REAL-TIME THEMATIC ROLE ASSIGNMENT IN CHILDREN AND ADULTS

The Influence of Case-Marking, Prosody, and Visual Cues

JULIA MARINA KRÖGER

A thesis submitted in partial fulfilment of the requirements for the degree of
Doktor der Philosophie

Faculty of Linguistics and Literature

Universität Bielefeld

2018

Printed on non-ageing paper °°ISO 9706

EVALUATION COMITTEE

PROF. DR. PIA KNOEFERLE
PROF. DR. MARTINA HIELSCHER-FASTABEND
PROF. DR. PETRA WAGNER
DR. JOANA CHOLIN
DR. ANNETT JORSCHICK

CONTENTS

ABSTRACT

Children rapidly acquire their first language in the early stages of their life. Not surprisingly, not all aspects of language are present from the beginning. For that reason, psycholinguistic research has focused on how children process language in comparison to adults. However, unlike research with adult participants which has identified a bi-directional relationship between spoken utterances and visual attention, research on children was often conducted offline (i.e., using offline measures only). With this thesis, we tried to extend existing findings on real-time child language comprehension which suggest pronounced adult-children differences, especially in the pragmatic domain.

In six eye-tracking studies (grouped into 3 Experiments), we investigated the influence of case-marking, prosody, and non-linguistic visual cues (depicted actions/a wiggling target character) on real-time thematic role assignment in children and directly compared the children's behaviour to that of adults (Exp 1, Exp 3, but Exp 2 adults only). In addition to monitoring eye-movements, we recorded accuracies for active and passive voice post-sentence comprehension questions. Participants listened to unambiguously case-marked subject-verb-object (SVO) and object-verb-subject (OVS) sentences (Exp 1, Exp 2a), unambiguously and ambiguously case-marked OVS sentences (Exp 2b), or ambiguously case-marked OVS sentences (Exp 3). Sentences were assigned an SVO- or OVS-biasing prosodic contour (SVO: L*+H accent on the NP1 and H* accent on the verb; OVS: L+H* accent on the NP1; as proposed by Weber, Grice, & Crocker, 2006). As a baseline, we added a neutral prosodic contour (Exp 1), or directly contrasted the SVO- with the OVS-biasing prosodic contour (Exp 2). In Exp 3, prosody was kept constant across conditions (OVS-biasing prosodic contour). Visual scenes contained three clipart animal characters, two of which were depicted as performing identical actions (Exp 1, Exp 2). Thus, the depicted actions did not disambiguate role relations early in the sentence. In Exp 3, we manipulated the target animal role filler: i.e., the target was either depicted as a) it was (no-cue baseline), b) performing an action (action-cue), c) wiggling up and down (wiggle-cue), or performing an action and wiggling (action plus wiggle cue). Visual cues were limited to the lifetime of the verb.

The results (Exp 1, Exp 2) revealed no clear effects of prosody on real-time thematic role assignment in children and adults but we observed slightly more target inspections in a) the biasing prosody conditions (Exp 1: Adults and children) and b) in ambiguous OVS sentences with an SVO-biasing prosody (Exp 2b: Adults). However, adults rapidly used case-marking to anticipate the target role filler as early as the verb (the patient in SVO sentences and the agent in OVS sentences) even though actions were depicted ambiguously (i.e., they did not disambiguate role relations). By contrast, children did not use case-marking for early target anticipation. In both sentence structures, they anticipated the patient (vs. agent). Responses to post-sentences questions corroborated our eye-movement result: Adults' responses were high independent of word order whereas children's response accuracies were only high for SVO sentences but below chance level for OVS sentences.

Action and/plus wiggle cues elicited clear effects on target anticipation during verb and adverb in adults and children (peaking during the verb in adults and the adverb in children). However, adults' offline responses were marginally higher in the action plus wiggle cue conditions (vs. action-cue, wiggle-cue, no-cue baseline) whereas children's responses were higher for active (vs. passive) voice questions. Within the subset of active-only questions, the wiggle-cue elicited marginally more correct responses than in the other conditions.

Taken together, differences between children and adults emerged during thematic role assignment in unambiguously case-marked SVO and OVS sentences (online and offline) but target anticipation was not clearly influenced by prosody in either of these two age groups. Visual cues boosted attentional responses but elicited different accuracy results in children versus adults.

ACKNOLEDGEMENTS

The last years have been an amazing journey. Working towards my Ph.D. has taught me more than I could have ever imagined, not only scientifically but also on a personal level. During my studies, I became more and more fascinated with psycholinguistics. Thanks to all the teachers and colleagues from Paderborn University who have sparkled my interest in psycholinguistic research. Being part of the XPrag community has been a great opportunity. Thanks for all the vivid and interesting discussions during the Annual Meetings. I consider myself lucky for having been able to travel to conferences all over the globe to present my research and exchange thoughts with other researchers.

A very big thank you goes to my supervisor, Pia Knoeferle. She has been amazing with her support and guidance. She was always there whenever I wanted to talk about ideas, problems, or worries. Her support throughout was more than I could have ever asked for. She always put me in the right direction or showed me ways of how to get there and I am very grateful for that.

Thanks to all the colleagues from the (former) Language and Cognition group for all the fruitful discussions during the lab meetings, for all their help and support, but also for the amazing working environment. A very special thanks to Katja Münster and Alba Rodriguez. They have been helpful in so many ways. Katja for guiding me through the first stages of experimental research but also for her dedicated support throughout. Alba for being there and supporting me day in and day out.

The biggest thank you goes to my mother, Barbara Kröger. Without her loving and caring upbringing and her endless support and love I would not be the person I am today and I would not be where I am now. Thanks to my brother, Philipp Kröger who has not only calmed me down on several occasions and had an open ear throughout my Ph.D. years but who also has been there for me all my life. Last but not least, thanks to all my friends for being patient, loving, caring, and encouraging. Thanks for bringing me up when I was feeling down, making me laugh when I was upset, saying the right things in the right moments, and for just being there.

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CIA Coordinated Interplay Account

sCIA social Coordinated Interplay Account

NP1 First Noun-Phrase

NP2 Second Noun-Phrase

SVO Subject-Verb-Object (word order)

OVS Object-Verb-Subject (word order)

SOV Subject-Object-Verb (word order)

OSV Object-Subject-Verb (word order)

aOSV ambiguously case-marked Object-Subject-Verb (word order)

VP Verb-Phrase

NP Noun-Phrase

PP Prepositional-Phrase

SI Scalar Implicature

subj Subject

obj Object

nom Nominative (case)

acc Accusative (case)

1

INTRODUCTION

Language is a means of human communication. When we talk to other people, we can share information. Understanding this information requires at least knowledge about a language's phonetics, phonology, morphology, syntax, semantics, and pragmatics, but communication is not isolated from our environment, what we see can also influence our understanding. Psycholinguistic research has thus focused on the comprehension of spoken utterances in real-time and highlighted the interplay between visual attention and utterance comprehension. A long line of research has provided rich evidence for the incrementality of language processing during which different sources of information (linguistic and non-linguistic cues[1]) can inform language comprehension. The unfolding language guides listeners' attention to visually present objects in the scene within a few hundred milliseconds and these objects can, in turn, influence comprehension processes: e.g., structural disambiguation, semantic interpretation, thematic role assignment (Allopenna, Magnuson, & Tanenhaus, 1998; Chambers, Tanenhaus, & Magnuson, 2004; Knoeferle, Crocker, Scheepers, & Pickering, 2005; Spivey, Tanenhaus, Eberhard, & Sedivy, 2002; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). However, evidence suggests that the online computation for at least some pragmatic (unlike semantic or syntactic) processes is delayed (Huang & Snedeker, 2009a). Additionally, when different types of information in scenes were available, some had similar but others different effects on language comprehension (e.g., Kreysa, Knoeferle, & Nunneman, 2014; Münster, 2016). Crucially, most of the evidence resulted from studies with adult participants.

For children, however, not all aspects of language are present from the beginning. Hence the question is to which extent children process language similar to adults. Over the past thirty years, researchers have investigated exactly this question. Children, much like adults, can relate words to referents in the visual world (Johnson & Huettig, 2011; Johnson, McQueen, & Huettig, 2011). However, a lot of studies with children have relied on offline measures,

1 In this thesis, we use the term *cue* to describe a source of information that can influence utterance comprehension. In the way we use the term, *cues* can be linguistic or non-linguistic (e.g., grammatical cues, syntactic cues, or visual cues). Note that the term does not imply theoretical aspects of language acquisition as it was suggested in the Competition Model by E. Bates and MacWhinney (1987, 1989).

used different materials, or provided no direct comparison between children and adults. Although these studies provide a fruitful insight into child language comprehension, real-time language comprehension in children in a direct comparison to adults has been widely understudied (see Knoeferle, 2015 for a discussion). Trueswell, Sekerina, Hill, and Logrip (1999) provided first evidence for real-time differences between child and adult language comprehension. Children based their interpretation of ambiguous instructions on their syntactic preferences. Unlike adults, children did not rapidly exploit the visual referential context for structural disambiguation, likely because they failed to compute pragmatic inferences online. Similarly, results by Huang and Snedeker (2009b) suggest that children failed to use the visual referential context specifically because they differ from adults in their ability to compute pragmatic inferences online. However, rapid effects of visual context emerged when pragmatic inferences were not required (Münster, 2016; Zhang & Knoeferle, 2012).

