurse arrested apparently by the sailor interviewed the vicar.

   (c) The nurse being apparently arrested by the sailor interviewed the vicar.

   (d) The nurse will apparently interview the vicar.

Item 5

1. (a) The princess cut evidently the pirate in the black vest.

   (b) The princess paid evidently by the golfer cut the pirate.

   (c) The princess being evidently paid by the golfer cut the pirate.

   (d) The princess will evidently cut the pirate.

2. (a) The princess paid evidently the golfer in the green trousers.
   (b) The princess cut evidently by the pirate paid the golfer.
   (c) The princess being evidently cut by the pirate paid the golfer.
   (d) The princess will evidently pay the golfer.

Item 6

1. (a) The construction worker attacked probably the clown in the colourful outfit.
   (b) The construction worker painted probably by the chef attacked the clown.
   (c) The construction worker being probably painted by the chef attacked the clown.
   (d) The construction worker will probably attack the clown.

2. (a) The construction worker painted probably the chef in the white apron.
   (b) The construction worker attacked probably by the clown painted the chef.
   (c) The construction worker being probably attacked by the clown painted the chef.
   (d) The construction worker will probably paint the chef.

Item 7

1. (a) The cleaning lady crowned obviously the dwarf in the red jumper.
   (b) The cleaning lady scorched obviously by the knight crowned the dwarf.
   (c) The cleaning lady being obviously scorched by the knight crowned the dwarf.
   (d) The cleaning lady will obviously crown the dwarf.

2. (a) The cleaning lady scorched obviously the knight in the grey armour.
   (b) The cleaning lady crowned obviously by the dwarf scorched the knight.
   (c) The cleaning lady being obviously crowned by the dwarf scorched the knight.
   (d) The cleaning lady will obviously scorch the knight.

Item 8

1. (a) The ice skater scrubbed seemingly the convict in the striped T-shirt.
   (b) The ice skater poked seemingly by the wizard scrubbed the convict.
   (c) The ice skater being seemingly poked by the wizard scrubbed the convict.

(d) The ice skater will seemingly scrub the convict.

2. (a) The ice skater poked seemingly the wizard in the blue robe.

(b) The ice skater scrubbed seemingly by the convict poked the wizard.

(c) The ice skater being seemingly scrubbed by the convict poked the wizard.

(d) The ice skater will seemingly poke the wizard.

Item 9

1. (a) The maid perfumed probably the gangster in the pinstripe suit.

(b) The maid bandaged probably by the drummer perfumed the gangster.

(c) The maid being probably bandaged by the drummer perfumed the gangster.

(d) The maid will probably perfume the gangster.

2. (a) The maid bandaged probably the drummer in the colourful dress.

(b) The maid perfumed probably by the gangster bandaged the drummer.

(c) The maid being probably perfumed by the gangster bandaged the drummer.

(d) The maid will probably bandage the drummer.

Item 10

1. (a) The tennis player punched clearly the dustman in the red outfit.

(b) The tennis player groomed clearly by the flutist punched the dustman.

(c) The tennis player being clearly groomed by the flutist punched the dustman.

(d) The tennis player will clearly punch the dustman.

2. (a) The tennis player groomed clearly the flautist in the grey trousers.

(b) The tennis player punched clearly by the dustman groomed the flautist.

(c) The tennis player being clearly punched by the dustman groomed the flautist.

(d) The tennis player will clearly groom the flautist.

Item 11

1. (a) The mermaid fanned evidently the teacher in the green jacket.
   (b) The mermaid pinched evidently by the soldier fanned the teacher.
   (c) The mermaid being evidently pinched by the soldier fanned the teacher.
   (d) The mermaid will evidently fan the teacher.
2. (a) The mermaid pinched evidently the soldier in the army uniform.
   (b) The mermaid fanned evidently by the teacher pinched the soldier.
   (c) The mermaid being evidently fanned by the teacher pinched the soldier.
   (d) The mermaid will evidently pinch the soldier.

Item 12

1. (a) The nun injected obviously the violinist in the red shirt.
   (b) The nun photographed obviously by the clarinettist injected the violinist.
   (c) The nun being obviously photographed by the clarinettist injected the violinist.
   (d) The nun will obviously inject the violinist.
2. (a) The nun photographed obviously the clarinettist in the green jacket.
   (b) The nun injected obviously by the violinist photographed the clarinettist.
   (c) The nun being obviously injected by the violinist photographed the clarinettist.
   (d) The nun will obviously photograph the clarinettist.

Item 13

1. (a) The grandmother slapped clearly the detective in the brown coat.
   (b) The grandmother filmed clearly by the musician slapped the detective.
   (c) The grandmother being clearly filmed by the musician slapped the detective.
   (d) The grandmother will clearly slap the detective.
2. (a) The grandmother filmed clearly the musician in the grey suit.
   (b) The grandmother slapped clearly by the detective filmed the musician.
   (c) The grandmother being clearly slapped by the detective filmed the musician.

(d) The grandmother will clearly film the musician.

Item 14

1. (a) The jogger pricked seemingly the hunter in the safari outfit.
   (b) The jogger drenched seemingly by the tourist pricked the hunter.
   (c) The jogger being seemingly drenched by the tourist pricked the hunter.
   (d) The jogger will seemingly prick the hunter.
2. (a) The jogger drenched seemingly the tourist in the holiday apparel.
   (b) The jogger pricked seemingly by the hunter drenched the tourist.
   (c) The jogger being seemingly pricked by the hunter drenched the tourist.
   (d) The jogger will seemingly drench the tourist.

Item 15

1. (a) The fairy prodded probably the athlete in the red shorts.
   (b) The fairy tickled probably by the angler singed the athlete.
   (c) The fairy being probably tickled by the angler singed the athlete.
   (d) The fairy will probably prod the athlete.
2. (a) The fairy tickled probably the angler in the brown vest.
   (b) The fairy prodded probably by the athlete tickled the angler.
   (c) The fairy being probably prodded by the athlete tickled the angler.
   (d) The fairy will probably tickle the angler.

Item 16

1. (a) The cheerleader hunted apparently the doctor in the white suit.
   (b) The cheerleader poisoned apparently by the student hunted the doctor.
   (c) The cheerleader being apparently poisoned by the student hunted the doctor.
   (d) The cheerleader will apparently hunt the doctor.
2. (a) The cheerleader poisoned apparently the student in the blue jeans.
   (b) The cheerleader hunted apparently by the doctor poisoned the student.
   (c) The cheerleader being apparently hunted by the doctor poisoned the student.
   (d) The cheerleader will apparently poison the student.

Item 17

1. (a) The bride clubbed obviously the boy-scout in the green uniform.
   (b) The bride jinxed obviously by the mailman clubbed the boy-scout.
   (c) The bride being obviously jinxed by the mailman clubbed the boy-scout.
   (d) The bride will obviously club the boy-scout.
2. (a) The bride jinxed obviously the mailman in the blue uniform.
   (b) The bride clubbed obviously by the boy-scout jinxed the mailman.
   (c) The bride being obviously clubbed by the boy-scout jinxed the mailman.
   (d) The bride will obviously jinx the mailman.

Item 18

1. (a) The stewardess scalded probably the warrior in the shiny amour.
   (b) The stewardess elbowed probably by the climber scalded the warrior.
   (c) The stewardess being probably elbowed by the climber scalded the warrior.
   (d) The stewardess will probably scald the warrior.
2. (a) The stewardess elbowed probably the climber in the Bavarian outfit.
   (b) The stewardess scalded probably by the warrior elbowed the climber.
   (c) The stewardess being probably scalded by the warrior elbowed the climber.
   (d) The stewardess will probably elbow the climber.

Item 19

1. (a) The witch brushed evidently the paperboy in the blue jeans.
   (b) The witch robbed evidently by the janitor brushed the paperboy.
   (c) The witch being evidently robbed by the janitor brushed the paperboy.
   (d) The witch will evidently brush the paperboy.
2. (a) The witch robbed evidently the janitor in the blue trousers.
   (b) The witch brushed evidently by the paperboy robbed the janitor.

(c) The witch being evidently brushed by the paperboy robbed the janitor.

(d) The witch will evidently rob the janitor.

Item 20

1. (a) The Japanese woman spotlighted seemingly the tailor in the blue shirt.

   (b) The Japanese woman blindfolded seemingly by the sheriff spotlighted the tailor.

   (c) The Japanese woman being seemingly blindfolded by the sheriff spotlighted the tailor.

   (d) The Japanese woman will seemingly spotlight the tailor.

2. (a) The Japanese woman blindfolded seemingly the sheriff in the yellow shirt.

   (b) The Japanese woman spotlighted seemingly by the tailor blindfolded the sheriff.

   (c) The Japanese woman being seemingly spotlighted by the tailor blindfolded the sheriff.

   (d) The Japanese woman will seemingly blindfold the sheriff.

Item 21

1. (a) The cowgirl assaulted evidently the plumber in the blue overall.

   (b) The cowgirl rewarded evidently by the umpire assaulted the plumber.

   (c) The cowgirl being evidently rewarded by the umpire assaulted the plumber.

   (d) The cowgirl will evidently assault the plumber.

2. (a) The cowgirl rewarded evidently the umpire in the striped outfit.

   (b) The cowgirl assaulted evidently by the plumber rewarded the umpire.

   (c) The cowgirl being evidently assaulted by the plumber rewarded the umpire.

   (d) The cowgirl will evidently reward the umpire.

Item 22

1. (a) The businesswoman dressed clearly the pilot in the yellow blouson.

   (b) The businesswoman washed clearly by the waiter dressed the pilot.

   (c) The businesswoman being clearly washed by the waiter dressed the pilot.

   (d) The businesswoman will clearly dress the pilot.

2. (a) The businesswoman washed clearly the waiter in the pink suit.
   (b) The businesswoman dressed clearly by the pilot washed the waiter.
   (c) The businesswoman being clearly dressed by the pilot washed the waiter.
   (d) The businesswoman will clearly wash the waiter.

Item 23

1. (a) The sunbather masked clearly the cosmonaut in the space apparel.
   (b) The sunbather bribed clearly by the musketeer masked the cosmonaut.
   (c) The sunbather being clearly bribed by the musketeer masked the cosmonaut.
   (d) The sunbather will clearly mask the cosmonaut.

2. (a) The sunbather bribed clearly the musketeer in the brown boots.
   (b) The sunbather masked clearly by the cosmonaut bribed the musketeer.
   (c) The sunbather being clearly masked by the cosmonaut bribed the musketeer.
   (d) The sunbather will clearly bribe the musketeer.

Item 24

1. (a) The aerobic instructor scratched obviously the conductor in the black tuxedo.
   (b) The aerobic instructor dried obviously by the carpenter scratched the conductor.
   (c) The aerobic instructor being obviously dried by the carpenter scratched the conductor.
   (d) The aerobic instructor will obviously scratch the conductor.

2. (a) The aerobic instructor dried obviously the carpenter in the blue overall.
   (b) The aerobic instructor scratched obviously by the conductor dried the carpenter.
   (c) The aerobic instructor being obviously scratched by the conductor dried the carpenter.
   (d) The aerobic instructor will obviously dry the carpenter.

A.4 SENTENCES FOR EXPERIMENT 5

Item 1

1. (a) Den Touristen bandagiert der Kneipenwirt.
   (b) Den Touristen bandagiert der Medikus.
   (c) Den Touristen verzaubert der Medikus.
   (d) Den Touristen verköstigt der Kneipenwirt.
2. (a) Den Touristen verköstigt der Medikus.
   (b) Den Touristen verköstigt der Kneipenwirt.
   (c) Den Touristen bespitzelt der Kneipenwirt.
   (d) Den Touristen bandagiert der Medikus.

