C E N T E R F O R R E S E A R C H I N L A N G U A G E

June 2007 **Vol. 19, No. 2**

CRL Technical Reports, University of California, San Diego, La Jolla CA 92093-0526
Tel: (858) 534-2536 • E-mail: editor@crl.ucsd.edu • WWW: http://crl.ucsd.edu/newsletter/current/TechReports/articles.html

TECHNICAL REPORT

The Coordinated Interplay Account of Utterance Comprehension, Attention, and the Use of Scene Information

Pia Knoeferle

Center for Research in Language, University of California, San Diego

EDITOR'S NOTE

The CRL Technical Report replaces the feature article previously published with every issue of the CRL Newsletter. The Newsletter is now limited to announcements and news concerning the **CENTER FOR RESEARCH IN LANGUAGE**. CRL is a research center at the University of California, San Diego that unites the efforts of fields such as Cognitive Science, Linguistics, Psychology, Computer Science, Sociology, and Philosophy, all who share an interest in language. The Newsletter can be found at
`http://crl.ucsd.edu/newsletter/current/TechReports/articles.html`.
The Technical Reports are also produced and published by CRL and feature papers related to language and cognition (distributed via the World Wide Web). We welcome response from friends and colleagues at UCSD as well as other institutions. Please visit our web site at `http://crl.ucsd.edu`.

SUBSCRIPTION INFORMATION

If you know of others who would be interested in receiving the Newsletter and the Technical Reports, you may add them to our email subscription list by sending an email to `majordomo@crl.ucsd.edu` with the line "`subscribe newsletter <email-address>`" in the body of the message (e.g., `subscribe newsletter jdoe@ucsd.edu`). Please forward correspondence to:

John Lewis and Jenny Staab, Editors
Center for Research in Language, 0526
9500 Gilman Drive, University of California, San Diego 92093-0526
Telephone: (858) 534-2536 • E-mail: editor@crl.ucsd.edu

Back issues of the the CRL Technical Reports are available on our website. Papers featured in recent issues include the following:

Syntactic Processing in High- and Low-skill Comprehenders Working under Normal and Stressful Conditions
Frederic Dick, Department of Cognitive Science, UCSD
Morton Ann Gernsbacher, Department of Psychology, University of Wisconsin
Rachel R. Robertson, Department of Psychology, Emory University
Vol. 14, No. 1, February 2002

Teasing Apart Actions and Objects: A Picture Naming Study
Analia L. Arevalo
Language & Communicative Disorders, SDSU & UCSD
Vol. 14, No. 2, May 2002

The Effects of Linguistic Mediation on the Identification of Environmental Sounds
Frederic Dick , Joseph Bussiere and Ayşe Pınar Saygın
Department of Cognitive Science and Center for Research in Language, UCSD
Vol. 14, No. 3, August 2002

On the Role of the Anterior Superior Temporal Lobe in Language Processing: Hints from Functional Neuroimaging Studies
Jenny Staab
Language & Communicative Disorders, SDSU & UCSD
Vol. 14, No. 4, December 2002

A Phonetic Study of Voiced, Voiceless, and Alternating Stops in Turkish
Stephen M. Wilson
Neuroscience Interdepartmental Program, UCLA
Vol. 15, No. 1, April 2003

New corpora, new tests, and new data for frequency-based corpus comparisons
Robert A. Liebscher
Cognitive Science, UCSD
Vol. 15, No.2; December 2003

The relationship between language and coverbal gesture in aphasia
Eva Schleicher
Psychology, University of Vienna & Cognitive Science, UCSD
Vol. 16, No. 1, January 2005

In search of Noun-Verb dissociations in aphasia across three processing tasks
Analía Arévalo, **Suzanne Moineau**
Language and Communicative Disorders, SDSU & UCSD, Center for Research in Language, UCSD
Ayşe Saygin
Cognitive Science & Center for Research in Language, UCSD
Carl Ludy
VA Medical Center Martinez
Elizabeth Bates
Cognitive Science & Center for Research in Language, UCSD
Vol. 17, No. 1, March 2005

Meaning in gestures: What event-related potentials reveal about processes underlying the comprehension of iconic gestures
Ying C. Wu
Cognitive Science Department, UCSD
Vol. 17, No. 2, August 2005

What age of acquisition effects reveal about the nature of phonological processing
Rachel I. Mayberry
Linguistics Department, UCSD
Pamela Witcher
School of Communication Sciences & Disorders, McGill University
Vol. 15, No.3, December 2005

Effects of Broca's aphasia and LIPC damage on the use of contextual information in sentence comprehension
Eileen R. Cardillo
CRL & Institute for Neural Computation, UCSD
Kim Plunkett
Experimental Psychology, University of Oxford
Jennifer Aydelott
Psychology, Birbeck College, University of London
Vol. 18, No. 1, June 2006

Avoid ambiguity! (If you can)
Victor S. Ferreira
Department of Psychology, UCSD
Vol. 18, No. 2, December 2006

Arab Sign Languages: A Lexical Comparison
Kinda Al-Fityani
Department of Communication, UCSD
Vol. 19, No. 1, March 2007

CRL Technical Reports, Vol. 19, No. 2, June 2007

The Coordinated Interplay Account of utterance comprehension, attention, and the use of scene information

Pia Knoeferle (pknoeferle@ucsd.edu)
Center for Research on Language,
University of California San Diego

Abstract

This paper reviews recent research on the interplay between language comprehension processes, attention to relevant objects and events, and the use of scene information for comprehension. It discusses, in particular, claims that information from scene events is prioritized over stereotypical thematic role knowledge during comprehension, and claims regarding a close temporal coordination between comprehension, visual attention, and the use of scene information. The discussion takes into account findings from different modalities (spoken comprehension versus reading), as well as insights regarding the decay of recently inspected scene information that is no longer present during language comprehension. We consider the implications of these findings for a recently proposed account of the coordinated interplay between utterance processing, attention, and the rapid use of scene information during comprehension.