1.1 MOTIVATION

In the following line of research, we investigated real-time language comprehension in five-year-old children in a direct comparison to adults. There has been an increasing number of studies on child language comprehension which suggest some similarities (e.g., in their visual attention or in the use of grammatical and syntactic cues; Arnold, Brown-Schmidt, & Trueswell, 2007; Gertner, Fisher, & Eisengart, 2006; Lew-Williams & Fernald, 2007; Mani & Huettig, 2012; Nation, Marshall, & Altmann, 2003) but also some differences (e.g., in the use of case-marking or prosody; Arnold, 2008; Dittmar, Abbot-Smith, Lieven, & Tomasello, 2008a; Grünloh, Lieven, & Tomasello, 2011; Schipke, Friederici, & Oberecker, 2011) between age groups. However, much of the evidence resulted from offline measures such as act-out or video-pointing tasks (e.g., Dittmar et al., 2008a; Grünloh et al., 2011; Meroni & Crain, 2003). Thus, the debate about the extent to which child language processing resembles adult language processing requires additional evidence from real-time measures. How do children compare to adults during real-time language comprehension? Evidence from online child language comprehension suggests differences between age groups in the computation of pragmatic inferences which also seem to be delayed in comparison to other comprehension processes in adults (Huang & Snedeker, 2009a, 2009b; Trueswell et al., 1999).

Another possibility to investigate whether children struggle with pragmatic inferences in real-time and whether pragmatic inferences are delayed in adults is to study the influence of prosody on thematic role assignment using eye-tracking and the visual-world paradigm. Prosody can mark focus via accentuation (i.e., prosodic prominence; e.g., Büring & Gutiérrez-Bravo, 2001; Jackendoff, 1972; Truckenbrodt, 1995). Listeners have to infer why a speaker placed focus on a constituent in the sentence. For example, a stressed first noun-phrase (NP1) is interpreted as object/patient in German, English, and Italian (Grünloh et al., 2011; MacWhinney, Bates, & Kliegl, 1984; Weber et al., 2006) but as subject/agent in Hungarian in NP-verb-NP sentences (MacWhinney, Pléh, & Bates, 1985). Thematic role relations are typically distinguished via case-marking in German. Existing findings from online sentence comprehension revealed that German children can use case-marking for the interpretation of subject-object-verb (SOV) and object-subject-verb (OSV) sentences (Özge, Münster, Knoeferle, Küntay, & Snedeker, 2016). Their scenes, however, contained world knowledge. Thus, it remains unclear whether children can use case-marking when the visual context does not provide world knowledge. Five-year-olds also assigned thematic roles in non-canonical sentences when depicted action events helped to determine 'who does what to whom' (Münster, 2016; Zhang & Knoeferle, 2012). However, it has not been established yet whether this is also true in more difficult-to-process (ambiguous) sentence structures.

Prior research on real-time adult language comprehension further suggests that different visual cues have similar effects (e.g., depicted actions and a speakers' gaze; Kreysa et al., 2014) whereas others have distinct effects (e.g., depicted actions and an emotional prime face; Münster, 2016). To our knowledge, the extent to which distinct visual cues guide visual attention and influence language comprehension has been largely understudied, especially in children. We introduce another visual cue (a wiggling motion) which provides an interesting comparison to the depicted action because it is not mediated by the verb but can function as a pragmatic/focusing cue highlighting the target character (similar to prosody). By introducing a wiggling motion, we can assess whether a visual cue can inform thematic role assignment when its presentation is similarly temporally limited as prosody (a short-lived supra-sentential cue).

The results of our research can potentially shed light on a) whether children and adults can rapidly exploit case-marking and/or prosody for incremental thematic role assignment in visual contexts that are void of world knowledge, b) whether temporally limited visual cues can inform thematic role assignment in difficult-to-process sentences, c) the extent to which distinct visual cues can influence language comprehension and thus provide a better understanding of whether the computation of pragmatic inferences is delayed in adults and one of the main differences between child and adult language processing.

Our results can also potentially influence accounts of visually situated language comprehension which model the interplay between utterance comprehension and visual attention such as the Coordinated Interplay Account (Knoeferle & Crocker, 2006, 2007) or the social Coordinated Interplay Account (Münster & Knoeferle, 2018). As mentioned above, many studies on child language comprehension have not relied on real-time measures. Although prior studies provided rich evidence for the outcome of the comprehension process in children versus adults, it remains unclear how the two age groups compare in real-time. Studying children's processing in a direct comparison with adults' using eye-tracking in the visual world in combination with post-sentence comprehension questions, gives us the opportunity to gain insight into the time course of processing and the outcome of the comprehension process, respectively.

1.2 OUTLINE

Chapter 2 briefly introduces eye-tracking as a measure for real-time language comprehension. We point out the main methodological aspects and describe how eye-tracking has become an important measure for psycholinguistic research. We further review experimental findings on situated language processing (i.e., the interplay between utterance comprehension and visual attention) which highlight the incrementality of language processing by providing evidence for the rapid mediation of visual attention via the semantic or phonological relation between spoken utterances and objects in the scene. We focus on the fact that non-linguistic information (referential context, contrast between objects) can inform syntactic structuring and referential interpretation. We then move towards the idea that listeners can also anticipate upcoming referents based on a verb's restrictions, affordances, and thematic role relations. A final section describes pragmatic processes and argues that some of these are delayed.

The next chapter (Chapter 3) reviews experimental evidence on the influence of different cues on situated language comprehension. We especially focus on literature that is relevant for the design of our experiments (i.e., studies on the influence of case-marking, prosody, and visual cues - depicted actions and a wiggling motion - on thematic role assignment). We discuss that most of the existing evidence has come from studies which investigated only one cue (visual or linguistic) and then highlight studies which have investigated the influence of more than one cue on language comprehension.

Chapter 4 provides the reader with a detailed review of the existing literature on child language comprehension. The first section discusses comprehension similarities between children and adults whereas the second section points out differences between the two age groups. We introduce how children, much like adults, can relate spoken utterances to objects in the real world, even in an anticipatory manner. We point out that, similarly to adults, children can integrate grammatical, syntactic, and prosodic cues during language comprehension. We then move towards the literature that has reported differences between child and adult language processing in their use of case-marking, and prosody and discuss potential reasons for these differences (e.g., the lack of additional visual information). This part is followed by one of the main debates about child language comprehension: Can children use non-linguistic referential context to resolve structural ambiguities of spoken utterances? We then argue that generating pragmatic inferences may be one of the reasons why child language processing does not resemble adult language processing.

In Chapter 5, we shortly summarise the Coordinated Interplay Account which models the close coordination between visual attention and utterance comprehension. We point out the different stages of processing which ultimately lead to the interpretation of an utterance with respect to the visual world. We then describe an extension of the CIA (sCIA) which includes factors such as age. Finally, since we also investigated child vs. adults language processing, we briefly discuss how the results of our research could be implemented in such an account.

Chapters 7, 8, and 9 report six visual world eye-tracking experiments which investigated the influence of morpho-syntactic (case-marking), suprasentential (prosody), and visual (depicted action events, a wiggling target character) cues on real-time thematic role assignment. Two experiments are grouped together within one experimental block (Experiment 1, 2, and 3). In Experiments 1 and 3, we compare young adults with five-year-old children. In Experiment 2, only young adults participated but the two experiments in itself are different. For each of the three experimental blocks, we provide a short introduction to the research questions and summarise existing research.

We then give a description of the participants, materials, procedure, analyses, and predictions. The methodological part is followed by a detailed description of the results. Towards the end of each block, we discuss the results with reference to existing findings and provide a short summary of the main findings.

Chapter 10 discusses the findings of our experiments in a broader context. The chapter is divided in five sections. In the first three sections (case-marking, prosody, and visual cues), we summarise and discuss our results in the context of existing findings. In the section following these three, we summarise the real-time age differences we observed in our experiments and compare them to previous findings. The last section provides implications of our results for the Coordinated Interplay Account (Knoeferle & Crocker, 2006; Knoeferle et al., 2005). We propose another possible component that could be included in the account.

In the last chapter (Chapter 11) we summarise our main points, discuss the implications of our findings, and propose ideas for future research.