Item 2

1. (a) Den Piloten bespitzelt der Zauberer.
   (b) Den Piloten bespitzelt der Detektiv.
   (c) Den Piloten verköstigt der Detektiv.
   (d) Den Piloten verzaubert der Zauberer.
2. (a) Den Piloten verzaubert der Detektiv.
   (b) Den Piloten verzaubert der Zauberer.
   (c) Den Piloten bandagiert der Zauberer.
   (d) Den Piloten bespitzelt der Detektiv.

Item 3

1. (a) Den Pagen jagt der Anstreicher.
   (b) Den Pagen jagt der Jaegermeister.
   (c) Den Pagen birgt der Jaegermeister.
   (d) Den Pagen malt der Anstreicher.
2. (a) Den Pagen malt der Jaegermeister.
   (b) Den Pagen malt der Anstreicher.

(c) Den Pagen lehrt der Anstreicher.

(d) Den Pagen jagt der Jaegermeister.

Item 4

1. (a) Den Skater lehrt der Feuerwehrmann.

   (b) Den Skater lehrt der Oberlehrer.

   (c) Den Skater malt der Oberlehrer.

   (d) Den Skater birgt der Feuerwehrmann.

2. (a) Den Skater birgt der Oberlehrer.

   (b) Den Skater birgt der Feuerwehrmann.

   (c) Den Skater jagt der Feuerwehrmann.

   (d) Den Skater lehrt der Oberlehrer.

Item 5

1. (a) Den Indianer bedient der Reporter.

   (b) Den Indianer bedient der Kellner.

   (c) Den Indianer trainiert der Kellner.

   (d) Den Indianer befragt der Reporter.

2. (a) Den Indianer befragt der Kellner.

   (b) Den Indianer befragt der Reporter.

   (c) Den Indianer beraubt der Reporter.

   (d) Den Indianer bedient der Kellner.

Item 6

1. (a) Den Afrikaner beraubt der Trainer.

   (b) Den Afrikaner beraubt der Gangster.

   (c) Den Afrikaner befragt der Gangster.

   (d) Den Afrikaner trainiert der Trainer.

2. (a) Den Afrikaner trainiert der Gangster.
   (b) Den Afrikaner trainiert der Trainer.
   (c) Den Afrikaner bedient der Trainer.
   (d) Den Afrikaner beraubt der Gangster.

Item 7

1. (a) Den Skifahrer unterwirft der Zauberkünstler.
   (b) Den Skifahrer unterwirft der Eroberer.
   (c) Den Skifahrer unterhält der Eroberer.
   (d) Den Skifahrer verwandelt der Zauberkünstler.

2. (a) Den Skifahrer verwandelt der Eroberer.
   (b) Den Skifahrer verwandelt der Zauberkünstler.
   (c) Den Skifahrer verhaftet der Zauberkünstler.
   (d) Den Skifahrer unterwirft der Eroberer.

Item 8

1. (a) Den Sportsmann verhaftet der Harlekin.
   (b) Den Sportsmann verhaftet der Inspektor.
   (c) Den Sportsmann verwandelt der Inspektor.
   (d) Den Sportsmann unterhält der Harlekin.

2. (a) Den Sportsmann unterhält der Inspektor.
   (b) Den Sportsmann unterhält der Harlekin.
   (c) Den Sportsmann erobert der Harlekin.
   (d) Den Sportsmann verhaftet der Inspektor.

Item 9

1. (a) Den Schüler besingt der Ganove.
   (b) Den Schüler besingt der Gitarrist.

(c) Den Schüler bekocht der Gitarrist.

(d) Den Schüler bestiehlt der Ganove.

2. (a) Den Schüler bestiehlt der Gitarrist.

(b) Den Schüler bestiehlt der Ganove.

(c) Den Schüler bekehrt der Ganove.

(d) Den Schüler besingt der Gitarrist.

Item 10

1. (a) Den Jungen bekehrt der Kuechenchef.

(b) Den Jungen bekehrt der Seelenhirt.

(c) Den Jungen bestiehlt der Seelenhirt.

(d) Den Jungen bekocht der Küchenchef.

2. (a) Den Jungen bekocht der Seelenhirt.

(b) Den Jungen bekocht der Küchenchef.

(c) Den Jungen besingt der Küchenchef.

(d) Den Jungen bekehrt der Seelenhirt.

Item 11

1. (a) Den Opa boxt der Kameramann.

(b) Den Opa boxt der Schwergewichtler.

(c) Den Opa kämmt der Schwergewichtler.

(d) Den Opa filmt der Kameramann.

2. (a) Den Opa filmt der Schwergewichtler.

(b) Den Opa filmt der Kameramann.

(c) Den Opa schnappt der Kameramann.

(d) Den Opa boxt der Schwergewichtler.

Item 12

1. (a) Den Renter schnappt der Frisoermeister.
   (b) Den Rentner schnappt der Ordnungshüter.
   (c) Den Rentner filmt der Ordnungshüter.
   (d) Den Rentner kämmt der Frisoermeister.
2. (a) Den Rentner kämmt der Ordnungshüter.
   (b) Den Rentner kämmt der Frisoermeister.
   (c) Den Rentner boxt der Frisoermeister.
   (d) Den Rentner schappt der Ordnungshüter.

Item 13

1. (a) Den Mexikaner beliefert der Hexenmeister.
   (b) Den Mexikaner beliefert der Postbote.
   (c) Den Mexikaner unterweist der Postbote.
   (d) Den Mexikaner vergiftet der Hexenmeister.
2. (a) Den Mexikaner vergiftet der Postbote.
   (b) Den Mexikaner vergiftet der Hexenmeister.
   (c) Den Mexikaner inhaftiert der Hexenmeister.
   (d) Den Mexikaner beliefert der Postbote.

Item 14

1. (a) Den Indonesen inhaftiert der Magister.
   (b) Den Indonesen inhaftiert der Wachtmeister.
   (c) Den Indonesen vergiftet der Wachtmeister.
   (d) Den Indonesen unterweist der Magister.
2. (a) Den Indonesen unterweist der Wachtmeister.
   (b) Den Indonesen unterweist der Magister.

(c) Den Indonesen beliefert der Magister.

(d) Den Indonesen inhaftiert der Wachtmeister.

Item 15

1. (a) Den Stallknecht bewirft der Pater.

   (b) Den Stallknecht bewirft der Sportfreund.

   (c) Den Stallknecht skizziert der Sportfreund.

   (d) Den Stallknecht segnet der Pater.

2. (a) Den Stallknecht segnet der Sportfreund.

   (b) Den Stallknecht segnet der Pater.

   (c) Den Stallknecht beschenkt der Pater.

   (d) Den Stallknecht bewirft der Sportfreund.

Item 16

1. (a) Den Hausmann beschenkt der Zeichner.

   (b) Den Hausmann beschenkt der Weihnachtsmann.

   (c) Den Hausmann segnet der Weihnachtsmann.

   (d) Den Hausmann skizziert der Zeichner.

2. (a) Den Hausmann skizziert der Weihnachtsmann.

   (b) Den Hausmann skizziert der Zeichner.

   (c) Den Hausmann bewirft der Zeichner.

   (d) Den Hausmann beschenkt der Weihnachtsmann.

Item 17

1. (a) Den Gärtner krönt der Schülerlotse.

   (b) Den Gärtner krönt der Wuerdentraeger.

   (c) Den Gärntner spritzt der Wuerdentraeger.

   (d) Den Gärtner stoppt der Schülerlotse.

2. (a) Den Gärtner stoppt der Wuerdentraeger.
   (b) Den Gärtner stoppt der Schülerlotse.
   (c) Den Gärtner haut der Schülerlotse.
   (d) Den Gärtner krönt der Wuerdentraeger.

Item 18

1. (a) Den Diener haut der Sanitäter.
   (b) Den Diener haut der Randalierer.
   (c) Den Diener stoppt der Randalierer.
   (d) Den Diener spritzt der Sanitäter.

2. (a) Den Diener spritzt der Randalierer.
   (b) Den Diener spritzt der Sanitäter.
   (c) Den Diener krönt der Sanitäter.
   (d) Den Diener haut der Randalierer.

Item 19

1. (a) Den Manager überfällt der Schneider.
   (b) Den Manager überfällt der Bandit.
   (c) Den Manager bewirtet der Bandit.
   (d) Den Manager bekleidet der Schneider.

2. (a) Den Manager bekleidet der Bandit.
   (b) Den Manager bekleidet der Schneider.
   (c) Den Manager kommandiert der Schneider.
   (d) Den Manager überfällt der Bandit.

Item 20

1. (a) Den Handwerker kommandiert der Hotelier.
   (b) Den Handwerker kommandiert der Leutnant.

(c) Den Handwerker bekleidet der Leutnant.

(d) Den Handwerker bewirtet der Hotelier.

2. (a) Den Handwerker bewirtet der Leutnant.

(b) Den Handwerker bewirtet der Hotelier.

(c) Den Handwerker überfällt der Hotelier.

(d) Den Handwerker kommandiert der Leutnant.

Item 21

1. (a) Den Turner interviewt der Wächter.

(b) Den Turner interviewt der Schreiber.

(c) Den Turner vermöbelt der Schreiber.

(d) Den Turner salutiert der Wächter.

2. (a) Den Turner salutiert der Schreiber.

(b) Den Turner salutiert der Wächter.

(c) Den Turner verknastet der Wächter.

(d) Den Turner interviewt der Schreiber.

Item 22

1. (a) Den Tänzer verknastet der Schlaeger.

(b) Den Tänzer verknastet der Amtsrichter.

(c) Den Tänzer salutiert der Amtsrichter.

(d) Den Tänzer vermöbelt der Schlaeger.

2. (a) Den Tänzer vermöbelt der Richter.

(b) Den Tänzer vermöbelt der Schlaeger.

(c) Den Tänzer interviewt der Schlaeger.

(d) Den Tänzer verknastet der Richter.

Item 23

1. (a) Den Pfadfinder verarztet der Spion.
   (b) Den Pfadfinder verarztet der Hausarzt.
   (c) Den Pfadfinder attackiert der Hausarzt.
   (d) Den Pfadfinder beschattet der Spion.
2. (a) Den Pfadfinder beschattet der Hausarzt.
   (b) Den Pfadfinder beschattet der Spion.
   (c) Den Pfadfinder frisiert der Spion.
   (d) Den Pfadfinder verarztet der Hausarzt.

Item 24

1. (a) Den Wanderer frisiert der Ritter.
   (b) Den Wanderer frisiert der Coiffeur.
   (c) Den Wanderer beschattet der Coiffeur.
   (d) Den Wanderer attackiert der Ritter.
2. (a) Den Wanderer attackiert der Coiffeur.
   (b) Den Wanderer attackiert der Ritter.
   (c) Den Wanderer verarztet der Ritter.
   (d) Den Wanderer frisiert der Coiffeur.

Appendix B

Item Images

Below we provide the item images for Experiments 1 to 5 of the present thesis. The corresponding item sentences are listed in Appendix A. It should be noted that while images are reproduced in black-and-white, the images employed in the actual experiments were colour images.