Introduction

"What is the time course with which we integrate what we see in a scene with a sentence that we read or hear" is one question that we might ask when studying the use of scene information during language comprehension. Answering this question is of interest in various types of comprehension situations such as when we read comic books (Carroll, Young, & Guertin, 1992), newspaper advertisements (Rayner, Rotello, Stewart, Keir, & Duffy, 2001), or inspect scientific diagrams (Feeney, Hola, Liversedge, Findley, & Metcalf, 2000). In addition to reading, this question has also been addressed in spoken comprehension, when carrying out instructions to manipulate objects in the environment (Sedivy, Tanenhaus, Chambers, & Carlson, 1999; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995) or in listening to sentences that describe pictures on a computer screen (e.g., Knoeferle, Crocker, Scheepers, & Pickering, 2005). The above issue has by now been largely resolved, and it is widely accepted that scene information rapidly affects syntactic structure building (e.g., Knoeferle et al., 2005; Spivey, Tanenhaus, Eberhard, & Sedivy, 2002; Tanenhaus et al., 1995) and incremental semantic interpretation (e.g., Sedivy et al., 1999).

In addition to showing that people rapidly use a non-linguistic (e.g., scene) context for online language comprehension, existing findings also provide evidence for the rapid influence of linguistic knowledge and discourse contexts during online language comprehension. 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., Traxler & Pickering, 1996; Trueswell, Tanenhaus, & Kello, 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, Spivey-Knowlton, & Tanenhaus, 1998; Trueswell, Tanenhaus, & Garnsey, 1994). In addition, research on auditory comprehension has demonstrated its rapid influence on the incremental interpretation of an utterance that describes a scene (e.g., Kamide, Scheepers, & Altmann, 2003). It has further been shown that people rapidly use a referential discourse con-

text to disambiguate locally structurally ambiguous sentences (e.g., Altmann & Steedman, 1988).

Thus, linguistic context and knowledge as well as scene contexts are rapidly used for online language comprehension. Psycholinguistic findings that underscore the importance of these two types of information - linguistic and scene context - receive support from developmental studies that show both linguistic and non-linguistic (e.g., information about objects and events in the environment) information are important for language acquisition. With respect to the influence of diverse informational sources, the acquisition accounts by Gleitman (1990), and Gillette, Gleitman, Gleitman, and Lederer (1999) emphasize the importance of linguistic knowledge for child language acquisition. At the same time, they acknowledge that no one informational source can account for full language acquisition: "The position I have been urging is that children usually succeed in ferreting out the forms and the meaning of the language just because they can play off these two imperfect and insufficient data bases(the saliently interpretable events and the syntactically interpreted utterances) against each other to derive the best fit between them (Gleitman, 1990, p. 50). Snow (1977) suggested that early language is largely about objects and events in the immediate environment of a child, thus corroborating this proposal. For the acquisition of more abstract terms, in contrast, Gillette et al.(1999) suggested that information from the immediate scene was less important.

Developmental findings further draw our attention to pre-requisites for the successful use of scene information in acquisition: Experimental findings suggest, for instance, that the influence of nonlinguistic information (e.g., objects and events) on language acquisition is highly dependent on the time-lock between a child's attention to, and child-directed speech about, these objects and events (e.g., Dunham, Dunham, & Curwin, 1993; Harris, Jones, Brookes, & Grant, 1986; Tomasello & Farrar, 1986). It has been suggested that similar pre-requisities as for child language acquisition also hold for adult language comprehension (e.g., Knoeferle & Crocker, 2006). Findings from Tomasello & Farrar (1986) support the importance of attention for the acquisition of concrete words: when a child's attention was already focused on an object, words referring to it were learned better than words that were presented when trying to redirect the child's attention.

Based on these findings - the importance of both linguistic contexts / knowledge and scene contexts for language comprehension and acquisition, as well as based on first insights into pre-requisites for the use of visual contexts in language acquisition - recent psycholinguistic research has moved on to examine in more detail how precisely scene information influences adult language comprehension (Knoeferle, 2007; Knoeferle & Crocker, 2006, in press; Knoeferle, Habets, Crocker, & Münte, in press).

The present article will present an overview and a discussion of this recent research by Knoeferle and colleagues. A first section below will draw attention to five key issues of their research that are motivated by both psycholinguistic and developmental findings. The reminder of the article will review individual experimental studies that have investigated these five research issues and discuss relevant findings in the context of a first account of the interplay between language-mediated visual attention to objects and events, and the rapid use of scene information for language comprehension (Knoeferle & Crocker, in press).

Five questions on the use of scene events for online language comprehension

When considering the developmental findings that we reviewed above, it is noteworthy that the influence of information about objects and events on language acquisition depended on the time-lock between a child's visual attention to objects and child-directed speech. It is further interesting to note that while both linguistic and non-linguistic (e.g., scene) information influenced comprehension and acquisition, their relative importance is so far unclear. Motivated by these two insights, we highlight five questions on how information from a visual scene (e.g., events) influences adult language comprehension. The first two questions pertain to the time-lock between language-mediated attention to objects / events and the influence of information from these objects / events on comprehension. A further three points draw attention to the relative importance of linguistic knowledge versus scene context.

Q1 Is the time course with which scene context affects utterance comprehension a function of when scene information is identified as relevant by the utterance?

Q2 Is scene context on a par with linguistic context / knowledge in resolving local structural ambiguity?

Q3 Is scene information - when it is clearly relevant for comprehension - potentially prioritized for language comprehension compared with linguistic knowledge?

Q4 Does the absence of an immediate scene or scene events during language comprehension diminish the preferred reliance on depicted events?

Q5 Is a close time-lock between utterance-mediated attention to objects and events prerequisite for the rapid use of visual context (e.g., information about objects and evens) during comprehension?

In the following sections we review in more detail the studies that have addressed the above five issues, and discuss findings that provide insight into when there is (and when there is not) a close temporal coordination between utterance comprehension, attention in a scene and the use of relevant scene information.

Q1: Do we use scene context time-locked to language-mediated attention?