Part I

LITERATURE REVIEW

2

SITUATED LANGUAGE PROCESSING

Communication takes place everywhere in the world. We talk about what we see, but what we see also influences what we say and how we say it. A large and growing body of literature has investigated how language is influenced by the visual world and also how the visual world influences our interpretation of spoken utterances (e.g., Altmann & Kamide, 1999; Chambers et al., 2004; Knoeferle et al., 2005; Spivey et al., 2002; Tanenhaus et al., 1995). This chapter provides an overview of eye-tracking as a measure for language comprehension and a review of studies on situated language processing which highlight the close coordination between language comprehension and visual attention.

2.1 EYE-TRACKING AND THE VISUAL WORLD PARADIGM

In eye-tracking studies, participants listen to a spoken utterance (e.g., sentences or words) and simultaneously inspect a visual display. The visual display can consist of real world objects or real world depictions of objects or scenes (e.g., Tanenhaus et al., 1995, Cooper1974; see Huettig, Rommers, & Meyer, 2011 for a discussion). During these studies, participants' eye-movements are monitored. Usually, eye-tracking systems determine the location of the pupil relative to the corneal reflection, generated via infra-red light. Eye-tracking is thus not a very intrusive measure (Goldberg & Wichansky, 2003). Participants engage in simple *look-and-listen* or *direct-action* tasks and, especially in developmental studies, comprehension questions (e.g., 'who does what to whom?') follow the spoken utterance (see Huettig et al., 2011 for a discussion). Clear and simple instructions at the beginning of an experiment are very important to create an understanding of the task at hand. Crucially, Altmann and Kamide (1999) observed differences in the timing of target object fixations when instructions were either a) to verify whether the sentence applies to the picture, or b) to pay no attention to the spoken utterance.

Research on online language comprehension has shown that non-linguistic context can inform language processing. Comprehenders rapidly relate spoken words to referents in the visual world when these referents are semantically or phonologically related to the linguistic input (Allopenna et al., 1998;

Cooper, 1974). In 1974, Cooper discovered that when people listen to spoken utterances, they move their eyes towards an object in the visual world that is semantically related to this utterance. Upon hearing the word *lion*, listeners directed their gaze towards a lion in the visual display. When participants heard the word *Africa*, they inspected the lion, the zebra, and the snake. Cooper (1974), therefore, argued that eye-movements can be used to study language comprehension. Thus, the foundation for a new methodology of studying language processing was laid. Huettig and Altmann (2005) suggested that eye-movements towards an object in the visual world can further be mediated by the conceptual knowledge of objects in the visual scene. Upon hearing the word *piano*, semantic information was activated and visual attention mediated towards a conceptually related object in the visual scene (e.g., a trumpet).

Allopenna and colleagues (1998) investigated the phonological relationship between eye-movements and the visual world. They used real world depictions of objects to examine whether cohort competitors are activated during instructions such as *Pick up the beaker...now put it above the triangle.* On a computer screen, participants saw scenes which depicted referents (e.g., a beaker), cohorts (e.g., a beetle), rhymes (e.g., a speaker), and unrelated objects (e.g., a dolphin) in a 5x5 grid. The results revealed more fixations on the referent, cohort, and rhyme object than on the unrelated object with more and earlier fixations on the cohort object than on the rhyme object. Crucially, these fixations started about 300 ms after word onset, underlining the incrementality of language processing. The results further provided evidence for the linking hypothesis between spoken utterances and eye-movements (Allopenna et al., 1998).

Taken together, during real-time language comprehension, spoken utterances mediated visual attention towards semantically or phonetically related objects in the immediate visual environment. Crucially, these utterance-mediated fixations occurred within 300 ms after word onset which underlined the incrementality of language processing. However, not only linguistic information can elicit rapid target fixations.

2.2 THE ROLE OF REFERENTIAL CONTEXT/CONTRAST

Although Cooper (1974) provided the first evidence for the referential link between language and the visual world, the close relation between eye-movements and language processing was only widely acknowledged after

Tanenhaus and colleagues published their paper in 1995 (Huettig et al., 2011). Using a real-world set-up, Tanenhaus et al. (1995) investigated whether visual referential context can mediate visual attention in interaction with the linguistic input. In more detail, they tried to find out whether referential context can help adult comprehenders to disambiguate spoken instructions online (e.g., *Put the apple on the towel in the box*). The prepositional-phrase (PP) *on the towel* is ambiguous such that it can either be interpreted as the location of *the apple* or as the destination for *the apple*. In the absence of visual context, comprehenders have a preference for a goal interpretation (attaching the prepositional-phrase *on the towel* into the verb-phrase (VP) *Put the apple*).

During the experiment, participants inspected either one or two referents in a visual display (one referent: An apple on a towel, an empty napkin, a box; two referents: One apple on a towel, another apple on a napkin, and a box). The one-referent context supports a destination interpretation whereas the two-referent context supports a location interpretation of the prepositional-phrase *on the towel*. Participants' gaze-pattern suggested that they immediately interpreted *on the towel* as a destination for *the apple* in the one-referent context and as a location of *the apple* in the two-referent context. Tanenhaus et al. (1995) argued that, in a realistic situation, comprehenders rapidly integrate non-linguistic information (the number of apples in the visual display) during online language comprehension. Since reference was established immediately, the results further supported the incrementality of language processing. Upon hearing an ambiguous prepositional-phrase, participants arguably inferred from the visual context that the prepositional-phrase is a modifier of one of the apples.

Eberhard, Spivey-Knowlton, Sedivy, and Tanenhaus (1995) further examined incremental reference resolution but explored complex noun-phrases. Participants listened to instructions (e.g., *Touch the starred yellow square*) whilst inspecting simple visual scenes which either disambiguated the instruction early (four squares but only one was yellow and starred), mid (four starred squares but only one was yellow), or late (two starred and yellow objects, one square and one rectangle). Disambiguating information about the target referent was either provided by the marking adjective (early), the colour adjective (mid), or the target noun (late). The results revealed earlier eye-movements towards the target in a) the early than in the mid and late condition and b) the mid than in the late condition. Thus, Eberhard and colleagues (1995) suggested that language processing is incremental (moment to moment) and influenced by the non-linguistic context. They further proposed that eye-tracking is a reliable measure for language processing since it monitors each moment in time.

Spivey et al. (2002) examined rapid referential context effects on ambiguity resolution but depending on the number of potential referents in the visual environment. The authors used a similar design to the one in Tanenhaus et al. (1995) but added a three-referent context (one apple on a towel and three other apples which were not on a towel or napkin) to the existing one- and two-referent contexts. The authors argued that in the two-referent context *the apple* is temporally ambiguous such that it could refer to the apple on the towel or the apple on the napkin which might have resulted in looks to the empty towel (goal interpretation) when the wrong referent was fixated at the beginning of the instruction (the apple on the napkin). In the three- and one-referent contexts, however, the apple on the towel can be immediately identified as the referent.

Upon hearing *Put the apple on the towel...* participants looked at the empty towel (goal interpretation) in the one-referent context. In the two-referent context participants looked at the box slightly later (modifier interpretation) in the ambiguous condition compared to the unambiguous condition. In the one- and three-referent context no such delay was found, indicating that the single apple on the towel was directly identified as the referent. However, in both the two- and three- and one-referent contexts, participants preferred a modifier interpretation of the instruction although the verb *put* is, in general, more often associated with a goal interpretation. The findings suggest that listeners rapidly use referential context (scene information) to overcome syntactic preferences (modifier interpretation instead of goal interpretation). Their research thus contributed to the idea that comprehenders use the number of referents in the visual context for incremental ambiguity resolution. In other words, non-linguistic information immediately influenced real-time language comprehension.

More evidence for the influence of non-linguistic information on real-time language comprehension comes from a study by Chambers et al. (2004). The authors examined whether the affordances[1] of objects also facilitate ambiguity resolution (similar to the objects themselves). While listening to instructions such as *Pour the egg in the bowl over the flour*, participants inspected scenes

1 According to Gibson (1977) "an affordance is an invariant combination of variables" (p. 134) which are based on experiences: The affordances of a glass, for example, can be that people can drink from it, pick it up, or that it can contain liquid, etc.

containing two eggs (one in a glass and one in a bowl), an empty bowl, and some flour. In one condition (compatible competitor condition) both eggs were in liquid form whereas in another condition (incompatible competitor condition) one egg was liquid (can be poured) and one was solid (cannot be poured). The results revealed more fixations on the incorrect goal object (i.e., the empty bowl) when the scene depicted one liquid and one solid egg (vs. two liquid eggs). The authors suggested that comprehenders do not only interpret sentences on the basis of the linguistic input but are also influenced by non-linguistic information (i.e., knowledge about an object's affordances) during ambiguity resolution.