B.1 Images for Experiments 1 and 3

Item 1

Image 1

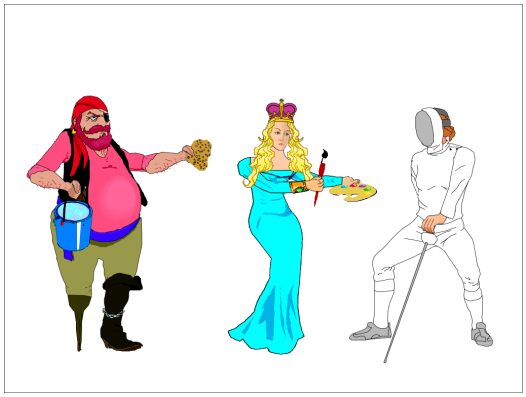


Image 2

Item 2

Image 1

Image 2

Item 3

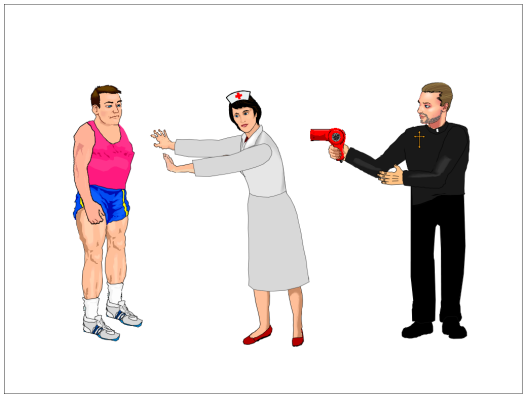


Image 1

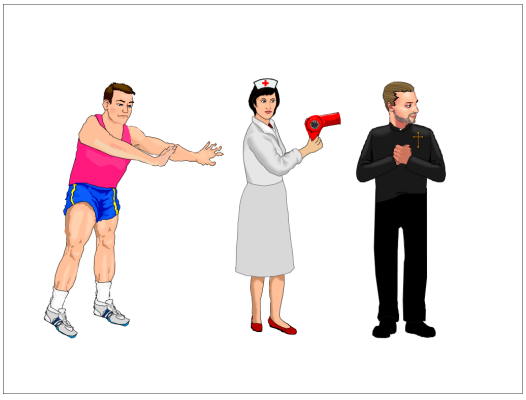


Image 2

Item 4

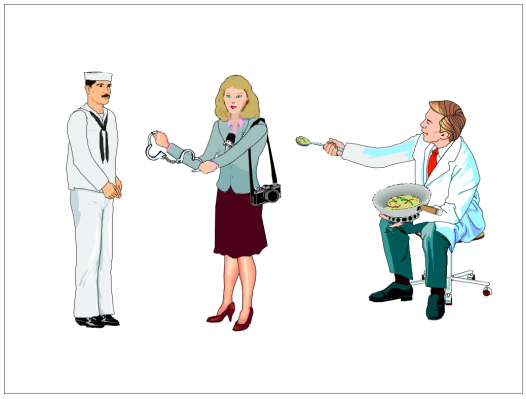


Image 1

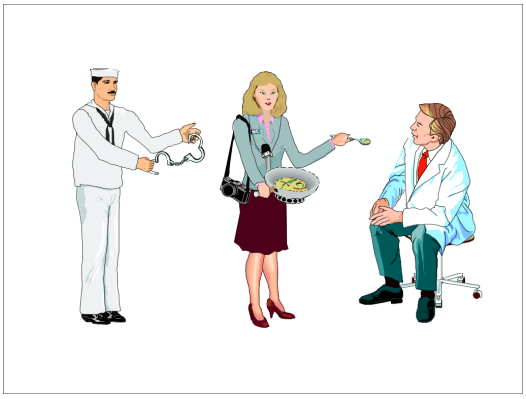


Image 2

Item 5

Image 1

Image 2

Item 6

Image 1

Image 2

Item 7

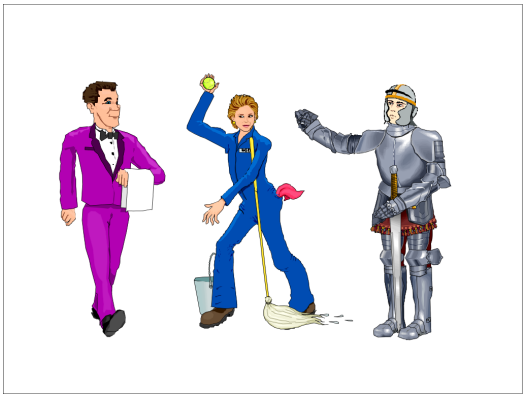


Image 1

Image 2

Item 8

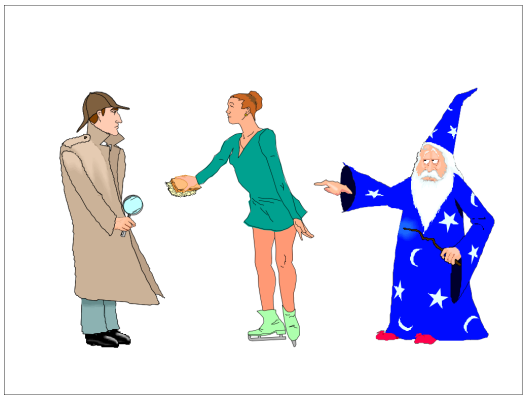


Image 1

Image 2

Item 9

Image 1

Image 2

Item 10

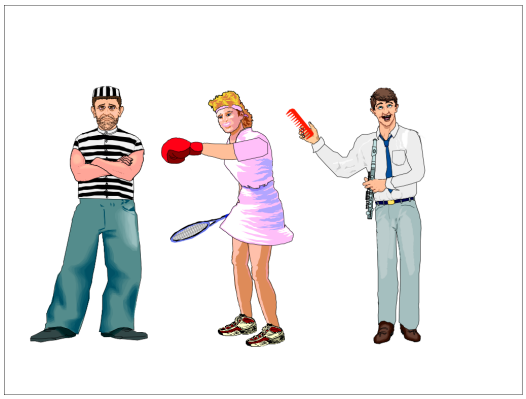


Image 1

Image 2

Item 11

Image 1

Image 2

Item 12

Image 1

Image 2

Item 13

Image 1

Image 2

Item 14

Image 1

Image 2

Item 15

Image 1

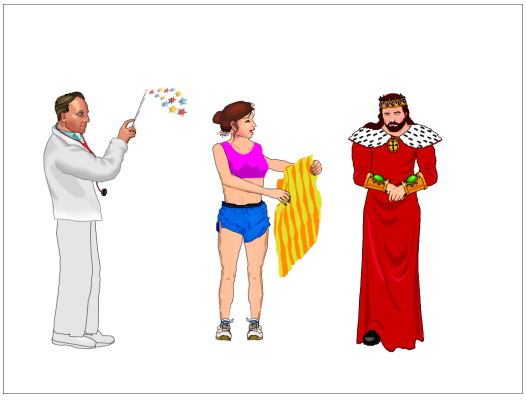


Image 2

Item 16

Image 1

Image 2

Item 17

Image 1

Image 2

Item 18

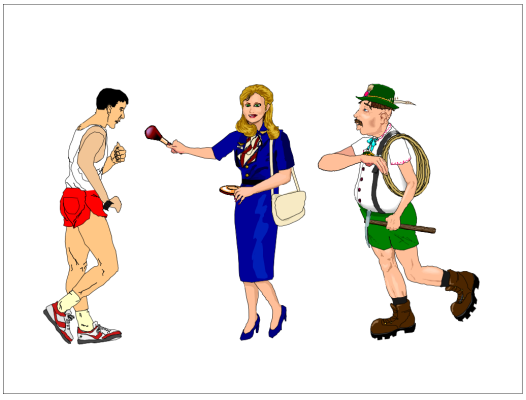


Image 1

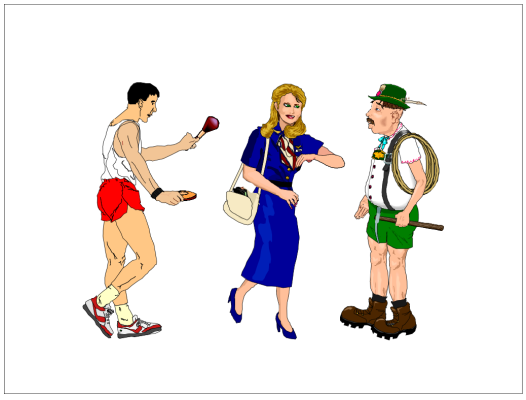


Image 2

Item 19

Image 1

Image 2

Item 20

Image 1

Image 2

Item 21

Image 1

Image 2

Item 22

Image 1

Image 2

Item 23

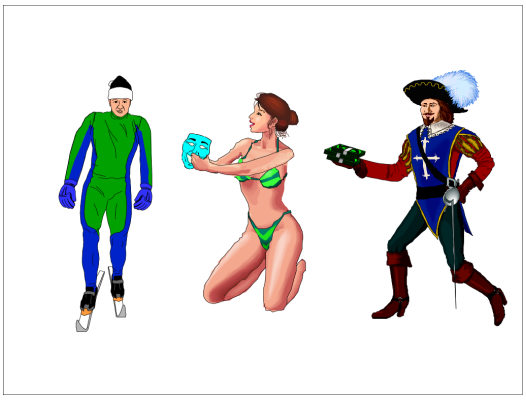


Image 1

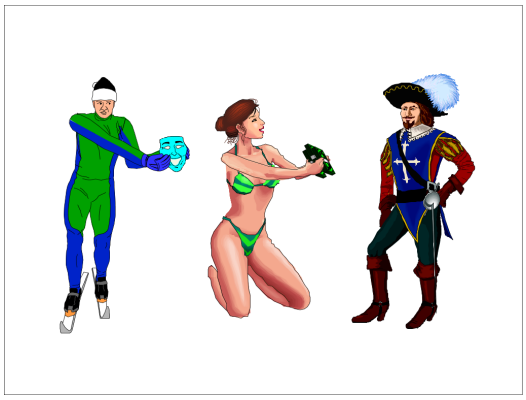


Image 2

Item 24

Image 1

Image 2

B.2 IMAGES FOR EXPERIMENT 2

Item 1

Image 1

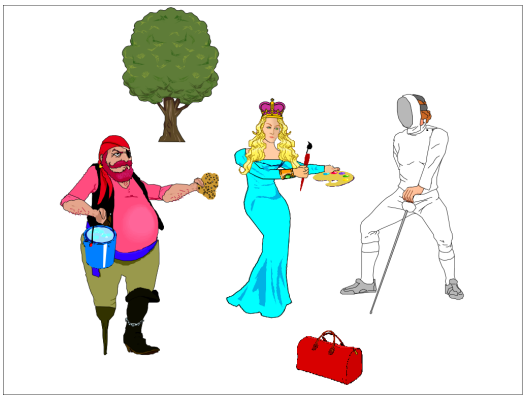


Image 2

Item 2

Image 1

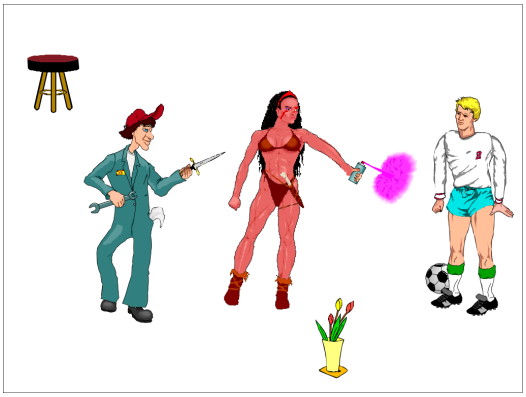


Image 2

Item 3

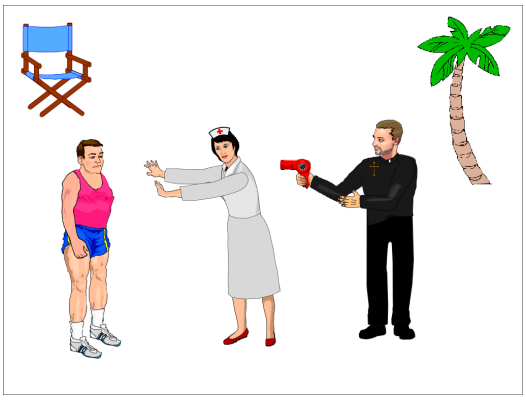


Image 1

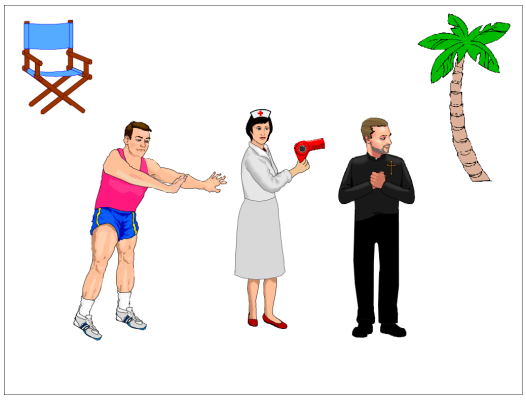


Image 2

Item 4

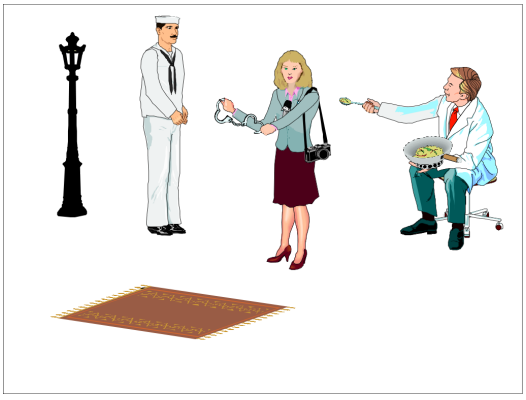


Image 1

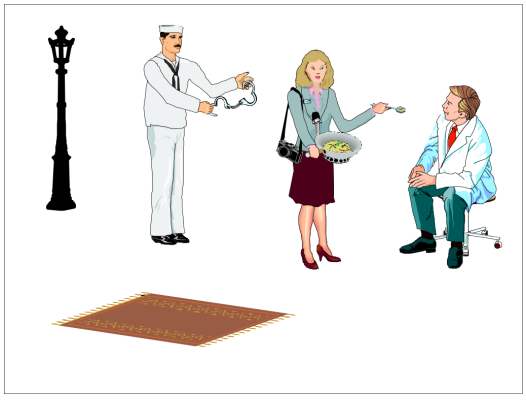


Image 2

Item 5

Image 1

Image 2

Item 6

Image 1

Image 2

Item 7

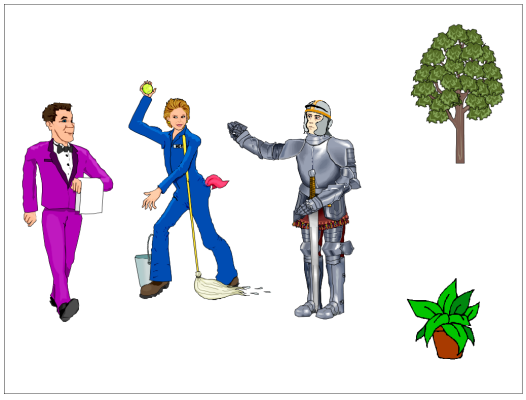


Image 1

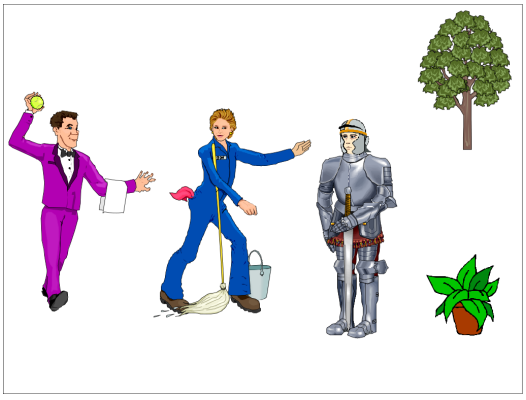


Image 2

Item 8

Image 1

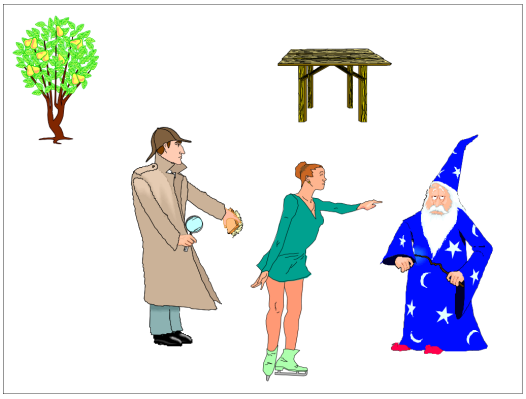


Image 2

Item 9

Image 1

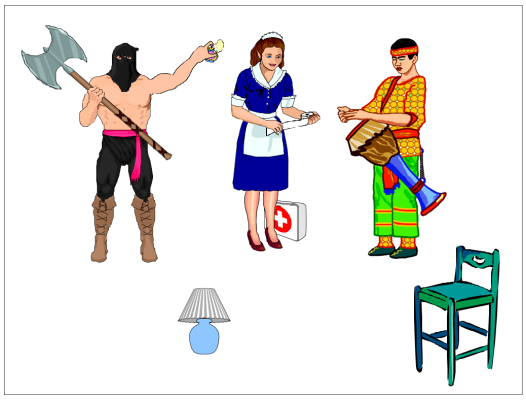


Image 2

Item 10

Image 1

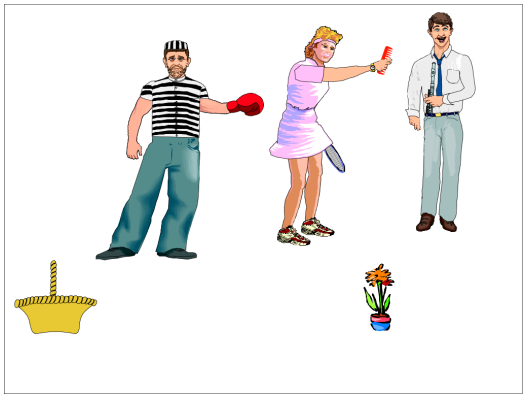


Image 2

Item 11

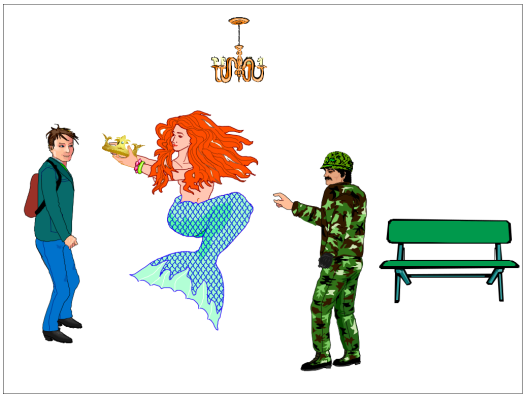