An important initial finding that relied on what has since been dubbed the "visual world" paradigm was relevant to both the coordination of utterance comprehension with scene processing and the rapid influence of scene information (Tanenhaus et al., 1995). Tanenhaus and colleagues presented people with instructions such as *Put the apple on the towel in the box*, in which the phrase *on the towel* can be temporarily analysed as modifier of the first noun phrase, identifying the location of the apple, or analysed as argument of the verb, indicating where to put the apple. As participants carried out the instructions, their eye gaze was monitored, and provided insights into online language comprehension. In a context containing one apple placed on a towel (location) and an empty towel (destination), people mostly inspected the apple when hearing *Put the apple*. Having heard *on the towel*, listeners mostly fixated the empty towel, suggesting interpretation of the ambiguous phrase as the destination. In a scene with two apples (of which one was on a towel) and an empty towel, fixation patterns differed from the start of the utterance in comparison with the one-apple context. People's eye movements - as they heard *the apple* - alternated between the two apples and settled on the apple on the towel shortly after hearing *on the towel*, indicating interpretation of the phrase *on the towel* as location of the apple. The core findings of this research are that utterance interpretation directs attention in the scene (see also Cooper, 1974; Spivey, Tyler, Eberhard, & Tanenhaus, 2001), and that the visual referential context rapidly influences the incremental structuring of the utterance. The rapidity of scene influence was confirmed by the fact that eye movements differed between the two visual context conditions from the onset of the utterance (see also Spivey et al., 2002).

The fixation patterns in the studies by Tanenhaus et al. (1995) do not, however, permit us to determine whether scene information influenced structuring and interpretation of the utterance in a manner that is closely time-locked to, or independent of, when the utterance identified that scene information as relevant. On a first interpretation, comprehension of the utterance (i.e., *the apple*) directed attention in the scene, and this triggered the construction and use of the appropriate referential context (one apple, two apples). A second possibility is that people acquired the referential context temporally independently of - maybe even prior to - hearing the utterance. They may then have accessed that context much as they would access a prior discourse context. The findings by Tanenhaus et al. are compatible with both interpretations. Since eye-movements differed from the start of the utterance between the two contexts, it is impossible to determine precisely whether or not identification of relevant scene information by the utterance was necessary for that scene information to affect structural disambiguation.

My own prior research that has relied on eye tracking in scenes during spoken comprehension has provided important insights into the temporal coordination of utterance-mediated attention and the use of scene information. Knoeferle et al. (2005) have extended insights on the rapid influence of visual contexts on comprehension by investigating comprehension in relatively rich scenes that contained depicted events in addition to objects. They examined the comprehension of subject-verb-object (SVO, see (1a)) and object-verb-subject (OVS, see (1b)) sentences that related

to event scenes. While both of these constituent orders are grammatical in German, the subject-initial order is preferred (e.g., Hemforth, 1993). The correct syntactic and thematic relations were ambiguous prior to the sentence-final accusative (SVO, 1a) or nominative (OVS, 1b) case-marked noun phrase.

Using such initially structurally ambiguous sentences (1) while people inspected a related event scene (Fig. 1) permitted the authors to examine whether people would use depicted events that show who-does-what-to-whom for early disambiguation (i.e., prior to hearing the second noun phrase that disambiguated the structural ambiguity through case marking on the determiner of that noun phrase).

The materials were constructed such that early disambiguation through depicted events was possible at the verb, if listeners exploited scene information (a princess washing a pirate; a fencer painting that princess, see Fig. 1). The verb for canonical ambiguous-verb-object sentences (1a) identified the princess as the agent of a washing event (princess-washing-pirate). In contrast, for non-canonical ambiguous-verb-subject sentences, the verb (*paints*) identified the princess as the patient of a painting event performed by another agent, the fencer. Such depicted role relations make it clear whether the referent of the initially ambiguous noun phrase (the princess) is the agent (SVO) or patient (OVS), and hence the subject or object of the sentence. If people use verb-mediated depicted events, then we would expect to see their use of scene events reflected in their eye-movements to relevant role fillers in the scene. Specifically, after hearing the verb *washes* and noticing that washing is the relevant event (1a, SVO), they may infer that the pirate is the patient of the event and look at him immediately after hearing the verb and before he is mentioned by the second noun phrase. After hearing *paints*, in contrast, the role relations make it clear that the princess is undergoing the event, and hence the object of the sentence, an inference that should result in people rapidly anticipating the correct agent (the fencer) if people rapidly exploit depicted events.

(1) (a) Die Prinzessin wäscht den Pirat.
'The princess (amb.) washes the pirate (ACC).'
(b) Die Prinzessin malt der Fechter.
'The princess (amb.) paints apparently the fencer (NOM)'.

Figure 1: Example image for sentences 1a and 1b

Indeed Knoeferle et al. found more eye movements to the patient (the pirate) than the agent (the fencer) of the action for SVO sentences (1) and more inspections of the agent (the fencer) than patient (the pirate) for OVS sentences (1) respectively. This gaze pattern occurred shortly after people had heard the verb and prior to people hearing the disambiguating second noun phrase. The time course of the gaze pattern permitted the authors to interpret this finding as revealing rapid thematic role assignment and structural disambiguation through verb-mediated depicted events. We replicated these findings in another language (English) and for another sentence structure (main clause/reduced relative clause ambiguity) (Knoeferle & Crocker, 2006).

The findings by Knoeferle et al. (2005) thus showed a rapid effect of depicted events on thematic role assignment as evidenced by eye movements to role fillers in the scene. These findings further suggest a close temporal coordination between when a depicted event is identified as relevant by the utterance (at the verb), and the point in time when that event affects comprehension. This view was supported by the observation that gaze patterns revealed thematic role assignment only after the verb had mediated a depicted event (washing/painting) but with a close temporal coordination to the verb (i.e., post-verbally).