In their early works, Cooper (1974), Tanenhaus et al. (1995), and Allopenna et al. (1998) established the interaction of linguistic and non-linguistic input. Not only do listeners rapidly inspect referents or objects on the basis of their semantic or phonological relation to spoken words (Allopenna et al., 1998; Cooper, 1974), they also exploit the number of referents in the visual world and their knowledge about these referents to disambiguate syntactic structures (Chambers et al., 2004; Spivey et al., 2002; Tanenhaus et al., 1995). Crucially, the authors have shown that language processing is incremental and that this moment-to-moment processing can be monitored using eye-tracking. Our interpretation of spoken utterances is influenced by the non-linguistic visual context when the utterance relates to this visual context.

2.3 anticipation

At the beginning of this chapter, we pointed out that visual attention in scenes can be rapidly mediated by spoken utterances. Crucially, a number of studies by Altmann and Kamide (1999, 2007) and Kamide, Altmann, and Haywood (2003) revealed that comprehenders direct their visual attention towards an object in the visual scene, not only rapidly but even before the target word is mentioned. Altmann and Kamide (1999) observed that participants directed their eye-movements towards an object, a *cake*, in the visual world (a scene depicting a boy, a cake, and other objects) even before it was mentioned in the sentence (e.g., *The boy will eat the cake*). The results further revealed that the onset of participants' eye-movements towards the *cake* were earlier in sentences containing the verb *eat* compared to sentences containing the verb *move*. In other words, verb information was rapidly integrated to anticipate the only edible object in the scene: the *cake*. The authors argued that listeners rapidly formed expectations about the upcoming linguistic input and that these expectations were influenced by the constraints of the verb and its argument structure in relation to the visual context.

Kamide, Altmann, and Haywood (2003) then investigated whether anticipation during language processing is based on more than just the verb's constraints. Participants heard future tense sentences (e.g., *The man will ride the motorbike* or *The girl will ride the carousel*) while inspecting a scene depicting a girl, a man, a motorbike, a carousel, and a distractor object. If the restrictions of the verb alone facilitated anticipatory eye-movements towards the target, then participants should have looked equally towards both ridable objects (the motorbike and the carousel). However, upon hearing *The man will ride...*, participants looked more towards the motorbike (vs. the carousel) and more towards the carousel upon hearing *The girl will ride...*. The authors thus suggested that not only verb information (the verb *ride* requires a theme which could be either the carousel or the motorbike) but also information about the agent of the sentence play a role in anticipatory processing. It is more plausible that a girl rides a carousel than a motorbike and that a man rather rides a motorbike than a carousel. Thus, anticipatory eye-movements towards an object or referent in the visual world are driven by the verb's restrictions and the plausible thematic roles in the visual display.

Additionally, anticipatory eye-movements are driven by the tense of the sentence in combination with the objects' affordances (Altmann & Kamide, 2007). In the verb region, participants looked more at a full glass of beer (vs. an empty glass of wine) when hearing the sentence *The man will drink...* . When they heard *The man has drunk...*, they looked more at the empty glass of wine (vs. the full glass of beer). The authors thus argued that the affordances of the empty wine glass (the wine has been drunk), the full glass of beer (the beer can still be drunken), and the tense information of the verb, motivated participants eye-movements towards a referent in the visual scene although this referent was not yet mentioned in the linguistic input. If it was only the restrictions of the verb, participants would have looked equally as much at the wine and the beer glass (both are objects we drink from). Moreover, mental representations of an objects' location can be updated by the linguistic input and, in turn, influence anticipatory eye-movements (e.g., towards the table; Altmann & Kamide, 2009). When participants heard sentences such as *The woman will put the glass onto the table...* (moved condition) or *The woman is too lazy to put the glass on the table...* (unmoved condition) *Then/Instead, she will pick up the bottle, and pour the wine carefully into the glass*, they inspected the table more in the moved (vs. unmoved) condition. However, the overall number of fixations on the glass was higher than on the table. The authors argued that the glass mentioned in the linguistic input and the one depicted in the scene biased more looks towards the glass than the table.

Taken together, visual attention is also influenced by linguistic and world knowledge and can be anticipatory. A verb's semantics or tense, plausibility, or an objects' affordances can mediate attention towards objects in the visual display even before the object is mentioned in the linguistic input. In other words, comprehenders form expectations about upcoming linguistic expressions based on diverse informational sources. However, during language comprehension, we may also want to consider the speakers intended meaning which can sometimes be different to what we initially assumed.

2.4 THE ROLE OF PRAGMATICS

During language comprehension, the semantic meaning of word plays an important role. Visual attention can be mediated by the semantic information in the linguistic input. However, meaning can be expressed beyond the semantic content of a spoken utterance. Another line of research thus investigated the role of a speaker's intended meaning and the immediate visual context on real-time language comprehension. The issue in question was whether pragmatic interpretation is similarly incremental as semantic interpretation.

Following the design by Eberhard et al. (1995), Sedivy, Tanenhaus, Chambers, and Carlson (1999) tried to find evidence for incremental semantic interpretation but with less predictable materials than in Eberhard et al. (1995). In their first experiment, participants were instructed to *Touch the blue pen* whilst inspecting a blue pen, a yellow rubber duck, a red notebook, and a pink comb (early disambiguation condition - adjective disambiguated), or a blue pen, a blue bowl, a yellow duck, and a red notebook (late disambiguation condition - noun-phrase disambiguated). The results corroborated the findings by Eberhard et al. (1995): The interpretation of adjectives is incremental with respect to the objects available in the visual display. Participants identified the target earlier when the display contained only one possible referent (early disambiguation condition) than when it contained two possible referents for the adjective *blue* (a blue pen and a blue bowl). However, when the instructions contained contrastive information (contrastive stress on the adjective, e.g., *Touch the pink comb. Now touch the YELLOW comb/bowl*), participants rather used the contrastive function of the adjectival modifier than the contrastive stress to disambiguate which object is being referred to. The authors argued that adjectival modifiers likely reach ceiling effects for a contrastive interpretation and therefore the contrastive stress did not additionally facilitate comprehension.

In a second experiment, the authors investigated whether the interpretation 'vague scalar adjectives' (e.g., *tall* or *short*) is similarly incremental than the interpretation of colour adjectives. Adjectives such as *tall* or *short* are vague in comparison to adjectives such as *blue* or *yellow*. The interpretation of vague adjectives requires a comparison to other things or objects. Participants were instructed to *Pick up the tall glass and put it below the pitcher* whilst inspecting a visual scene that contained a contrasting object (a small glass), a distractor object (a key), a target object (a tall glass next to a shorter glass), and a competitor object (a pitcher). In half of the trials, a competitor object was present, whereas in the other half, an unrelated object was presented. The results showed that participants looked to the target object earlier when a contrastive object was present (vs. absent). When the contrastive object was absent, early looks to the competitor object were observed. Comprehenders established a contrastive interpretation immediately by using information from the visual display (e.g., the small glass) or representative information in memory (e.g., the knowledge that a pitcher is a tall glass) even before the head noun was mentioned (Sedivy et al., 1999). Listeners generated pragmatic inferences such that the target object belongs to a group of objects from the same category (see Huang & Snedeker, 2009a for a discussion). Sedivy et al. (1999) argued that semantic interpretation, syntactic processing, and pragmatic inferences happen immediately upon hearing the adjective.

By contrast, Huang and Snedeker (2009a) argued that it is yet unclear whether semantic interpretation occurs before or simultaneously to pragmatic inferences. The authors suggested that in Sedivy et al. (1999), participants first had to establish that *tall* is a scalar adjective and could then use the contrasting objects in the visual display. Perhaps there is a point in time when the semantic interpretation is available but the pragmatic inference is not. Therefore, Huang and Snedeker (2009a) investigated the interface of semantic and pragmatic processing using scalar quantifiers. Participants were instructed to *Point to the girl that has some/two/three/all of the socks*. The display contained two boys and two girls. One of the girls had two of four socks and another girl who had three of three soccer-balls. Thus, the items distributed among the characters had the same onset (socks - soccer-balls) which created a temporal ambiguity of the instruction. In another experiment, the distribution of objects was varied such that in a two-referent context, a girl had three socks and another had three soccer-balls, and in a one-referent context, one girl had three socks and the other girl had no socks or soccer-balls. In Experiments 1 and 2, the authors observed that the spoken utterance had a strong influence on participants' reference resolution. In more detail, participants were quicker to look at the target character when semantics clearly identified the target (two, three, and all trials). However, in some trials, in which the semantic interpretation was not enough to identify the target, participants took longer to establish reference. Such a delay in reference resolution was not found in one-referent contexts. The authors thus argued that a temporal lag between semantic and pragmatic processing exists: Semantic interpretation occurs prior to pragmatic inferences.