Image 1

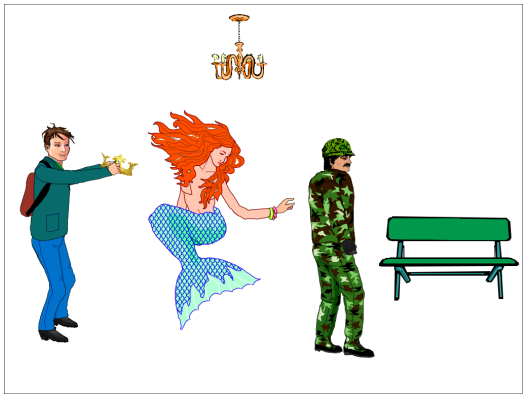


Image 2

Item 12

Image 1

Image 2

Item 13

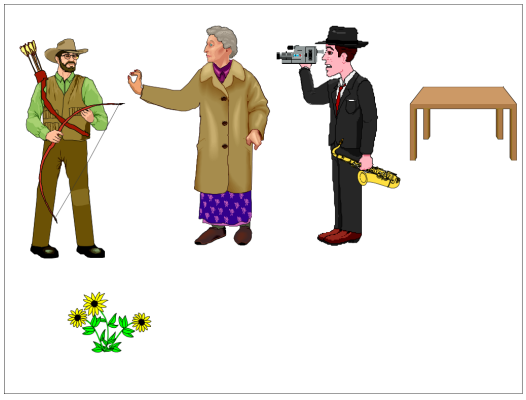


Image 1

Image 2

Item 14

Image 1

Image 2

Item 15

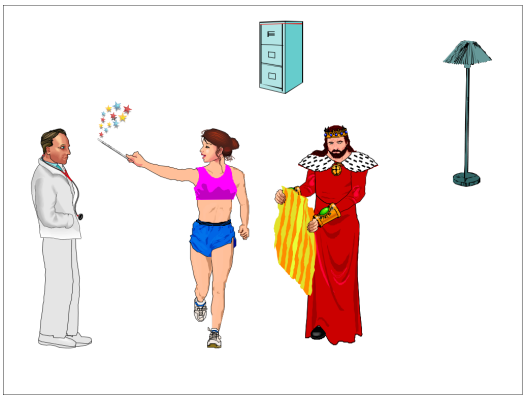


Image 1

Image 2

Item 16

Image 1

Image 2

Item 17

Image 1

Image 2

Item 18

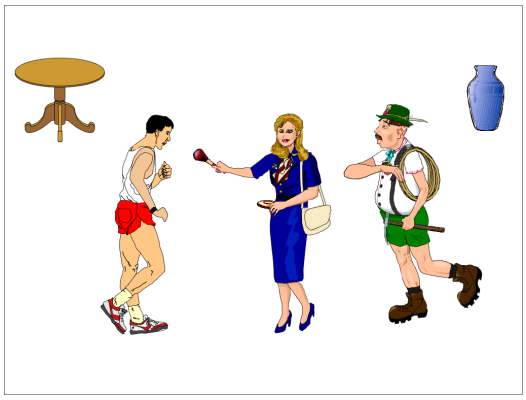


Image 1

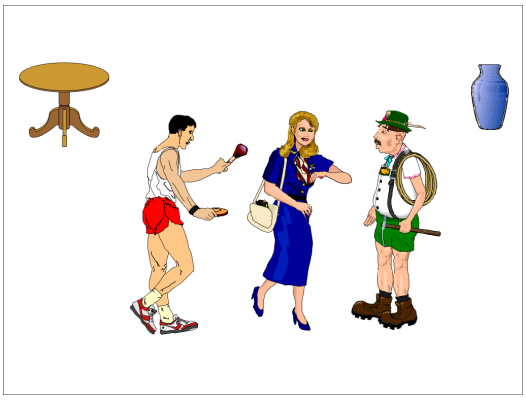


Image 2

Item 19

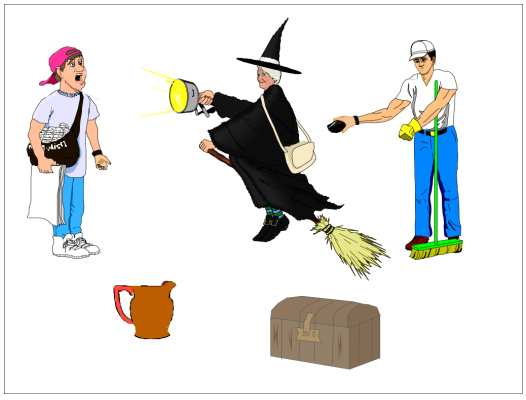


Image 1

Image 2

Item 20

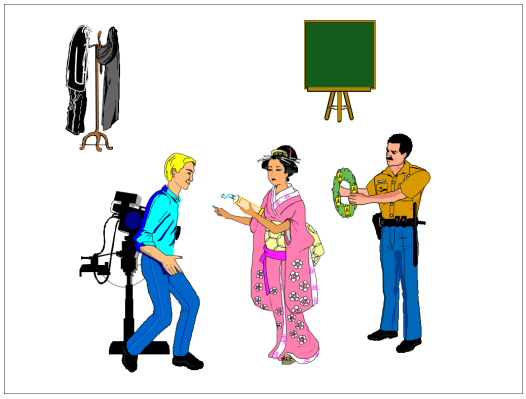


Image 1

Image 2

Item 21

Image 1

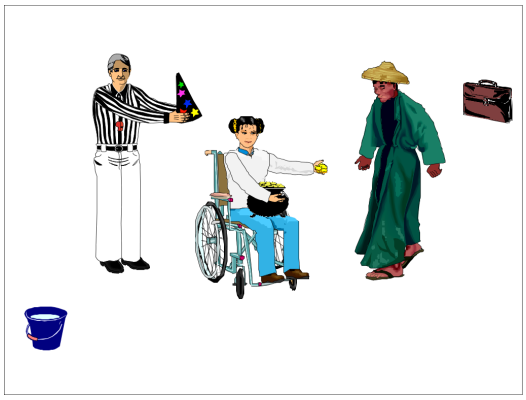


Image 2

Item 22

Image 1

Image 2

Item 23

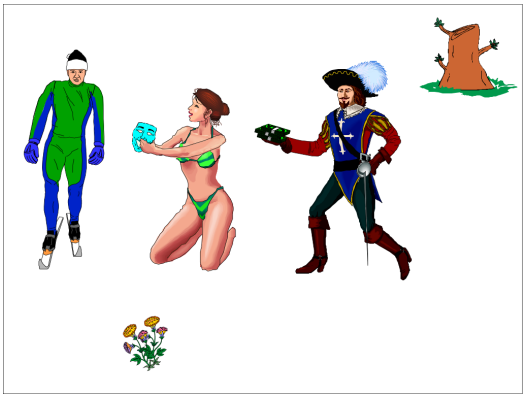


Image 1

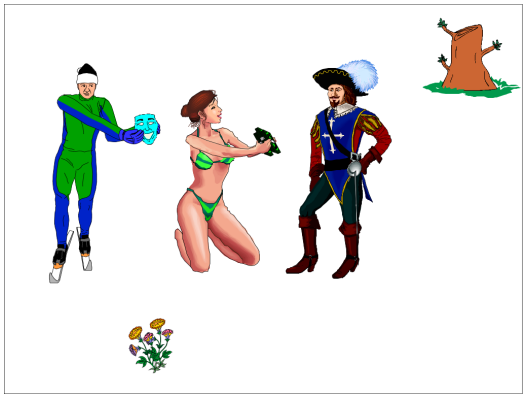


Image 2

Item 24

Image 1

Image 2

B.3 IMAGES FOR EXPERIMENT 4

Item 1

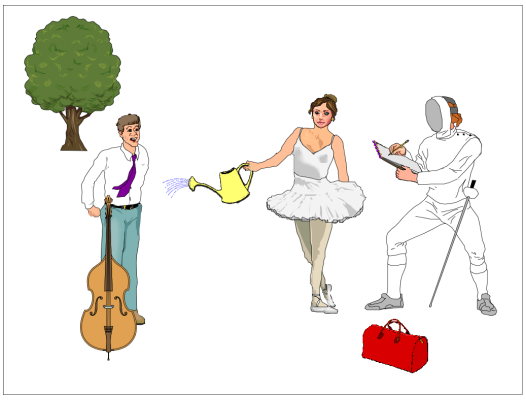


Image 1

Image 2

Item 2

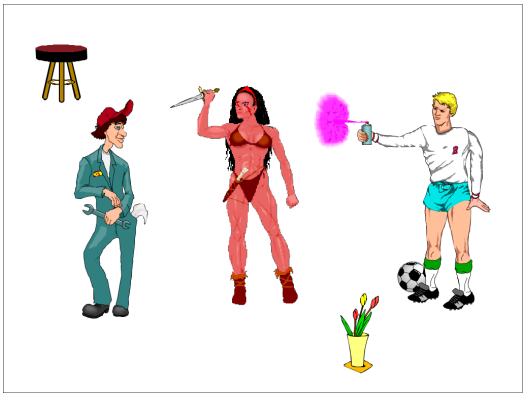


Image 1

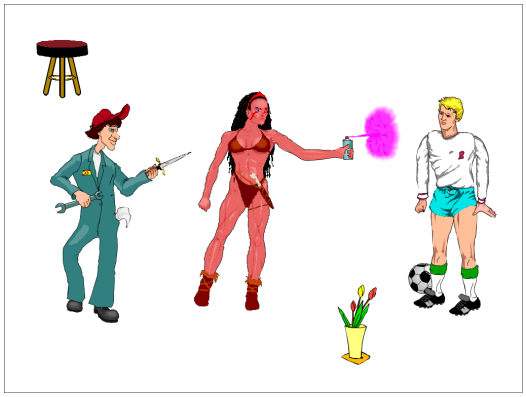


Image 2

Item 3

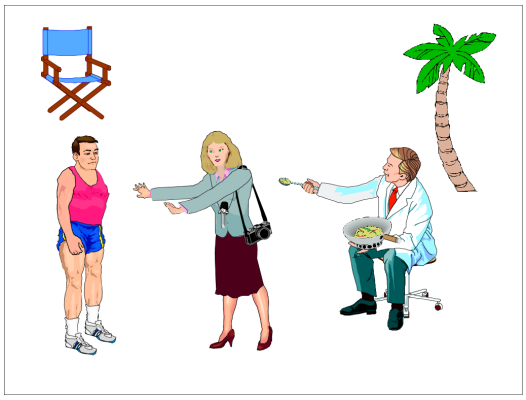


Image 1

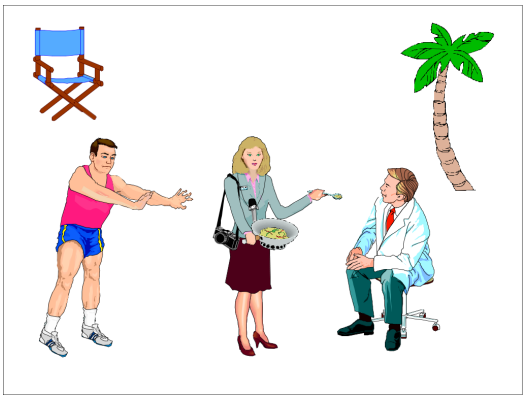


Image 2

Item 4

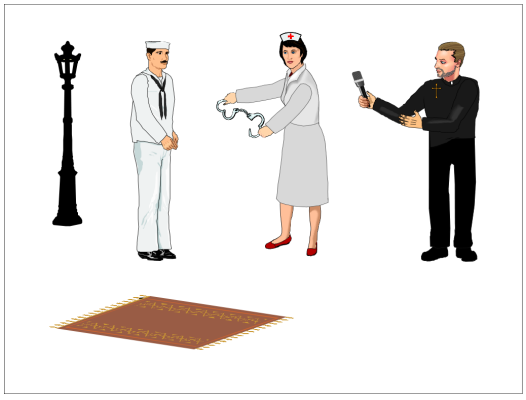


Image 1

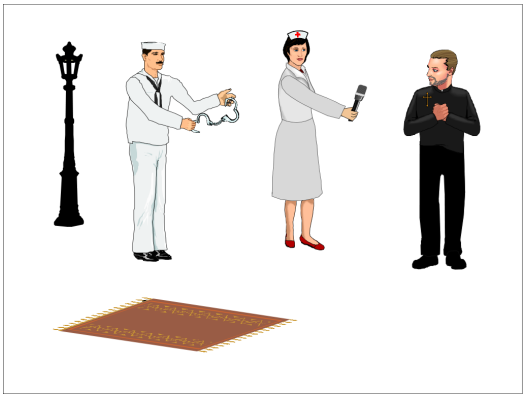


Image 2

Item 5

Image 1

Image 2

Item 6

Image 1

Image 2

Item 7

Image 1

Image 2

Item 8

Image 1

Image 2

Item 9

Image 1

Image 2

Item 10

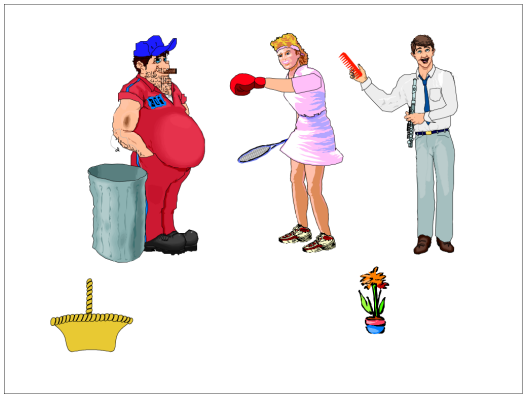


Image 1

Image 2

Item 11

Image 1

Image 2

Item 12

Image 1

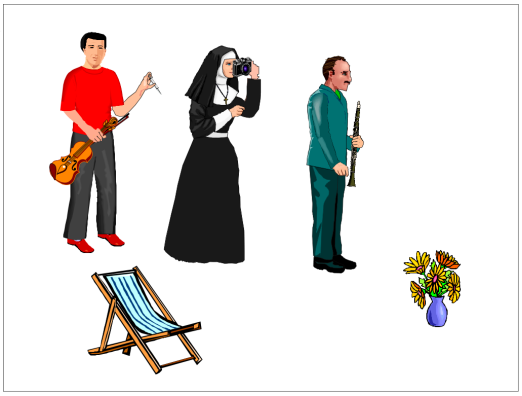


Image 2

Item 13

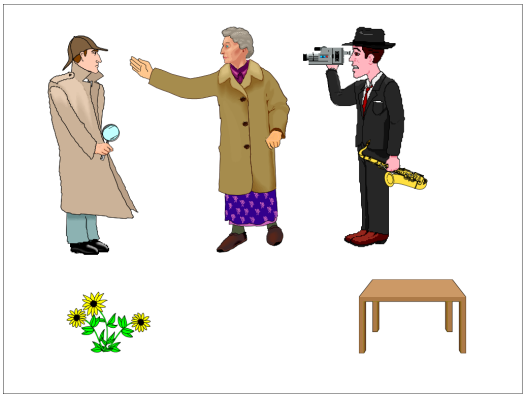


Image 1

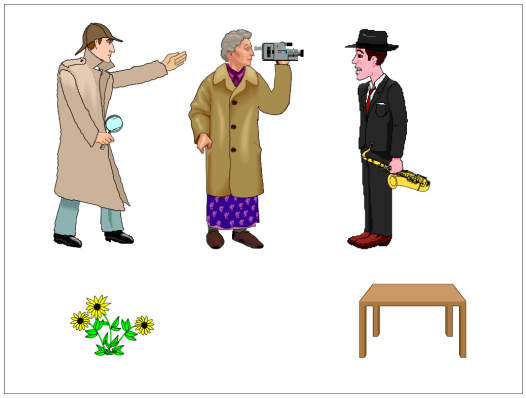


Image 2

Item 14

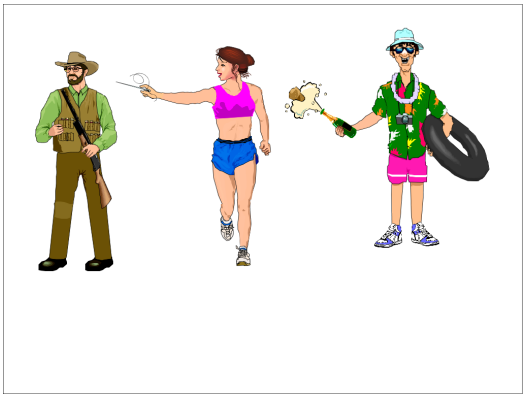


Image 1

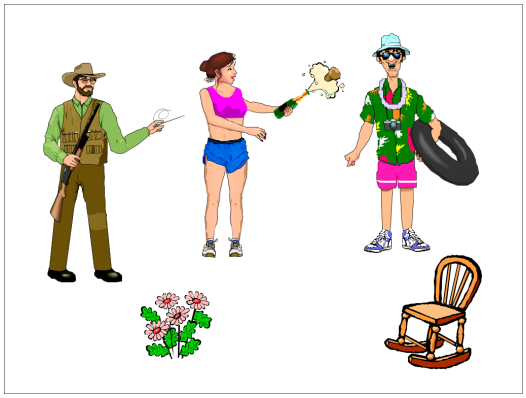


Image 2

Item 15

Image 1

Image 2

Item 16

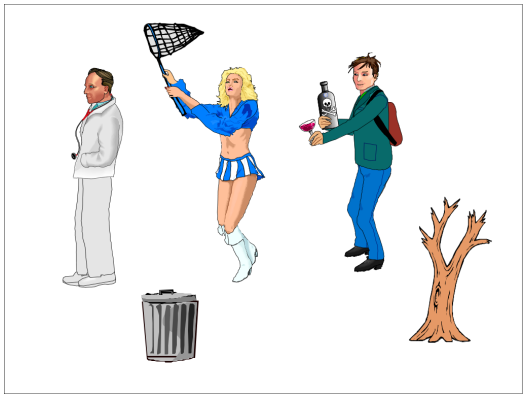


Image 1

Image 2

Item 17

Image 1

Image 2

Item 18

Image 1

Image 2

Item 19

Image 1

Image 2

Item 20

Image 1

Image 2

Item 21

Image 1

Image 2

Item 22

Image 1

Image 2

Item 23