Knoeferle (2007) tested this hypothesis. The authors examined thematic role assignment and the structuring of an utterance for two types of German sentences that differ with respect to when the utterance identifies relevant role relations in the scene. If the above "temporal coordina-

tion" account is correct, then the time-course with which a depicted event can affect structural disambiguation and thematic role assignment depends on when that depicted event is identified as relevant for comprehension by the utterance. For earlier identification of a relevant depicted event, an earlier influence of that event on structure-building and thematic role assignment would be expected than for cases where identification of a relevant event takes place comparatively later. Gaze pattern in the scene revealed indeed that the point in time when the utterance identified a relevant depicted event was closely temporally coordinated with the point in time when that depicted event triggered thematic role assignment and structuring of the utterance. An early mediation of the relevant depicted event (before people heard the verb) triggered an earlier thematic role assignment than a comparatively later, verb-based mediation of the relevant depicted event in initially ambiguous sentences (Knoeferle, 2007). The take-home message from this set of studies is that (a) depicted events affect structural disambiguation rapidly, and (b) that scene events are used for online comprehension in close temporal coordination with utterance-mediated attention to the events.

Q2: Are scene events on a par with linguistic knowledge for disambiguation?

The above section has reviewed findings that concerned the first question (Q1). In what follows, we will consider findings that pertain to the first two questions (Q1 and Q2). The findings by Knoeferle et al. (2005) lend support to the view of a temporally coordinated interplay between utterance-mediated attention and the use of information about scene events. Their findings and those by Tanenhaus et al. (1995) furthermore provide behavioral evidence for the claim that scene information affects the incremental structuring of initially structurally ambiguous utterances. Utterance-mediated attention in scenes is also known to reflect various other underlying linguistic and non-linguistic processes such as semantic interpretation (Sedivy et al., 1999) thematic interpretation (e.g., Altmann and Kamide 1999), or visual search (e.g., Spivey et al., 2001). Furthermore, eye movement measures alone do not clarify whether the processes involved in resolving local structural ambiguity through scene information are similar to the processes that are at work when linguistic cues resolve a temporary structural ambiguity.

To better understand the influence of visual contexts (depicted events) on structural revision during spoken sentence comprehension, Knoeferle et al. (in press) conducted event-related potential (ERP) studies. Measures such as ERPs have in the past been used to examine the processing of syntactic violations (e.g., Friederici, Pfeifer, and Hahne 1993; Hagoort, Brown, and Groothusen 1993; Osterhout, Holcomb, and Swinney 1994) and, in particular, the resolution of temporary structural ambiguity through linguistic cues: When linguistic cues triggered structural revision towards a non-canonical structure during reading in the absence of scenes, the difficulty of this revision has typically been associated with a positivity that has a maximum at approximately 600 ms (P600, e.g., beim Graben, Saddy, and Schlesewsky, 2000; Frisch et al., 2002; Matzke et al., 2002).

Based on these findings (beim Graben, Saddy, and Schlesewsky, 2000; Frisch et al., 2002; Matzke et al., 2002) Knoeferle et al. (in press) investigated the structural revision of locally structurally ambiguous German utterances through linguistic cues (e.g., case marking on the determiner of a noun phrase) and through verb-mediated depicted events in sentences such as (1a/b). In the ERP study, sentences such as (1a/b) were presented with the image in Fig. 1 just as in the eye-tracking study. In addition, the sentences from the eye-tracking study (e.g,. 1a/b) were presented in the absence of scenes, thus forcing late disambiguation through linguistic cues (case marking on the determiner of the second noun phrase disambiguated (1a) towards SVO and (1b) towards OVS structure). This permitted the authors to examine two issues: A first important question that the authors addressed in the ERP study was whether brain potentials - just as eye-movements - would reveal the use of depicted events for structural disambiguation shortly after the verb. Evidence in favor of this view would be if we were to find a positivity for non-canonical relative to canonical conditions time-locked to the onset of the verb in cases when the verb mediates relevant depicted events.

In addition, we wanted to compare disambiguation through depicted events - mediated at the verb - with linguistic disambiguation through the

case marked determiner of the second, post-verbal noun phrase (i.e., when no scenes were present). When the verb does not mediate relevant events (e.g., when no scenes are present during utterance presentation), no early disambiguation through verb-mediated events should be possible. As a result, when no scenes are present, there should be no P600-like component for non-canonical relative to canonical conditions time-locked to the verb. Rather, we should observe late disambiguation on the post-verbal second noun phrase.

Findings confirmed that depicted events are immediately used for structural disambiguation: When relevant depicted events were available for inspection during utterance presentation, findings showed that scenes containing depicted events rapidly influence the disambiguation of a locally structurally ambiguous utterance. This was revealed by an earlier P600, time-locked to the verb that mediated the events in auditory event-related potentials when depicted events were simultaneously present while people listened to the utterance. When no scenes were present, in contrast, and disambiguation of local structural ambiguity towards either an object-verb-subject or a subject-verb-object order could only occur through a case-marked determiner on the second noun phrase, Knoeferle and colleagues found - as expected - no P600 time-locked to the verb. Rather, they replicated previous findings of later disambiguation through cues in the linguistic input (e.g., Matzke et al., 2002): They observed a positivity with a peak at approximately 600 ms time-locked to the onset of the second noun phrase in response to linguistic disambiguation towards the non-canonical structure.

With respect to the role of scene information in disambiguation it is further interesting to note that the distribution of the P600-like component was similar - regardless of whether disambiguation was triggered by depicted events at the verb or by linguistic marking on the second noun phrase. This suggests that the kinds of cues (e.g., non-linguistic depicted events versus linguistic cues such as case marking) that trigger disambiguation do not fundamentally modulate the neural correlates underlying disambiguation mechanisms. It appears that scene context (e.g., depicted events) is used on a par with linguistic cues (e.g., case marking) for structural disambiguation.

Q3: Are scene events preferred in situated comprehension over linguistic knowledge?

In fact, are scene events maybe even prioritized over linguistic context and knowledge in situated comprehension? A further finding that provided evidence for both the great importance of depicted events in comprehension and for the close temporal coordination between utterance comprehension and the use of scene information comes from Knoeferle and Crocker (2006). The authors directly compared the importance of verb-mediated world knowledge about likely role-fillers (Kamide et al., 2003) with that of verb-mediated depicted events (Knoeferle et al., 2005).