In summary, the literature reviewed in this section provided evidence for incremental pragmatic interpretation. However, the results of the two studies were somewhat contradictory. Sedivy et al. (1999) concluded that pragmatic inferences were computed immediately (similar to the semantic interpretation) whereas the findings by Huang and Snedeker (2009a) suggested a temporal lag between the two. Similar to Sedivy et al. (1999), Grodner, Klein, Carbary, and Tanenhaus (2010) argued that when the context supports the interpretation sufficiently, this temporal lag between semantic processing and pragmatic inference disappears. All in all, the extent to which pragmatic inferences are delayed remains an open question.

2.5 SUMMARY

In summary, the studies reviewed in this chapter provided evidence for the interplay between spoken utterances and visual attention. Linguistic and world-knowledge can influence visual attention and once attended non-linguistic (visual) information can influence utterance interpretation. Crucially, comprehenders direct their attention towards objects or referents in the visually present scene within a few hundred milliseconds which emphasises the incrementality (moment to moment) of language processing; i.e., visual attention is closely time-locked to utterance comprehension. In more detail, we have pointed out that comprehenders relate words to objects or referents in the visual display when this object or referent is semantically or phonologically related to the utterance. Non-linguistic information such as referential context, or contrast between objects can inform online syntactic structuring and contrastive interpretation. Listeners further form expectations about upcoming thematic roles on the basis of a verbs' restrictions, thematic plausibility, and verb tense. Although the core findings are that syntactic and semantic processing are incremental, other experimental evidence suggests that the online computation of (at least some) pragmatic processes is delayed. This tension between semantic and pragmatic processing remains an open issue at this point. What we also do not know is to which extent other kinds of information such as prosody, case-marking, and visual cues influence other processes such as thematic role assignment in real-time. The following chapter thus provides a review of existing findings on how different cues (case-marking, prosody, and visual cues) can be rapidly integrated during real-time thematic role assignment/language comprehension.

3

CASE-MARKING, PROSODY, AND (NON-LINGUISTIC) VISUAL CUES

The assignment of thematic roles ('who does what to whom') can be challenging, especially in a language with a relatively free word order such as German. In general, German transitive sentences consist of a subject, a transitive verb (predicate), and a direct object. Nouns in German are inflected for case (nominative, accusative, dative, genitive) and number (singular and plural) which is realised mostly on the determiner. Case-marking of masculine nouns is clearly distinguishable in nominative (*der*) and accusative (*den*) case. For feminine and neuter nouns, case-marking is identical in both nominative (*die*, *das*) and accusative (*die*, *das*) case (Haider, 2010). Thus, in a sentence such as *Die Katze jagt die Biene* ('The cat chases the bee') it is not clear 'who does what to whom'. The cat can be either the subject and the agent (SVO sentence structure), or object and patient (OVS sentence structure).

In the sentence *Die Katze jagt der Biber* ('The cat chases the beaver') masculine case-marking on the second noun-phrase disambiguates 'who does what to whom'. Case-marking in German can thus create local and global ambiguities. When case-marking on the determiner of the first noun-phrase is ambiguous, other types of information can be used to distinguish 'who does what to whom'. In English word order is less flexible (subject-verb-other argument) and, therefore, comprehenders can rely more on word order to distinguish 'who does what to whom' (Hemforth, 1993; Hemforth & Konieczny, 2013). In this chapter, we review empirical evidence which suggests the rapid influence of case-marking, prosody, and visual cues on situated language comprehension. Most of the evidence comes from studies on the influence of one single cue (visual or linguistic). We thus further review literature examining the effects of more than one cue (visual and linguistic) on utterance comprehension.

3.1 CASE-MARKING

Thematic role ambiguity is typically resolved via case-marking in German. The subject mostly takes nominative case and stands in agreement with the verb whereas the object takes accusative case and does not stand in agreement with the verb (Dürr & Schlobinski, 2006). When case-marking is ambiguous (nominative/accusative), comprehenders tend to interpret case-marking on the first noun-phrase as nominative because SVO sentence structure is generally preferred over OVS (Dittmar et al., 2008a; Hemforth, 1993; Hemforth & Konieczny, 2013). A number of studies suggest that OVS sentence structure is more difficult to process than SVO (e.g., Fanselow, 2000; Fiebach, Schlesewsky, & Friederici, 2001, see Bornkessel, Schlesewsky, & Friederici, 2002 for a discussion). Frequency, linguistic distinction, and patient before agent ordering are possible reasons for why OVS (vs. SVO) sentence structure may be more difficult to process (Bornkessel et al., 2002; Ferreira, 2003). Nevertheless, when unambiguous, case-marking is believed to be a strong cue for thematic role assignment in German in both SVO and OVS sentence structures (Hemforth, 1993; Kamide, Scheepers, & Altmann, 2003; Matzke, Mai, Nager, Rüsseler, & Münte, 2002).

In an ERP-study[1], Matzke et al. (2002) examined how case-marking can inform sentence processing during reading. Participants read unambiguous and ambiguous German SVO and OVS sentences (e.g., unambiguous: *Der/Den begabte(n) Sänger entdeckte den/der talentierte(n) Gitarrist(en)* - 'The gifted$_{\text{Masc.Nom./Acc.}}$ singer discovered the talented guitar player$_{\text{Masc.Acc./Nom.}}$'; ambiguous: *Die begabte Sängerin entdeckte den/der talentierte(n) Gitarrist(en)* - 'The gifted$_{\text{Fem.Nom.?/Acc.}}$ singer discovered the talented guitar player$_{\text{Masc.Acc./Nom.}}$'). The sentences were presented word by word on a video monitor. The results indicated that case-marking in non-canonical word order (OVS) has an immediate effect on how sentences are processed (they require a greater load of working memory than SVO sentences). One of the reasons for this might be that SVO is, in general, the preferred word order in German (Dittmar et al., 2008a; Hemforth, 1993; Hemforth & Konieczny, 2013).

In a visual world eye-tracking study, Kamide, Scheepers, and Altmann (2003) investigated the influence of syntactic (case-marking) and semantic (verb) constraints on thematic role assignment. Participants listened to unambiguously case-marked German SVO or OVS sentence (e.g., SVO: *Der Hase frißt gleich den Kohl* - 'The hare-nom eats shortly the cabbage-acc'; OVS: *Den Hasen frißt gleich der Fuchs* - 'The hare-acc eats shortly the fox-nom') accom-

1 "An ERP-component can be operationally defined as a set of voltage changes that are consistent with a single neural generator site and systematically vary in amplitude across conditions, time, individuals and so forth." (Luck, 2014, page 68). During language comprehension: The N400 is a negativity that peaks 400 ms after word onset and represents the processing of semantic incongruity (Kutas & Hillyard, 1980). The P600 is a positivity which occurs in sentences with: Syntactic violations, non-preferred structures, or complex syntactic structures (Osterhout & Holcomb, 1992)

panied by a scene that contained a hare, a fox, and a cabbage. Participants looked more at the cabbage (vs. fox) in SVO sentences and more at the fox (vs. cabbage) in OVS sentences. The results suggest that case-marking, in combination with constraints of the verb and world knowledge, influence thematic role assignment in real-time.

Furthermore, in Grünloh et al. (2011), the adult control group (N=10) used case-marking (rather than prosody) for thematic role assignment but not online. In unambiguously case-marked OVS sentences adult participants interpreted the first noun-phrase as the patient whereas in ambiguously case-marked OVS sentences, they interpreted the first noun-phrase as the agent (SVO interpretation).

Overall, case-marking seems to be a strong cue for thematic role assignment in combination with or without world knowledge. Although the previously described studies used different measures, the effects were similar: Morpho-syntactic information can rapidly influence thematic role assignment. However, case-marking in German can also be ambiguous (feminine and neuter gender are identical in nominative and accusative case). When such ambiguity occurs, other, possibly more subtle cues, inform language comprehension.