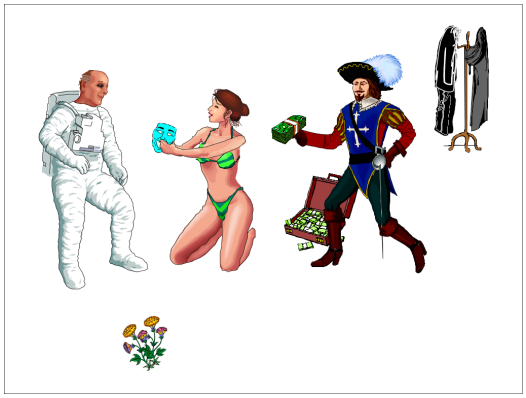


Image 1

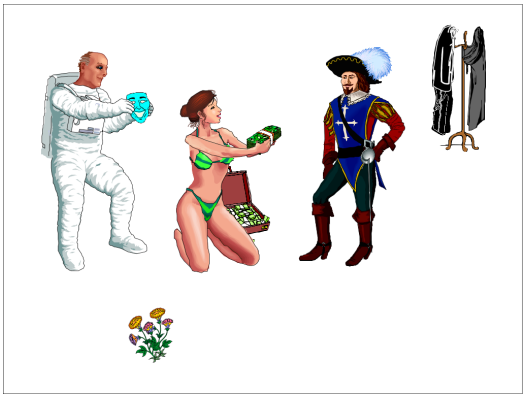


Image 2

Item 24

Image 1

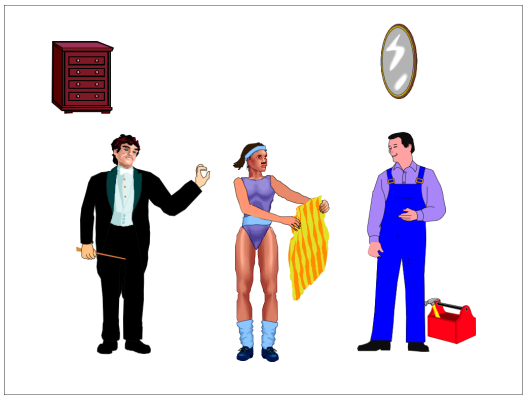


Image 2

B.4 IMAGES FOR EXPERIMENT 5

Item 1

Image 1

Image 2

Item 2

Image 1

Image 2

Item 3

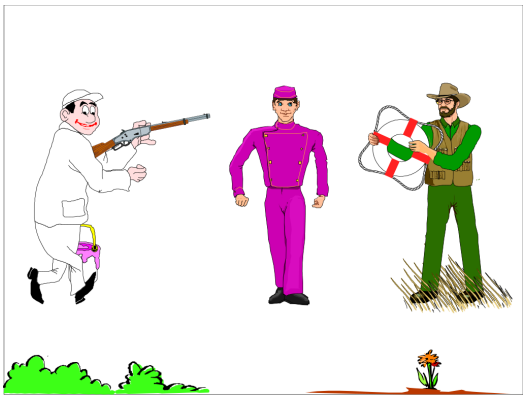


Image 1

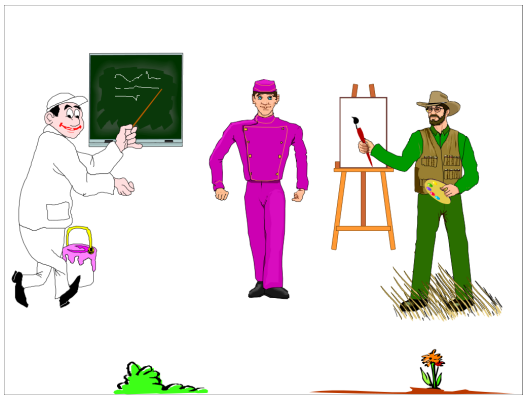


Image 2

Item 4

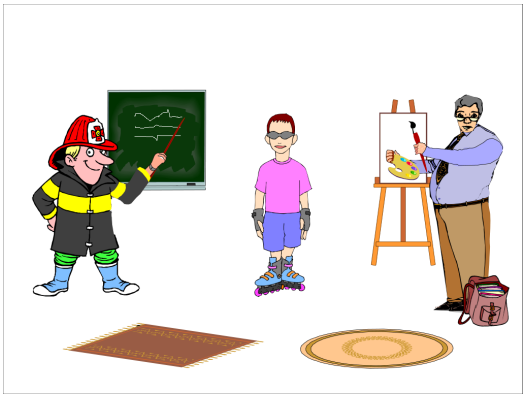


Image 1

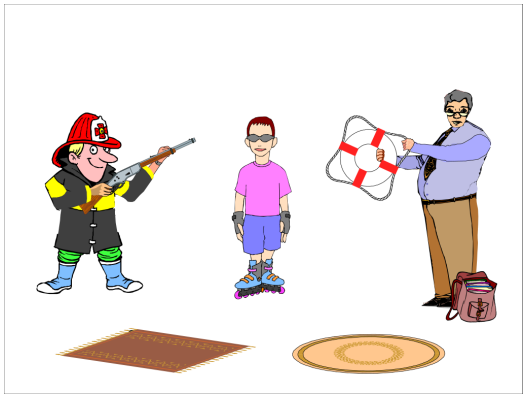


Image 2

Item 5

Image 1

Image 2

Item 6

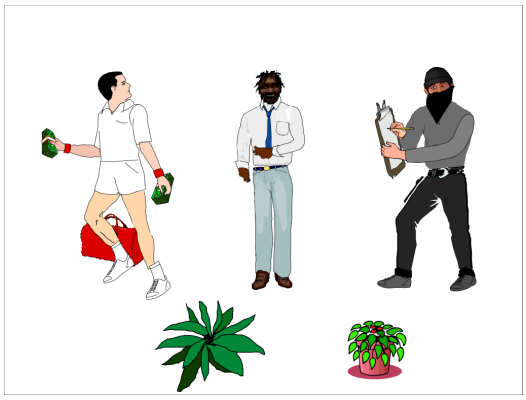


Image 1

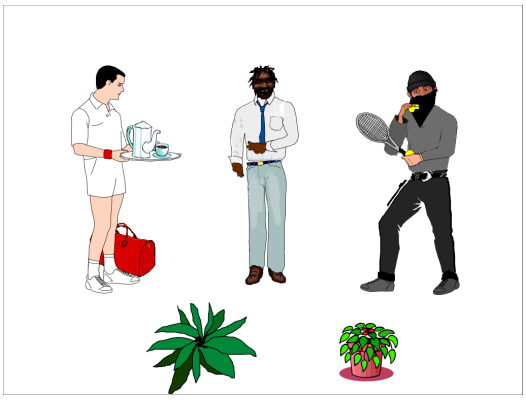


Image 2

Item 7

Image 1

Image 2

Item 8

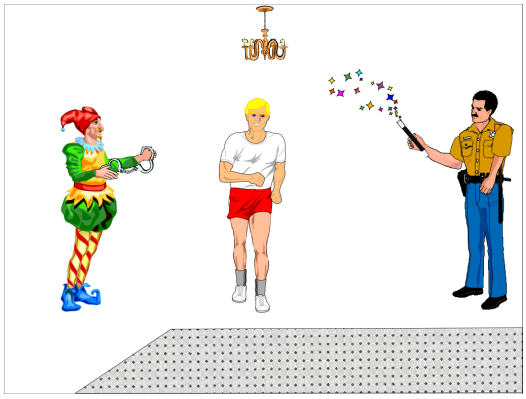


Image 1

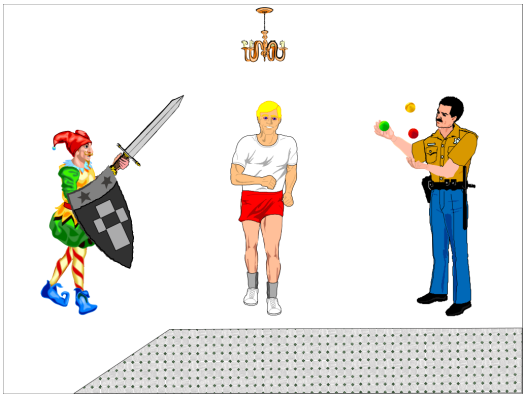


Image 2

Item 9

Image 1

Image 2

Item 10

Image 1

Image 2

Item 11

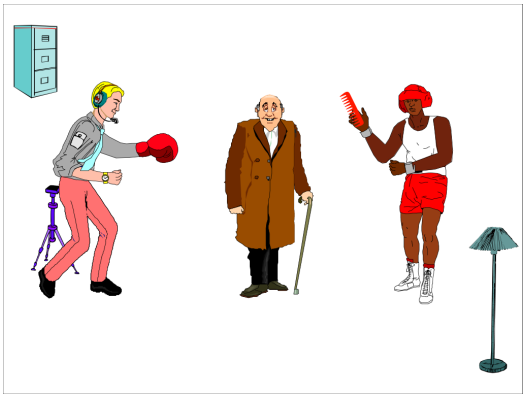


Image 1

Image 2

Item 12

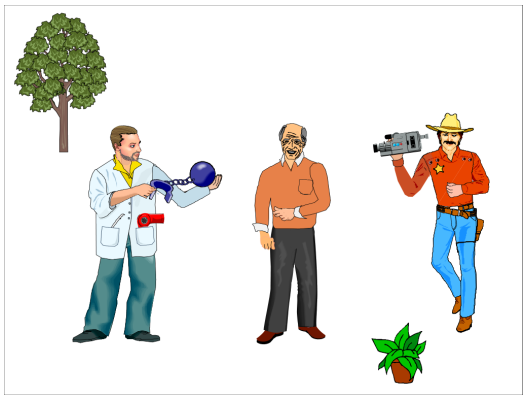


Image 1

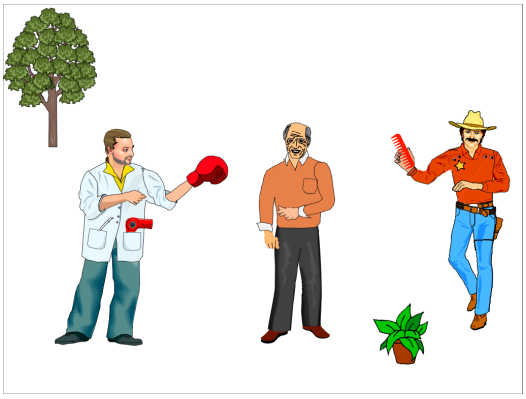


Image 2

Item 13

Image 1

Image 2

Item 14

Image 1

Image 2

Item 15

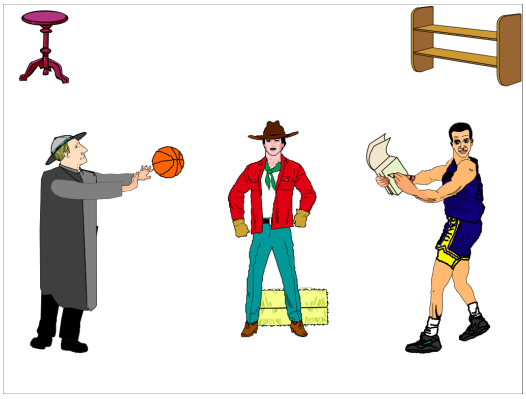


Image 1

Image 2

Item 16

Image 1

Image 2

Item 17

Image 1

Image 2

Item 18

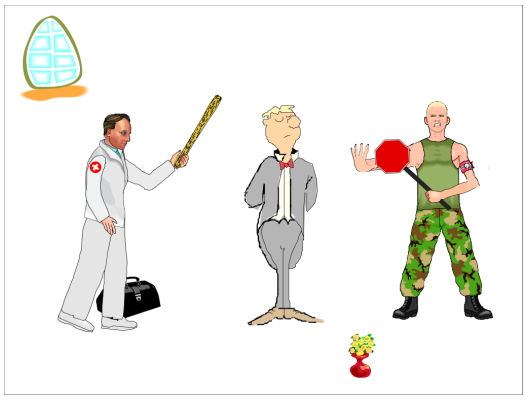


Image 1

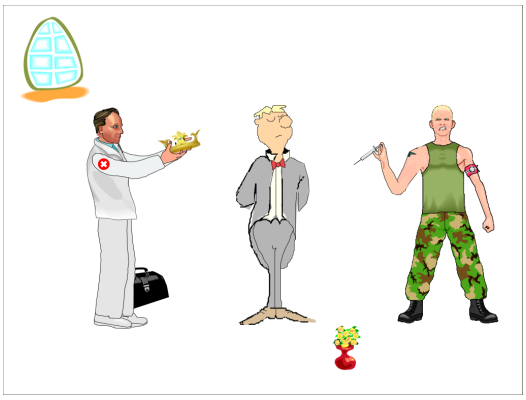


Image 2

Item 19

Image 1

Image 2

Item 20

Image 1

Image 2

Item 21

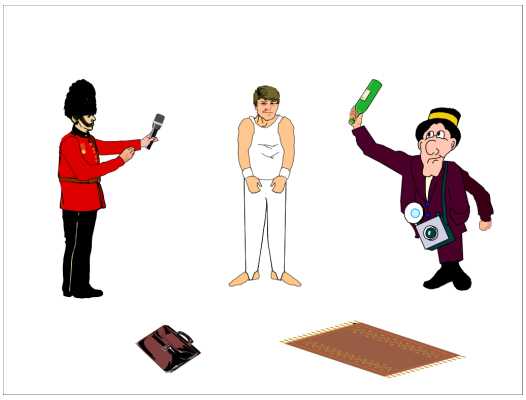


Image 1

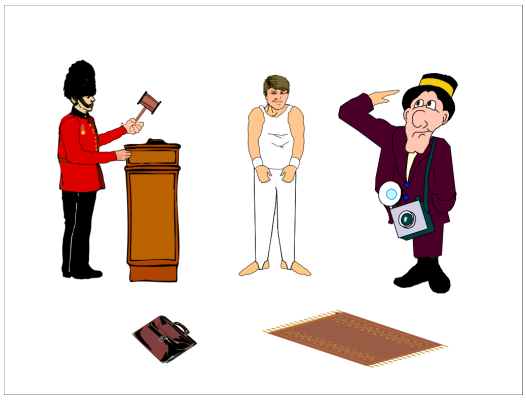


Image 2

Item 22

Image 1

Image 2

Item 23

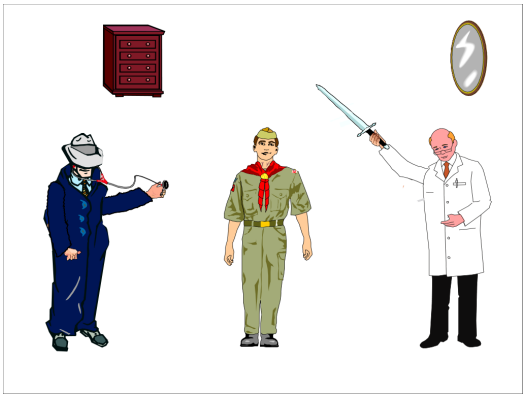


Image 1

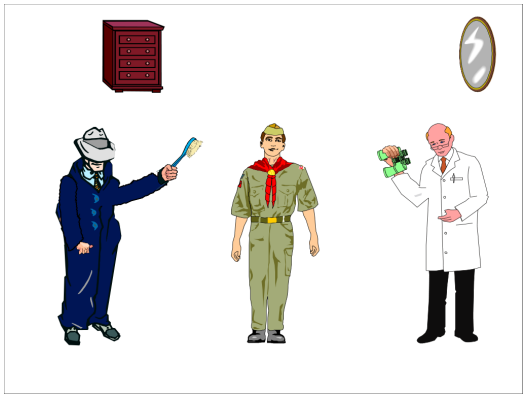


Image 2

Item 24

Image 1

Image 2

References

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 at verbs: restricting the domain of subsequent reference. *Cognition*, *73*, 247–264.

Altmann, G. T. M., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, *30*, 191–238.

Anderson, J. R. (1990). *The adaptive character of thought.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Baayen, R. H., Pipenbrock, R., & Gulikers, L. (1995). *The celex lexical database (cd-rom).* University of Pennsylvania, Philadelphia, PA: Linguistic Data Consortium.