To directly compare these two sources of information they relied upon characters in a clipart scene that could be identified through a verb in the utterance - relying either on people's world knowledge or alternatively on events that the characters were depicted as performing. An example image in the eye tracking study showed two agents (e.g., a doctor, and a cook), each performing an action upon a patient (e.g., the tourist, see Fig. 2). The depicted events provided information about role relations (e.g., doctor-jinxing-tourist and cook-bandaging-tourist). In addition, each agent provided stereotypical thematic role knowledge (e.g., a cook is a stereotypical agent of a cooking action, and a doctor a stereotypical agent of a bandaging action). Sentences always started with an unambiguously accusative case-marked noun phrase referring to a patient role-filler (Fig. 2, the tourist).

(2) (a) Den Touristen (ACC) verköstigt gleich der Koch (NOM)
'The tourist (ACC) serves-food-to soon the cook (NOM).'

(b) Den Touristen (ACC) verzaubert gleich der Arzt (NOM)
'The tourist (ACC) jinxes soon the doctor (NOM).'

(c) Den Touristen (ACC) bandagiert gleich der Arzt (NOM)
'The tourist (ACC) bandages soon the doctor (NOM). '

(d) Den Touristen (ACC) bandagiert gleich der Koch (NOM)
'The tourist (ACC) bandages soon the cook (NOM). '

A first condition pair was designed to ensure that depicted events and verb-based stereotypical

Figure 2: Example image for sentences 2a to 2d

role knowledge each rapidly influence thematic role assignment when uniquely identified. For the image in Fig. 2 and sentence 2a, accusative case-marking on the determiner of the first noun phrase marked the first noun phrase as the patient. This together with the verb, *verköstigt* ('serves-food-to'), should bias expectations towards an upcoming agent. World knowledge associated with the verb, in particular, identified the cook as the likely agent of a cooking-event. The fact that there were no other agents that performed a cooking event or that were plausible agents for such an event made the cook a unique target based on world knowledge. In contrast, in a second condition, another character - the physician - was identified as a unique target by means of the action depiction (rather than world knowledge): When people heard sentence 2b, the verb uniquely identified a (nontypical) agent of the verb as relevant target (the physician) through a jinxing action that the physician was depicted as performing. No other character was either a plausible agent for a jinxing action or depicted as performing such an action.

A second condition pair contrasted stereotypical knowledge and depicted events: For Fig. 2 and sentences (2c) and (2d), the verb ('bandages') identified two agents as likely: a stereotypical agent (the physician), and a second participant depicted as the agent of a bandaging event (the cook) (Fig. 2). Prior to hearing the second noun phrase, the comprehension system is forced to chose between relying on the immediate event depiction or on stereotypical thematic role knowledge for anticipating an agent in the scene since these two types of information suggest each a different entity as likely agent.

Gaze pattern for conditions (2a) and (2b) in Knoeferle and Crocker (2006) showed that people relied on each of these two informational sources when they uniquely identified a relevant scene agent: For condition (2a), this claim was confirmed by more inspections to the stereotypical agent (the cook) than to the other agent in the scene. For sentence (2b), in contrast, people looked more often at the agent depicted as performing a jinxing action (the physician) than at the second agent in the scene (the cook).

In contrast, for conditions (2c) and (2d), we expected no such interaction but rather a main effect of inspections to either the agent that was depicted as performing a non-stereotypical action (cook-bandaging) or to the agent that was compatible with expectations based on stereotypical knowledge (physician-bandaging). The interesting question was which agent people would inspect most for conditions (2c) and (2d) shortly after the verb. The authors observed a higher proportion of anticipatory eye movements (prior to the second noun phrase) to the agent depicted as performing the verb action (the cook) in comparison with the agent that was stereotypical for the verb (the doctor). This suggests a strong preference of the comprehension system to rapidly rely on depicted events over stored thematic knowledge for incremental thematic role assignment. Note that the finding of a greater relative priority of depicted events over stereotypical event knowledge does not depend on specifics of the images and how the event depiction was realized. The image was the same for all four conditions and since findings for (2a) and (2b) clearly showed that people can use each type of information - depicted action and stereotypical event knowledge - equally well provided it is uniquely identified. It is only when the utterance is ambiguous regarding which character - the agent depicted as performing a non-stereotypical action or the stereotypical agent - is the most likely agent of the verb that we observed the preference to rely on the depicted events and the associated agent over knowledge of stereotypical events.

Q 4 and 5: Do *recent* objects & events rapidly influence spoken comprehension?

Together, the findings that I reviewed above provide strong evidence for a tight temporal coordination between utterance-mediated attention and the use of scene events for comprehension when those events are simultaneously present during comprehension. In addition, they underscore the great importance of scene events for utterance comprehension when scenes are both co-present and relevant to comprehension.

A crucial point in both the coordinated interplay and the rapid use of scene information, however, is the co-presence of an immediate scene: The close time-lock of utterance and attention in a scene has importantly been extended to serial picture-utterance presentation at least for scenes that contain objects. In Altmann (2004) people inspected an image with a man, a woman, a cake, and a newspaper, and then the screen went blank (see also Richardson & Spivey, 2000; Spivey, Richardson, & Fitneva, 2004). Two seconds later, people heard a sentence that described part of the previously-inspected scene (e.g., *The man will eat the cake*). Once people had heard the verb in the sentence, they rapidly looked at the location on the blank screen where previously there had been a cake. The time-course of eye-movements in the serial-presentation study closely ressembled the time-course of gaze-patterns in an earlier study (Altmann & Kamide, 1999) with concurrent scene and utterance presentation. The data by Altmann (2004) suggest that even when a prior scene is no longer immediately available, mental representations of the spatial configuration of objects in the scene may remain accessible for comprehension.

It was unclear whether findings by Altmann (2004) extend to complex scenes that contain agent-action-patient events in addition to objects and characters, and to scene information that is non-stereotypical, or even implausible, as with the depicted events in the studies by Knoeferle and Crocker (2006). Furthermore, even if people rapidly *exploit* depicted event information when those events are both non-stereotypical and no longer visually present during utterance presentation, it is unclear whether they still *prioritize* them over long-term knowledge of stereotypical agents for thematic interpretation.