3.2 PROSODIC CUES

Language comprehension involves more than just a knowledge about grammar. We often have to consider a speaker's intended meaning. Prosody, among others, is a tool that can help us with our interpretation of speech. It can, however, have different functions. Prosodic boundaries, for example, can influence ambiguity resolution (attachment and consticuency) whereas pitch accents can be exploited for contrast and reference (Carlson, 2009). Crucially, prosody can help listeners to distinguish 'who does what to whom'. Among others, prosody can mark focus. Focus can, in turn, be broad, narrow, or contrastive. Depending on the type, focus can indicate new, given, or contrastive information via phonetic variation or duration (e.g., Baumann, Becker, Grice, & Mücke, 2007; Hanssen, Peters, & Gussenhoven, 2008; Kügler, 2008; see Sauermann, Höhle, Chen, & Järvikivi, 2011 for a discussion). Focus marking can also help listeners to distinguish between different word orders. In German unmarked constructions, a clause initial subject is a Topic and a clause final object a Focus (e.g., Lambrecht, 1996). If the first noun-phrase of a sentence is accented, listeners are likely to interpret this as a signal for a deviation from canonical word order (e.g., Grünloh et al., 2011; MacWhinney et al., 1984; Weber et al., 2006). However, a direct mapping between focus (a stressed NP1) and object/patient does not apply across languages (e.g., Du Bois, Kumpf, & Ashby, 2003; Skopeteas & Fanselow, 2009). Evidence from Hungarian, for example, suggests that focus of a pre-verbal constituent signals default sentence structure, therefore a subject/agent and not an object/patient (MacWhinney et al., 1985). Given this asymmetry between information structure and thematic roles, prosodically marked focus initiates pragmatic inferences. Existing findings on real-time language comprehension have revealed that listeners rapidly integrate supra-segmental (prosody) information for syntactic disambiguation, or reference resolution (Dahan, Tanenhaus, & Chambers, 2002; Schafer, Speer, Warren, & White, 2000; Snedeker & Trueswell, 2003; Weber et al., 2006). However, prosodic cues seem to be more effective in some instances and less in others. This is likely the case because the contribution of prosody on speech processing is diverse and less fixed than that of, for example, case-marking.

For example, when Sedivy et al. (1999) investigated the influence of prosody on reference resolution, they observed no clear effects. However, Dahan et al. (2002) provided evidence for prosodic effects on reference resolution. If we look at the designs of the two studies, we may be able to explain why their results differed. In Sedivy et al. (1999), participants followed instructions such as *Touch the pink comb...* then *Now touch the YELLOW comb* whilst inspecting a visual display which included a yellow comb, a pink comb, a yellow bowl, and a knife. The authors argued that the lack of effects of prosody on reference resolution might have resulted from the contrast provided in the visual display and the linguistic input. The first instruction already included a comb which likely influenced looks to the yellow comb in the second instruction. It seems that participants rather used the contrastive function of the adjectival modifier than the contrastive stress to disambiguate the object being referred to.

In Dahan et al. (2002), eye-movements were monitored while participants listened to instructions such as *Put the candle/candy below the triangle* (candle - anaphoric, candy - non-anaphoric), followed by *Now put the CANDLE above the square*, or *Now put the candle ABOVE THE SQUARE*. Thus, the accent in the second instruction either referred to an unmentioned (above the square) or an already mentioned (candle) referent. Visual displays contained four fixed geometric shapes and four moveable objects, among them the target and competitor objects (candle and candy). Fixation patterns suggest that listeners looked more towards the cohort competitor object (candy - onset overlaps onset of candle) in the anaphoric condition (vs. non-anaphoric) when it was introduced by a pitch accent (candle-CANDLE). In the non-anaphoric condition, participants inspected the previously mentioned object more when it carried no pitch accent (candy/candle). Dahan et al. (2002) demonstrated that pitch accents influence reference resolution such that accentuation leads to a non-anaphoric interpretation and de-accentuation leads to an anaphoric interpretation.

Furthermore, evidence by Snedeker and Trueswell (2003) suggests that listeners use prosodic cues to disambiguate sentences such as *Tap the frog with the flower*. In a referential communication task, speakers used prosody only in sentences that were ambiguous (vs. unambiguous: e.g., *Tap the frog by using the flower*) and in referential scenes which supported both, an instrument or modifier interpretation (i.e., a flower, a frog holding a flower, another frog, a giraffe holding a different object, and a lego block). Listeners then used this prosody to anticipate the correct upcoming referent. The authors argued, that prosodic cues are unreliable because they were only produced when no other disambiguating information was available. However, listeners exploited them when they were available. Prosodic cues might be less frequent (they were only produced in ambiguous structures) but nevertheless valid for disambiguation (Snedeker & Trueswell, 2003). The authors pointed out that the results of this study contradict the existing findings by Schafer et al. (2000) who showed that prosody was used as a cue for syntactic disambiguation in both ambiguous and unambiguous constructions.

Weber et al. (2006) demonstrated that prosody can rapidly influence the assignment of grammatical functions in German. In a visual world eye-tracking study, participants listened to ambiguously case-marked German SVO (e.g., *Die Katze jagt womöglich den Vogel* - 'The cat$_{\text{ambiguous nominative}}$ chases possibly the bird$_{\text{accusative}}$') and OVS sentences (e.g., *Die Katze jagt womöglich der Hund* - 'The cat$_{\text{ambiguous accusative}}$ chases possibly the dog $_{\text{nominative}}$). Feminine case-marking is identical in nominative and accusative case in German, creating local structural ambiguity. The determiner on the second noun-phrase, however, disambiguated the sentence - masculine and neuter case-marking is different in nominative (*der*) and accusative (*den*) case. The visual scenes depicted a cat, a dog, and a bird. Thus, world knowledge implied stereotypical role relations (i.e., cat chase birds and dogs chase cats).

Sentences were assigned SVO- and OVS-biasing prosodic contours following GToBI (German Tones and Break Indices) which is a tool to describe the phonological structure of German intonation. It is related to the English ToBI which is based on autosegmental- and metrical phonology (distinguishing between two basic levels of intonation: H for high tones and L for low tones). In more detail, GToBI includes the description of tones, break indices, and words (Baumann, Grice, & Benzmüller, 2000). For this thesis, the focus was placed on the description of tonal events (i.e., pitch accents) for the description of the prosodic structure used in the experiments. GToBI includes two monotonal (H*, L*) and four bitonal pitch accents (L+H*, L*+H, H+L*, H+!H*; Baumann et al., 2000). The SVO-biasing prosodic contour contained an L*+H accent on the first noun-phrase and an H* accent on the verb and the OVS-biasing prosodic contour an H* accent on the first noun-phrase. In the adverb region of the sentence, participants inspected the patient (vs. agent) more in SVO sentences with an SVO-biasing prosody. In OVS sentences assigned an OVS-biasing prosody, more looks to the agent (vs. patient) were observed. Thus, prior to the disambiguating second noun-phrase (NP2), prosody helped participants to overcome their strong SVO preference in OVS sentences carrying an OVS-biasing prosody.

In summary, comprehenders can use prosody to infer an intended meaning of an utterance. Especially in ambiguous constructions, prosody seems to be helpful to single out an intended referent or 'who does what to whom'. Although prosody is likely a more subtle cue than, for example, case-marking (which is tightly linked to the syntactic structure), in the absence (although in some instances also in the presence) of other disambiguating information, listeners rapidly integrate supra-sentential information during online language comprehension. In chapter 2, we reviewed literature that provided evidence for the influence of both linguistic and non-linguistic information on situated language comprehension. The extent to which scene information is closely time-locked to utterance interpretation is another question which we discuss in the following section.

3.3 VISUAL CUES

Information from the visual world can also help listeners determine 'who does what to whom'. Imagine someone tells you *Guck mal, die Katze jagt die Hündin* ('Look, the (feminine) cat chases the (feminine) dog'), from word order alone, you would assume that the cat is the agent and the dog the patient (SVO word order is preferred). However, it is very unlikely that a cat chases a dog. Thus, if you then see a dog chasing a cat, you can infer that the speaker varied the word order (object-verb-subject word order is possible in German) and with the help of the context, it is clear 'who does what to whom'. In Chapter 2, we saw that in rich contexts, the visual context can be supportive for incremental reference resolution (Spivey et al., 2002), thematic role assignment (Kamide, Altmann, & Haywood, 2003), or syntactic structuring (Tanenhaus et al., 1995). However, in these studies, the objects in the visual scene were depicted as one entity each. Settings in the real world, however, often provide more information than just objects (Knoeferle et al., 2005).

Knoeferle et al. (2005) thus investigated the combined influence of the visual (depicted action events) and the linguistic (verb information) input on incremental thematic role assignment. Unlike studies reported by Tanenhaus et al. (1995) and Sedivy et al. (1999), in this study, the reference to the object in the visual scene was unambiguous (e.g., only one princess was depicted). In contrast to Altmann and Kamide (1999), Altmann and Kamide (2007), Kamide, Altmann, and Haywood (2003), and Kamide, Scheepers, and Altmann (2003), the scenes depicted action events (e.g., for the verb washing a sponge and a bucket were depicted) and not only the objects which identified thematic role relations by the constraints of the linguistic input (e.g., agent and verb information).