Bergen, B., Narayan, S., & Feldman, J. (2003). Embodied verbal semantics: evidence from an image-verb matching task. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the 25th Annual Conference of the Cognitive Science Society* (pp. 139–144). Boston, Massachusetts, USA: Cognitive Science Society.

Bever, T. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and the development of language* (pp. 279–362). New York: Wiley.

Buccino, G., Binkofsky, F., Fink, G. R., Fadiga, L., Fogassi, L., Gallese, V., et al. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *European Journal of Neuroscience*, *13*, 400–404.

Carroll, P. J., Young, J. R., & Guertin, M. S. (1992). Visual analysis of cartoons: a view from the far side. In K. Rayner (Ed.), *Eye-movements and visual cognition: scene perception and reading* (pp. 444–461). New York: Springer-Verlag.

Chambers, C. G., Tanenhaus, M. K., Filip, H., & Carlson, G. N. (2002). Circumscribing referential domains during real time language comprehension. *Journal of Memory and Language*, *47*, 30–49.

Chater, N., Crocker, M. W., & Pickering, M. J. (1998). The rational analysis of inquiry: The case for parsing. In N. Chater & M. Oaksford (Eds.), *Rational analysis of cognition* (pp. 441–468). Oxford: Oxford University Press.

Coltheart, M. (1999). Modularity and cognition. *Trends in Cognitive Science*, *3*, 115–120.

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.

Crain, S., & Steedman, M. (1985). On not being led up the garden path: the use of context by the psychological parser. In D. Dowty, L. Karttunnen, & A. Zwicky (Eds.), *Natural language parsing* (pp. 320–358). Cambridge, MA: Cambridge University Press.

Crocker, M. W. (1996). *Computational psycholinguistics: An interdisciplinary approach to the study of language*. Dordrecht: Kluwer.

Crocker, M. W. (1999). Mechanisms for sentence processing. In S. Garrod & M. J. Pickering (Eds.), *Language processing* (pp. 191–232). London: Psychology Press.