Knoeferle and Crocker (in press) examined these issues by modifying scene presentation. In real life, for instance, scene events may be at a greater disadvantage than entities regarding their co-presence. Real-world actions are often fleeting, completed, and part of the recent past, and may as a result only be briefly available for inspection, and not necessarily concurrently with utterance comprehension. Event participants, however, often remain for some time after performing an action. To put the depicted event preference to a test, they used the stimuli from Knoeferle and Crocker (2006, Exp. 2), but presented them in sequences of four images to briefly depict the actions, then remove them with only the characters remaining as people listened to the utterance. This presentation was a first, coarse-grained approximation of more dynamic event sequences. It further disadvantaged the depicted events alone since the characters remained visible during comprehension and were thus in the scene for performing future actions.

Gaze patterns confirmed that - even when the depicted actions were absent while the characters remained in the scene - people were able to use either verb-based stereotypical thematic knowledge or non-stereotypical depicted events when the verb uniquely identified these information types as relevant. However, when the verb identified two different agents as relevant, there no longer was a preferential anticipation of the agent that had previously been depicted as performing an action. There was a tendency to anticipate the stereotypical agent more often than the agent of the (non-stereotypical) depicted event, but this effect was not significant. Regarding the temporal coordination, the time course of gaze patterns was similar to the original study in which the events had been simultaneously during utterance comprehension.

The analyses of gaze patterns from this study suggest that the visual co-presence of an event is not required to explain the findings by Knoeferle and Crocker (2006). If visual co-presence were crucial, we should not have replicated the finding that people still rely on depicted events with dynamic action presentation and when only the characters were simultaneously present during utterance presentation. On the other hand, if modulating the co-presence of events did not affect the *priority* of depicted events for comprehension, then the preference to rely on depicted events should have been replicated with dynamic action presentation.

Summary of the experimental findings

The findings from eye tracking gaze on a display during comprehension of a concurrently presented utterance have shown that depicted events rapidly influence structural disambiguation; that there is a close temporal coordination between the identification of relevant scene information through the utterance, attention to relevant scene information, and the use of associated scene events for structural disambiguation. Our data further provide evidence for the view that scene events are on a par with linguistic cues for structural disambiguation, and that scene events are even prioritized for comprehension over stereotypical thematic role knowledge.

The close temporal coordination was not interrupted by the absence of recently inspected scene events during utterance comprehension. The preferred reliance on scene events, in contrast, was diminished when scene events were not simultaneously present during utterance presentation while stereotypical thematic role fillers remained in the scene during utterance presentation.

This suggests that the close temporal coordination plays a great role in the use of scene information during comprehension. For the greater relative importance of scene events our findings suggest that their influence relative to that of other information sources such as stereotypical thematic role knowledge decays when non-stereotypical scene events are no longer present during utterance presentation while characters that proffer stereotypical knowledge are present in the scene. We now discuss the findings from the preceding sections regarding their theoretical implications for theories of language comprehension.

Implications for frameworks of language comprehension

Findings from the above studies have provided evidence for a fluid interaction between linguistic processes and the output of the visual perceptual system. To account for these findings, Knoeferle and colleagues have considered existing frameworks of the language system and language processing.

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 instance, 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 3 (see Coltheart, 1999, p.116). Transducers convert physical stimuli into neural activity that is then passed on to the input modules. These interpret the transduced signal. Examples for input modules are the language or visual perception systems, and their characterizing property 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 informational encapsuation. This term means that distinct input modules such as language and visual perception have only access to the output of another module, but cannot influence its internal processes. In Figure 3 this is illustrated by the lack of 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. As a result of this assumption, direct consideration of the influence of scene perception on such core comprehension processes is not possible (Knoeferle, 2005).

The findings that we have presented in this paper, as well as existing results in the literature (e.g., Tanenhaus et al., 1995) exclude strictly modular accounts that postulate informational encapsulation of processes internal to the language system (e.g., Frazier & Clifton, 1996). Nonetheless, it appears that under the influence of a Fodorian view of the mind, many psycholinguistic theories of online language comprehension have been developed on the basis of the assumption that the mechanisms underlying language comprehension can be examined in isolation from the perception of scene information. This becomes apparent in the fact that scene information has not been explicitly included in most psycholinguistic frameworks of online language comprehension (e.g., Forster, 1979; Frazier & Fodor, 1979; Frazier & Clifton, 1996;

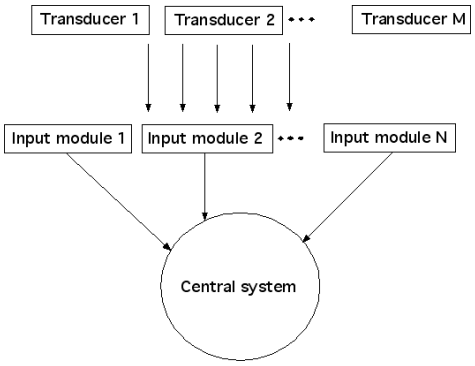

Figure 3: Schematic outline of a Fodorian organization of the mind, Coltheart, 1999, p. 116

MacDonald, Pearlmutter, & Seidenberg, 1994; Pickering, Traxler, & Crocker, 2000; Townsend & Bever, 2001). Furthermore, the visual perceptual system has not been explicitly included as a cognitive system that might proffer important information for comprehension. We argue that as a result such theories do not make sufficiently detailed predictions for language comprehension in situations in which language is about the scene (i.e., detailing the time course, temporal coordination, and relative importance of scene information).

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., Altmann & Steedman, 1988; Tanenhaus et al., 1995). Interactionist theories and the Referential Theory of sentence processing (Crain & Steedman, 1985; Altmann & Steedman, 1988) account for the findings by Tanenhaus et al. (1995) in settings that contain objects. They do not, however, provide the rich inventory of ontological categories, referential expressions and mental representations required to account for the influence of depicted events.