In Knoeferle et al.'s study (2005), participants listened to locally ambiguous German SVO and OVS sentences (e.g., SVO: *Die Prinzessin wäscht / malt offensichtlich den Pirat / der Fechter* - 'The (agent/patient) princess washes / paints apparently the (patient) pirate / (agent) fencer'). Case-marking on the second noun-phrase disambiguated who the agent or patient of the sentence was. Scenes depicted three characters (a princess, a fencer, and a pirate). The princess in the scene was role ambiguous because she acted upon the pirate and was being acted upon by the fencer (agent and patient role). During the verb region of the sentence, participants anticipated the agent (the fencer) in OVS sentences - the princess was being acted upon by the fencer - and the patient (the pirate) in SVO sentences - the princess was acting upon the pirate. Knoeferle et al. (2005) suggested that participants disambiguated sentence and role ambiguity via depicted action events in the immediate visual scene. However, when the scenes depicted action events and included thematic knowledge (e.g., a non-stereotypical agent, a wizard, performing a spying action and a stereotypical agent, a detective, performing a serving action) participants relied on depicted action events rather than stereotypical knowledge about likely agents. Upon hearing *Den Piloten bespitzelt...* ('The pilot (ACC) spies...'), participants anticipated the non-stereotypical agent but depicted as performing the spying action (the wizard) more than the stereotypical agent depicted as performing a serving action (the detective; Knoeferle & Crocker, 2006).

Münster (2016) contributed to the idea that depicted actions can inform thematic role assignment in German. Participants heard unambiguous OVS sentences (e.g., *Den Marienkäfer kitzelt vergnügt der Kater* - 'The (agent) ladybug tickles happily the (patient) cat') while inspecting a visual scene depicting three clip-art animal characters: One target character (a cat), one role-ambiguous character (a ladybug), and one distractor character (a rat). Actions were either depicted (for the verb 'tickle' the cat was holding a feather towards the ladybug) or not depicted. Post-sentence questions elicited participants offline performance. Early anticipatory eye-movements towards the target characters were found across different age groups (younger adults, five-year-old children, and older adults). However, depicted actions only increased correct post-sentence answers in children and older adults. The author argued that adults had no difficulty determining 'who does what to whom', resulting in ceiling effects for comprehension questions.

Although case-marking, prosody, and visual cues can all inform incremental thematic role assignment, the processes involved are arguably different. Unambiguous case-marking in German immediately signals whether the first noun-phrase of the sentence is subject and agent (nominative case-marking) or object and patient (accusative case-marking). Prosody, however, requires pragmatic inferences. When something is prosodically marked, comprehenders infer that the intended meaning deviates from their initial interpretation. Depicted action events are mediated by the verb and thus tightly time-locked to the spoken utterance. Comprehenders can rapidly integrate this information to distinguish 'who does what to whom'. Overall, listeners can exploit a variety of different cues for the interpretation of an utterance. But, most of the evidence comes from studies investigating the influence of one source of information. In the following chapter, we thus review existing findings on the influence of more than one source of information on situated language comprehension.

3.4 VISUAL AND LINGUISTIC CUES

In sections 3.1, 3.2, and 3.3, we outlined how different cues such as morpho-syntactic, semantic, supra-segmental, or visual cues can rapidly influence language comprehension. Among others, we have seen that linguistic cues (syntactic, grammatical, and pragmatic) and visual cues (depicted action events, contrast between objects, referential context) can help with the disambiguation of the linguistic input. The results of the studies that we reviewed suggest that these cues elicit the anticipation of a target character, referent, or object. However, the studies mostly focused on one single cue. This leaves us with a question: What happens if more than one cue that directs people's attention towards a target is available? Is one cue stronger than the other and thus used exclusively or do two cues together enhance anticipation?

In an ERP study, Knoeferle, Habets, Crocker, and Münte (2007) examined whether visual cues elicit immediate reanalysis of syntactic structures. In two experiments, participants heard German SVO and OVS sentences which were either unambiguously or ambiguously case-marked (e.g., SVO: *Die Prinzessin/ Der Musiker malt offensichtlich den Fechter* - 'The princess$_{\text{amb.}}$/ The musician$_{\text{unamb.}}$ paints apparently the fencer $_{\text{object}}$)'; OVS: *Die Prinzessin Den Musiker wäscht offensichtlich der Pirat* - 'The princess$_{\text{amb.}}$/ The musician$_{\text{unamb.}}$ washes apparently the pirate$_{\text{subject}}$'). In Experiment 1, the sentences were accompanied by a visual scene containing three characters (musician/princess in the middle, pirate on one side, fencer on the other). The musician/princess performed a painting action (holding a brush in one hand and a colour selection chart in the other hand) and the pirate a washing action (holding a bucket in one hand and a sponge in the other).

In Experiment 2, the sentences were not accompanied by a visual display. Thus, this design allowed for a direct comparison between linguistic input only (case-marked NP2) and linguistic and visual input (case-marked NP2 and depicted action events). The results showed a P600 at the verb for ambiguous SVO and OVS sentences when sentences were accompanied by a scene. When the scene was absent, the P600 occurred later, namely at the NP2. For unambiguous sentences, no P600 effects were observed in both scene or no-scene conditions. Knoeferle et al. (2007) argued that linguistic (case-marking) and visual (depicted actions) cues contribute to syntactic disambiguation, although at different points in time.

Although linguistic and visual cues can both influence language processing, it is yet unclear whether there is a difference in the time course and mechanisms that influence the effects of the visual context. Knoeferle et al. (2014) investigated this issue using two crucially different picture-sentence mismatches: Verb-action mismatches and thematic role relation mismatches. For the sentence *The gymnast punches the journalist*, four different visual scenes depicting a gymnast and a journalist were created: 1) The gymnast punches the journalist (full match), 2) the gymnast applauds the journalist (action mismatch), 3) the journalist punches the gymnast (role mismatch), and 4) the journalist applauds the gymnast (combined mismatch).

Participants' ERPs were monitored while they inspected each scene for at least 3000 ms and then read English subject-verb-object sentences. After they read the sentence, participants verified quickly whether the sentence matched the previously seen picture or not (via button press). Action mismatches elicited different ERP responses than role mismatches. While role mismatches elicited a larger negativity for the mismatch (vs. match) at the first noun (200-400 ms after noun onset), action mismatches occured in form of a greater negativity to mismatches (vs. matches) at the verb (300-500 ms after verb onset). Furthermore, reaction-time patterns indicated faster and more accurate responses to role mismatches than to action mismatches. On the basis of these findings, Knoeferle et al. (2014) concluded that picture-sentence processing features more than one cognitive/neural mechanism.

In another study, Staudte, Crocker, Heloir, and Kipp (2014) compared a virtual speaker's gaze with another visual cue, namely an arrow pointing at the target object. While participants listened to spoken utterances such as *Das Ei ist größer als der Quarder* ('The egg is taller than the box'), they watched video clips which depicted a virtual character behind a table upon which differently coloured and shaped objects were placed. Gaze cues and arrow cues aligned either congruently, reversely, or neutrally with speech cues. Reaction-time results revealed lower response accuracy for gaze cues than for arrow cues. Crucially, eye-gaze results revealed an almost identical behaviour for the two cues: Mean first fixations to N2 objects were launched shortly after the onset of *box* (within 92 ms for the arrow cues and 182 ms for the gaze cues). The authors thus argued that the two cues facilitated comprehension in a similar manner (no quantitative differences between the two). However, this was only the case when the gaze cue was as precise as the arrow cue.

Kreysa et al. (2014) investigated whether two cues (speaker gaze and depicted action events) influence language comprehension in an additive or interactive manner. Participants listened to SVO sentences (e.g., *Der Kellner beglückwünscht den Millionär am Nachmittag* - 'The waiter congratulates the millionaire in the afternoon'). They used videos, which showed (vs. did not show) a human speaker inspecting a visual scene on a computer screen. This scene included three characters with an agent character in the middle of the scene, a patient character on one side of the agent, and a distractor character on the other side of the agent. In between the agent and the patient character, a verb-related depicted action appeared (vs. did not appear) after the onset of the verb. The human speaker first inspected the whole scene, looking at each character. At sentence onset, the speaker looked at the agent character and shifted the gaze from the agent to the patient after verb onset. Overall, conditions included no cue, one cue (speaker gaze or depicted action event), and two cues (speaker gaze and depicted action event).

The results revealed that, in the NP2 word region, participants used the gaze-cue and 200 ms later also the action-cue to fixate the patient character. The combined gaze and action cue did not contribute to more fixations on the patient character than the gaze only cue but to more fixations on the patient character than the action only cue. Kreysa et al. (2014) argued that one cue is enough to facilitate more looks to the patient character. Two cues together did not lead to a greater facilitation of looks to the patient character. Importantly, the two single cues differed in terms of how they linked to the target character. For the gaze-cue, participants simply followed the speaker gaze which led them to the target character. For the action-cue, participants had to first recognise that a) the verb relates to the depicted action and then b) the semantic relationship between the agent performing the action upon the patient.