Crystal, D. (2003). *A dictionary of linguistics and phonetics* (5th ed.). Oxford: Blackwell.

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., Swingley, D., Tanenhaus, M. K., & Magnuson, J. S. (2000). Linguistic gender and spoken-word recognition in French. *Journal of Memory and Language*, *42*, 465–480.

Davidson, D. (1967). Truth and meaning. *Synthese*, *XVII*, 304–323.

Davidson, D. (1986a). The individuation of events. In D. Davidson (Ed.), *Essays on actions and events* (pp. 163–180). Oxford, NY: Oxford University Press.

Davidson, D. (1986b). The logical form of action sentences. In D. Davidson (Ed.), *Essays on actions and events* (pp. 105–122). Oxford, NY: Oxford University Press.

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

Feldman, J., & Narayanan, S. (2003). Embodied meaning in a neural theory of language. *Brain and Language*, *89*, 385–392.

Ferreira, F. (2003). The misinterpretation of canonical sentences. *Cognitive Psychology*, *47*, 164-203.

Ferretti, T. R., McRae, K., & Hatherell, A. (2001). Integrating verbs, situation schemas, and thematic role concepts. *Journal of Memory and Language*, *44*, 516–547.

Fillmore, C. J. (1968). The case for case. In E. Bach & R. T. Harms (Eds.), *Universals in linguistic theory* (pp. 1–88). New York: Holt, Reinhart and Winston.

Fillmore, C. J. (1982). Frame semantics. In The Linguistic Society of Korea (Ed.), *Linguistics in the morning calm* (pp. 111–137). Seoul: Hanshin Pub. Co.

Fillmore, C. J. (1985). Frames and the semantics of understanding. *Quaderni di Semantica*, *6.2*, 222–254.

Fodor, J. (1983). *Modularity of mind*. Cambridge, MA: MIT Press.

Foley, W. A., & Van Valin, R. D. (1984). *Functional syntax and universal grammar*. Cambridge: Cambridge University Press.

Forster, K. (1979). Levels of processing and the structure of the language processor. In W. E. Cooper & E. C. T. Waler (Eds.), *Sentence processing: psycholinguistic studies presented to Merrill Garrett* (pp. 27–85). Hillsdale, NJ: Lawrence Erlbaum.