To account for the rapid, verb-mediated influence of depicted events, Knoeferle et al. (2005) have adopted the Jackendoff architecture (Jackendoff, 1997, 2002) as a suitable basis for describing the processing mechanisms that underlie on-line comprehension (see Fig. 4). Jackendoff (2002, p. 198) proposes a parallel constraint-based processing architecture and includes a suitably rich inventory of representations. Interface processors allow information exchange between semantic/conceptual structure and perception or action, permitting us to describe a rapid, incremental interplay between scene information and sentence comprehension. The individual levels in Jackendoffs architecture are modular in the sense of being domain-specific (i.e., their representational vocabulary is specialized), but unlike Fodorian modularity, linguistic structures are linked among themselves and to other cognitive sub-systems (see Fig. 4). 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. Jackendoff s framework provides an interface at which conceptual structure and the visual system can be linked, and thus provides a means for describing how comprehension proceeds when these different types of information are available and interact (see Knoeferle, 2005).

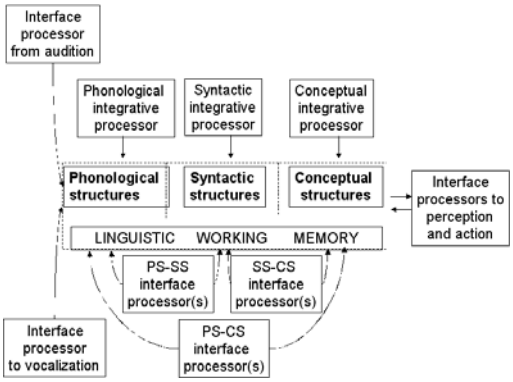

Figure 4: Schematic outline of the Jackendoff architecture (Jackendoff, 2002, p. 1999)

While Jackendoff outlines an architecture for situated comprehension, he does not provide a processing account of the coordinated interplay between utterance comprehension, attention, and the use of scene information. Based on the above

findings, and further motivated by insights concerning the importance of both joint attention and scene information in language acquisition (e.g., Harris et al., 1986; Dunham et al., 1993; Snow, 1977), Knoeferle and Crocker (2006) and Knoeferle and Crocker (in press) outlined the Coordinated Interplay Account of scene-sentence interaction in adult utterance comprehension. Briefly recall the findings that the authors had to account for:

F1 Rapid effect of scene objects / events on incremental structure building

F2 Close temporal coordination between the point in time when scene information is identified as relevant by the utterance and when it affects online comprehension.

F3 Scene context is on a par with linguistic cues for structural disambiguation and may even be prioritized for online comprehension

F4 The absence of recently inspected events during language comprehension does not disrupt the close temporal coordination between utterance processing and the influence of those scene events on comprehension

F5 The preferred reliance on scene events is a function of their decay in working memory

We first outline at a general level how the Coordinated Interplay Account by Knoeferle and Crocker operates. Subsequent to that we will provide two illustrative examples. The Coordinated Interplay Account identifies two basic steps in situated comprehension. Note that we assume these processes overlap modulo informational dependencies. First, comprehension of the unfolding utterance guides attention in the scene. This permits establishing reference to objects (e.g., Tanenhaus et al., 1995) and events (e.g., Knoeferle et al., 2005) and anticipating likely referents (e.g., Altmann & Kamide, 1999). The currently processed word is integrated into the structure and interpretation of the utterance, and the listener derives linguistic expectations in the process. The listener then inspects the scene for objects/events that the current input refers to, and he may further explore the scene, anticipating objects and / or events based on her linguistic and world-knowledge expectations. Upon finding an object / event that is the referent for the current input, reference to that object/event is established by forging a link between the meaning of the referring expression in the input and the designated object/event in the scene/world. In the model, this is achieved through coindexing of referring expressions in the utterance (e.g., nouns and verbs, see Knoeferle et al., 2005) with corresponding scene objects and events (see also Jackendoff, 2002).

In the second step, once attention has shifted to the most relevant object or event, the scene information that has been inspected rapidly influences utterance comprehension (Knoeferle & Crocker, 2006; Knoeferle, 2007). After any revisions, the next word is integrated with the revised interpretation, yielding a new interpretation and derived linguistic expectations. The Coordinated Interplay Account crucially suggests that the close timelock between comprehension and attention in the scene is at the origin of the relative priority of immediately depicted events over knowledge of stereotypical events in comprehension.

Findings F4 and F5 furthermore point to an account in which prior scene information can inform subsequent comprehension, but also in which objects and events that are no longer present during comprehension experience some decay. These findings (F4/5) have motivated a revision of the original Coordinated Interplay Account as outlined above. This revised version differs from the account described above in that it incorporates an explicit working memory from which recently inspected scene information can be accessed for rapid use during online language comprehensio. Knoeferle and Crocker (in press) furthermore revised the original mechanism such that representations only available in working memory can still influence subsequent comprehension but are less salient than representations that still receive support from corresponding objects/events that are present during utterance presentation. The revised mechanism thus assumes that entities and events in working memory, which are no longer in the scene, decay. This mechanism permits the authors to account for the diminished priority of recent scene events while it ensures that recent scene events are available during online comprehension.

I now present an example for how structural disambiguation through depicted events takes place according to the account. Consider the example sentence 1a, *Die Prinzessin wäscht den Pirat*. When people process the first noun phrase,

they look for an appropriate referent and establish reference to the relevant entity in the corresponding scene, the princess (Fig. 1). People during that time mostly inspect the princess and probably notice proximal scene information such as the action that she performs. Based on both linguistic preferences of subject-initial constituent order and the scene depiction, people likely interpret the noun phrase 'the princess' as the subject and agent of the sentence. For SVO sentences, people then hear the verb 'washes'. They search for a matching action referent. Their expectations of the princess being the agent is reinforced when the verb matches the action that the princess performs. As a result people develop expectations about a likely patient of the washing action and anticipate the pirate shortly after the verb more often than the other role-filler (the fencer). For OVS sentences, in contrast, people hear the verb 'paints'. When they try to find a referent, they notice that the action performed by the princess is incongruent with the verb 'paints'. The model assumes people then engage in a search for an appropriate referent, and notice the painting action. They establish reference between the verb 'paints' and the appropriate painting action. Upon inspecting that action they may notice proximal scene information such as the agent (fencer) associated with the action, and its patient, the princess. The event information can then be used to revise the initial thematic interpretation and syntactic structure of the utterance.