Frazier, L. (1987). Syntactic processing: evidence from Dutch. *Natural Language and Linguistic Theory*, *5*, 505–526.

Frazier, L., & Clifton, C. (1996). *Construal*. Cambridge, MA: MIT Press.

Frazier, L., & Fodor, J. D. (1979). The sausage machine: a new two-stage parsing model. *Cognition*, *6*, 291–325.

Garrod, S., & Pickering, M. J. (2003). Linguistics fit for dialogue. *Brain and Behavioural Sciences*, *26*, 678–678.

Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition*, *68*, 1–76.

Gibson, J. (1966). *The senses considered as perceptual systems*. Boston, MA: Houghton-Mifflin Co.

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, *1*, 3–55.

Gorrell, P. (1995). *Syntax and parsing*. Cambridge: Cambridge University Press.

Griffin, Z., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, *11*, 274–279.

Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language*, *49*, 43–61.

Hemforth, B. (1993). *Kognitives Parsing: Repräsentation und Verarbeitung sprachlichen Wissens*. Sankt Augustin: Infix-Verlag.

Henderson, J. M., & Ferreira, F. (2004). Scene perception for psycholinguists. In J. M.

Henderson & F. Ferreira (Eds.), *The interface of language, vision, and action: eye movements and the visual world.* New York: Psychology Press.

Howell, D. C. (2002). *Statistical methods for psychology* (5th ed.). Pacific Grove: Duxbury.

Huettig, F., & Altmann, G. T. M. (2004). The online processing of ambiguous and unambiguous words in context: Evidence from head-mounted eye-tracking. In M. Carreiras & C. Clifton (Eds.), *The on-line study of sentence comprehension: eyetracking,ERP and beyond.* New York: Psychology Press.

Hyönä, J., & Hujanen, H. (1997). Effects of case marking and word order on sentence parsing in Finnish: an eye fixation analysis. *Quarterly Journal of Experimental Psychology*, *50A*, 841–858.

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

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

Jackendoff, R. (1997). *The architecture of the language faculty.* Cambridge, MA: MIT Press.

Jackendoff, R. (2002). *Foundations of language: brain, meaning, grammar, evolution.* Oxford, UK: Oxford University Press.

Jackendoff, R. (2003). Precis of: Foundations of language: Brain, meaning, grammar, evolution. *Brain and Behavioural Sciences*, *26*, 651–665.

Kaiser, E., & Trueswell, J. C. (in press). The role of discourse context in the processing of a flexible word-order language. *Cognition.*

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.

Kamide, Y., Scheepers, C., & Altmann, G. T. M. (2003). Integration of syntactic and semantic information in predictive processing: cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research*, *32*, 37–55.

Katz, J. J. (1996). Semantics in linguistics and philosophy: an intensionalist perspective. In S. Lappin (Ed.), *The handbook of contemporary semantic theory* (pp. 599–651). Oxford: Blackwell.

Kimball, J. (1973). Seven principles of surface structure parsing in natural languages. *Cognition*, *2*, 15–47.

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

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*, 676–703.

MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behaviour*, *23*, 127–150.

Martin, A., Haxby, J. V., Lalonde, F. M., Wiggs, C. L., & Ungerleider, L. G. (1995). Discrete cortical regions associated with knowledge of colour and knowledge of action. *Science*, *270*, 102–105.

Matin, E., Shao, K., & Boff, K. (1993). Saccadic overhead: information processing time with and without saccades. *Perceptual Psychophysics*, *53*, 372–380.

McRae, K., Ferretti, T. R., & Amyote, L. (1997). Thematic roles as verb-specific concepts. *Language and Cognitive Processes*, *12*, 137–176.

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.

Mecklinger, A., Schriefers, H., Steinhauer, K., & Friederici, A. D. (1995). Processing relative clauses varying on syntactic and semantic dimensions: an analysis with event-related potentials. *Memory & Cognition*, *23*, 477–497.

Mitchell, D. C., Cuetos, F., Corley, M., & Brysbaert, M. (1995). Exposure-based models of human parsing: evidence for the use of coarse-grained (nonlexical) statistical records. *Journal of Psycholinguistic Research*, *24*, 469–488.

Parsons, T. (1990). *Events in the grammar of English.* Cambridge, MA: MIT Press.

Pickering, M. J., Clifton, C., & Crocker, M. W. (2000). Architectures and mechanisms in sentence comprehension. In M. W. Crocker, M. J. Pickering, & C. Clifton (Eds.), *Architectures and mechanisms for language processing* (pp. 1–28). Cambridge: Cambridge University Press.

Pickering, M. J., Traxler, M. J., & Crocker, M. W. (2000). Ambiguity resolution in sentence processing: Evidence against frequency-based accounts. *Journal of Memory and Language*, *43*, 447–475.

Pulvermüller, F., Härle, M., & Hummel, F. (2001). Walking or talking?: behavioural and neurophysiological correlates of action verb processing. *Brain and Language*, *78*, 143–168.

Quine, W. V. (1953). Notes on the theory of reference. In W. V. Quine (Ed.), *From a logical point of view* (pp. 130–138). New York: Harper and Row.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372–422.

Rayner, K., & Raney, G. E. (1996). Eye-movement control in reading and visual search: effects of word frequency. *Psychonomic Bulletin & Review*, *3*, 245–248.

Rayner, K., Rotello, C. M., Stewart, A. J., Keir, J., & Duffy, S. A. (2001). Integrating text and pictorial information: eye movements when looking at print advertisements. *Journal of Experimental Psychology: Applied*, *7*, 219–226.

Richards, D., & Goldfarb, J. (1986). The episodic memory model of conceptual development: An integrative viewpoint. *Cognitive Development*, *1*, 183–219.

Roy, D., & Pentland, A. (2002). Learning words from sights and sounds: A computational model. *Cognitive Science*, *26*, 113–146.

Runner, T. R., Sussman, R. S., & Tanenhaus, M. K. (2003). Assignment of reference to reflexives and pronouns in picture noun phrases: evidence from eye-movements. *Cognition*, *89*, B1–B13.

Scheepers, C., & Crocker, M. W. (2004). Constituent order priming from reading to listening: a visual-world study. In M. Carreiras & J. C. Clifton (Eds.), *The on-line study of sentence comprehension: Eyetracking, ERP, and beyond.* United Kingdom: Psychology Press.

Schriefers, H., Friederici, A. D., & Kühn, K. (1995). The processing of locally ambiguous relative clauses in German. *Journal of Memory and Language*, *34*, 227–246.

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, *71*, 109–148.

Spivey, M. J., & Gonzalez-Marquez, M. (2003). Rescuing generative linguistics: too little, too late? *Brain and Behavioural Sciences*, *26*, 690–691.

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 ambiguity resolution. *Cognitive Psychology*, *45*, 447–481.

Spivey-Knowlton, M. J. (1994). Quantitative predictions from a constraint-based theory of syntactic ambiguity resolution. In M. Mozer, J. Elman, P. Smolensky, D. Touretzky, & A. Weigand (Eds.), *The 1993 Connectionist Models Summer School* (pp. 130–137). Hillsdale, NJ: Erlbaum.

Spivey-Knowlton, M. J. (1996). *Integration of visual and linguistic information: human data and model simulations.* Unpublished doctoral dissertation, University of

Rochester, Rochester, N. J.

Steedman, M. (2002a). Formalizing affordance. In *Proceedings of the 24th Annual Conference of the Cognitive Science Society* (pp. 834–839). Mahwah, NJ: Lawrence Erlbaum.

Steedman, M. (2002b). Plans, affordances, and combinatory grammar. *Linguistics and Philosophy*, *5–6*, 723–753.

Stowe, L. A., Paans, A. M. J., Wijers, A. A., & Zwarts, F. (2004). Activations of "motor" and other non-language structures during sentence comprehension. *Brain and Language*, *89*, 290–299.

Sussman, S. R., & Sedivy, J. C. (2003). The time-course of processing syntactic dependencies: evidence from eye-movements. *Language and Cognitive Processes*, *18*, 143–163.

Tanenhaus, M. K., Carlson, G. N., & Trueswell, J. C. (1989). The role of thematic structures in interpretation and parsing. *Language and Cognitive Processes*, *4*, 211–234.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*, 632–634.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1996). Using eye movements to study spoken language comprehension: evidence for visually mediated incremental interpretation. In T. Inui & J. McClelland (Eds.), *Attention & performance XVI: integration in perception and communication* (pp. 457–478). Cambridge, MA: MIT Press.

Tanenhaus, M. K., Trueswell, J. C., & Hanna, J. E. (2000). Modeling thematic and discourse context effects with a multiple constraints approach: implications for the architecture of the language comprehension system. In M. W. Crocker, M. J. Pickering, & C. Clifton (Eds.), *Architectures and mechanism for language processing* (pp. 90–118). Cambridge: Cambridge University Press.

Taraban, R., & McClelland, J. (1988). Constituent attachment and thematic role assignment in sentence processing: influences of content-based expectations. *Journal of Memory and Language*, *27*, 597–632.

Townsend, D. J., & Bever, T. G. (2001). *Sentence comprehension: the integration of habits and rules*. Cambridge, MA: MIT Press.

Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the processing of unbounded

dependencies: an eye-tracking study. *Journal of Memory and Language*, *35*, 454–475.

Trueswell, J. C. (1996). The role of lexical frequency in syntactic ambiguity resolution. *Journal of Memory and Language*, *35*, 566–585.

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

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *19*, 528–553.

Tyler, L. K., & Marslen-Wilson, W. D. (1977). The on-line effects of semantic context on sytactic parsing. *Journal of Verbal Learning and Verbal Behavior*, *16*, 683–692.

Underwood, G., Jebbett, L., & Roberts, K. (2004). Inspecting pictures for information to verify a sentence: eye movements in general encoding and in focused search. *The Quarterly Journal of Experimental Psychology*, *56*, 165–182.

Webster, M. (1994). *Webster's new encyclopedic dictionary.* New York: Black Dog & Leventhal.

Zelinsky, G. J., & Murphy, G. L. (2000). Synchronizing visual and language processing: an effect of object name length on eye movements. *Psychological Science*, *11*, 125–131.

Curriculum Vitae

Name	Knoeferle
Vornamen	Pia Stefanie
Geburtstag	28. Februar 1976
Geburtsort	Ingolstadt
Adresse	Hafenstr. 6-8, 66111 Saarbrücken
	Tel.: 0681 302 6557
	E-mail: knoeferle@coli.uni-sb.de

1982–1986	Grundschule Hitzhofen
1986–1995	Katharinen-Gymnasium Ingolstadt
30/06/1995	Abitur, Katharinen Gymnasium Ingolstadt
1995–1997	Universität Regensburg: Lehramt Englisch, Französisch, Sport
1997–1998	University of Aberdeen: Englisch und Französisch
1998–1999	Universität Regensburg: Lehramt Englisch, Französisch, und Sport; Magister: Englische Philologie, Romanische Philologie, und Philsophie
1999–2000	Teaching Assistant in Landerneau (Bretagne, Frankreich)
2000–2001	Universität Regensburg: Lehramt Englisch, Französisch, und Sport; Magister: Englische Philologie, Romanische Philologie, und Philsophie
06/02/2001	Magister in Englischer Philologie, Romanischer Philologie, und Philsophie
19/07/2001	Erstes Staatsexamen in Englisch, Französisch, und Sport
2001–2005	Promotion an der Universität des Saarlandes im Fach Computerlinguistik