As a further example, consider an outline of how the model accounts for the greater relative priority of depicted events. Consider the example sentence *Den Touristen bandagiert gleich der Doktor* ('The tourist$_{acc/obj}$ bandages soon the physician$_{nom/subj}$'). Recall that the corresponding scene showed a tourist, a cook bandaging the tourist, and a physician jinxing the tourist (see Fig. 2). When people encounter the first noun phrase *Den Touristen* ('the tourist'), the meaning of the noun *Tourist* is accessed and integrated with linguistic constraints (e.g., accusative case marking on the determiner of the noun phrase). People begin to build an interpretation with the tourist as the patient of the action. Most eye movements are directed towards the tourist during referential processing, and the comprehension system establishes reference to the tourist in the scene. People may also explore scene regions that are close by the tourist, and thus notice that the tourist is the patient of two events: a wizard that is spying on the tourist, and a detective that is serving food to the tourist.

Then people encounter the verb *bandagiert* ('bandages'). Its meaning is accessed, and the interpretation that is built consists of a bandaging action of which the tourist is the patient. At this point, verb meaning creates linguistic expectations of a physician as a stereotypical agent of 'bandages'. The scene events, in contrast, suggest that the cook is depicted as the agent of a bandaging action provided people have perceived this event. Based on verb meaning, people attempt to establish reference between the verb an an appropriate action, prompting them to look at the depicted bandaging action. In the process they also look at proximal objects such as the agent of that event (the cook). Informed by the scene event, the cook is now taken to be the most relevant agent for the bandaging action. For more detailed examples also on how the account deals with the influence of recent scene events see Knoeferle et al. (2005).

Conclusions

To summarize, depicted events influenced spoken comprehension rapidly both when a relevant scene was present and when it was absent. The greater relative priority of scene events over stereotypical thematic role knowledge, however, was diminished when the events were no longer copresent while characters and their stereotypical affordances were supported by the immediate visual context.

While accounting for the above findings through the coordinated interplay mechanism and decay in working memory, there are obviously other factors that have not yet been explicitly included in the Coordinated Interplay Account and that may influence the use and relative importance of scene objects and events. One example is the extent of referential success (see Knoeferle and Crocker (2006, in press)); Another issue is locational or temporal cues in the utterance that clarify the (ir)relevance of the immediate scene. Finally, the availability of social interaction (e.g., gestures) for explicitly directing attention to scene information will likely modulate the extent to which people prefer to rely on scene events.

Imagine a situation in which you sit in front of the television with a friend, and your friend says "You know, I really loved those cookies at aunt

Marge's yesterday". At the same time, the TV screen displays a newscaster. In this situation, none of the referring expressions are present on the screen. Furthermore, the utterance refers to an event in the past. Both of these constraints will obviously decrease reliance on information about objects and events from the immediate screen. In the absence of cues to bias against a strong relative importance of scene information, however, the Coordinated Interplay Account predicts that referents and associated proximal scene objects/events that one pays attention to will play a dominant role in guiding situated comprehension processes (see Knoeferle & Crocker, in press).

Acknowledgements

This paper was written while the author was on a postdoctoral fellowship awarded by the German Research Foundation (DFG). Thanks go to one anonymous reviewer for her / his comments. We also would like to thank Roger Levy for helpful comments on the manuscript.

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.

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.

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

Cooper, R. (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.

Dunham, P. J., Dunham, F., & Curwin, A. (1993). Joint-attentional states and lexical acquisition at 18 months. *Developmental Psychology*, *29*, 827–831.

Feeney, A., Hola, A. K. W., Liversedge, S. P., Findley, J. M., & Metcalf, R. (2000). How people extract information from graphs: evidence from a sentence-graph verification paradigm. In M. Anderson, P. Cheng, & V. Haarslev (Eds.), *Diagrams* (pp. 149–161). Berlin: Springer Verlag.

Fodor, J. (1983). *Modularity of mind*. Cambridge, MA: MIT 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., & 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.

Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*, *73*, 135–176.

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, *1*, 3–55.

Harris, M., Jones, D., Brookes, S., & Grant, J. (1986). Relations between the non-verbal context of maternal speech and rate of language development. *British Journal of Developmental Psychology*, *4*, 261–268.

Hemforth, B. (1993). *Kognitives Parsing: Repräsentation und Verarbeitung sprachlichen Wissens*. Sankt Augustin: Infix-Verlag.

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.

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.

Knoeferle, P. (2005). *The role of visual scenes in spoken language comprehension: Evidence from eye-tracking*. http://scidok.sulb.uni-saarland.de/volltexte/2005/438: Saarland University. (Doctoral Dissertation in Computational Linguistics)

Knoeferle, P. (2007). Comparing the time-course of processing initially ambiguous and unambiguous german SVO/OVS sentences in depicted events. In R. van Gompel, M. Fischer, W. Murray, & R. Hill (Eds.), *Eye movement research: insights into mind and brain.* Oxford: Elsevier.

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

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

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.

Knoeferle, P., Habets, B., Crocker, M. W., & Münte, T. F. (in press). Visual scenes trigger immediate syntactic reanalysis: evidence from ERPs during situated spoken comprehension. *Cerebral Cortex*.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*, 676–703.

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.

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.

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.

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, *71*, 109–148.

Snow, C. (1977). Mothers' speech research: from input to interaction. In C. Snow & C. A. Ferguson (Eds.), *Talking to children: language input and acquisition.* Cambridge, MA: Cambridge University Press.

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, M. J., Tyler, M. J., Eberhard, K. M., & Tanenhaus, M. K. (2001). Linguistically mediated visual search. *Psychological Science*, *12*, 282–286.

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.

Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child Development*, *57*, 1454–1463.

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., 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. *Jour-*

nal of Experimental Psychology: Learning, Memory and Cognition, *19*, 528–553.