Another study containing two different cues (emotional prime face, depicted action events) was conducted by Münster, Carminati, and Knoeferle (2015). Participants listened to positively emotionally valenced German OVS sentences (e.g., *Den Marienkäfer kitzelt vergnügt der Kater* - 'The ladybug (patient) tickles happily the cat (agent)'). Scenes contained three clipart animal characters (the patient/ladybug in the middle, the smiling agent/cat on one side, and a distractor character on the other side). The prime face was either positive (a smiley, a natural smiling face) or incongruent (a star, a natural sad face) with the sentence. Actions were either depicted (the agent performed a verb related action on the patient, the distractor performed an unrelated action on the patient) or not depicted. Participants anticipated the agent (vs. distractor) as early as the verb region of the sentence when actions were depicted (vs. when they were not). The natural smiling face only influenced looks to the target character when the action was also present (vs. absent). Münster et al. (2015) argued that the emotional prime face may only have additional beneficial effects on anticipatory eye-movements: It supports the depicted action events. Crucially, they discussed the processing differences of the two types of cues. Depicted action events are mediated by the verb whereas the link between an emotional prime face, a smiling target character, and a positively valenced sentence is more indirect.

The results of the studies by Knoeferle et al. (2007), Knoeferle et al. (2014), Kreysa et al. (2014), Münster et al. (2015), and Staudte et al. (2014) described the effects of two different cues that point towards the same target on language processing. Knoeferle et al. (2007) argued that non-linguistic and linguistic cues work equally well (for syntactic reanalysis) but the processes involved may vary (see Knoeferle et al., 2014 for a discussion). Staudte et al. (2014) provided evidence for similar effects on language comprehension for an arrow and a virtual speakers' gaze when they are equally precise. Furthermore, Kreysa et al. (2014) pointed out that target anticipation is faster for easier-to-process non-linguistic cues (speaker gaze) than for other non-linguistic cues (depicted action events). Adding both cues, however, did not have beneficial effects on anticipatory eye-movements. Münster et al. (2015), however, provided evidence for combinatorial effects of depicted actions and an emotional prime face. Overall, it seems that different cues can elicit similar effects, although not additive, and some cues do not elicit the same effects as others.

3.5 SUMMARY

Taken together, chapters 2 and 3 provided an overview of how adults process language in real-time. Specifically, previous research has demonstrated the interplay of linguistic and visual information during utterance comprehension. First of all, listeners relate spoken words to semantically or phonologically related objects or referents in the visual display (Allopenna et al., 1998; Cooper, 1974). Non-linguistic information such as referential context, or contrast between objects rapidly influences the interpretation of an utterance (Sedivy et al., 1999; Spivey et al., 2002; Tanenhaus et al., 1995). Third, listeners can anticipate upcoming linguistic input based on a verb's selectional restrictions, plausibility, and verb tense, and the affordances of objects (Altmann & Kamide, 1999, 2007; Chambers et al., 2004; Kamide, Altmann, & Haywood, 2003). Moreover, comprehenders rapidly integrate syntactic, supra-segmental, visual information as a cue to information structure (Kamide, Scheepers, & Altmann, 2003; Matzke et al., 2002; Snedeker & Trueswell, 2003; Weber et al., 2006). Finally, Knoeferle et al. (2005) provided evidence for the influence of depicted action events on incremental thematic role assignment.

In summary, evidence provided by the reviewed literature suggests that visual attention is closely time-locked with utterance comprehension. Linguistic information and world-knowldge, as well as visual information, can rapidly influence a range of comprehension processes: The interpretation, syntactic disambiguation, or thematic role assignment of spoken utterances. The utterance can further guide attention towards objects in the scene, often before a word is mentioned. Thus, the existing findings underlined that language processing is incremental. However, there seems to be a tension between semantic processing and pragmatic inferencing. Crucially, most of the evidence comes from studies with young adults. The extent to which the previously described findings generalise to other age groups, remains an open question at this point. The following chapter thus provides an overview of studies on child language comprehension.

4

CHILDREN VS ADULTS

In recent years, there has been an increasing number of studies on child language comprehension. From studies on adult language comprehension we know that language can guide (visual) attention and non-linguistic information such as referential context, contrast between objects, or depicted action events can inform incremental syntactic structuring, thematic role assignment, or semantic interpretation (e.g., Altmann & Kamide, 1999; Chambers et al., 2004; Knoeferle et al., 2005; Tanenhaus et al., 1995). Children acquire their native language throughout their first years of life. Consequently, not all aspects of their language are present from the beginning. From the age of twelve to eighteen months, children start to produce single words (e.g., *ball, mum, dad*), followed by two word sentences (e.g., *play ball, mum cookie*) soon after (O'Grady & Cho, 2001; Rothweiler, 2015; Schulz, 2007). However, whether children process language similarly to adults has been a controversial debate in the literature. Data from several studies suggests similarities between child and adult language comprehension (e.g., Gertner et al., 2006; Nation et al., 2003). Others argued for a developmental difference in child versus adult language comprehension (e.g., Huang & Snedeker, 2009b; Trueswell et al., 1999).

However, many studies used offline measures (e.g., an act-out or video-pointing task) which can only account for the outcome of language processing but not possible real-time processing differences or similarities. Other studies have focused on children only or used different materials for children versus adults. They thus failed to directly compare children and adults. Knoeferle (2015) argued that only a direct comparison between the two age groups can eliminate effects of stimulus variation on the time course or manner of processing. Thus, the debate of whether children process language similar to adults, leaves us with a puzzle: How do children actually compare to adults? How can we test for differences or similarities during real-time language comprehension? This chapter provides a review of literature on child language comprehension. We point out that some aspects of child versus adult language comprehension are similar whereas others demonstrate fundamental developmental differences during real-time language comprehension between children and adults (e.g., the use of non-linguistic context/pragmatic inferences).

4.1 SIMILARITIES

To begin with, vocabulary knowledge in children at the age of two can vary a lot (from 50 to 500 words). For some children, word learning is a very gradual process, whereas for others, it seems rather discontinuous (Rothweiler, 2015). Language acquisition is a process that takes time. However, certain aspects of child language processing resemble that of adult language processing. In this section, we discuss how language guides visual attention in children in comparison to adults.

4.1.1 *Anticipation*

Fernald, Pinto, Swingley, Weinbergy, and McRoberts (1998) explored whether children at the age of 15, 18, and 24 months can relate familiar spoken words to a visual referent. Children saw two familiar objects on two adjacent computer screens (e.g., a ball and a shoe, or a dog and a baby) and responded to questions (e.g., *Where's the baby?*). The results revealed that children rapidly fixated the visual referent of the familiar word. 24-month-olds already shifted their gaze to the target object before the end of the word. By contrast, 15-month-old children only shifted their gaze after the offset of the word. Existing findings on adult language processing have shown more fixations to a *beetle* - within a few hundred milliseconds upon hearing a word with a similar onset (e.g., a *beaker*) - than to unrelated objects (Allopenna et al., 1998).

Children have further been found to shift their (visual) attention towards a known object that shares colour features with the object mentioned in the linguistic input (Johnson & Huettig, 2011). 36-month-olds responded to questions such as *Can you find the frog?* while inspecting a scene in which a) the target was present, b) a colour matching object was present, or c) an unrelated object was presented adjacent to another object without typical colour features (e.g., a cup). Participants fixated the target significantly more than the other object in the target trials (a) and in the colour-matching trials (b) but not in the unrelated trials (c). This fixation pattern indicated that children are able to rapidly integrate the representations of objects attributes (i.e., their colour) during real-time language comprehension.

In a later study, young children shifted their attention towards lexically related objects, even though they did not produce words for this referent yet (Johnson et al., 2011). In an eye-tracking study, two-year-old children listened to questions such as *Can you find the banana?* while inspecting a scene that contained either a) a banana (target) and a red flower, b) a yellow shoe (colour or semantically related distractor) and a red flower, or c) a pink shoe (unrelated distractor) and a red flower. In a second part of the experiment, children were instructed to *Look at the yellow one* while inspecting smiley faces which only differed in colour. The eye-movement results revealed that children directed their gaze to target objects (banana) and related objects (yellow shoe) upon hearing *Can you find the banana?* but not upon hearing *Look at the yellow one*. Crucially, the data suggests that children's visual attention can be mediated by their perceptual-conceptual knowledge (i.e., bananas are yellow). Taken together, the two studies reviewed here pointed out that very young children are able to integrate lexical information in combination with referents in the visual world during language comprehension.

Crucially, children cannot only relate spoken words to objects in the visual display like adults (Altmann & Kamide, 1999), they can also rapidly anticipate upcoming target objects (Mani & Huettig, 2012; Nation et al., 2003). Ten- to eleven-year-old children listened to sentences in which the verb was restrictive (e.g., *Jane watched her mother eat the cake*) or neutral (e.g., *Jane watched her mother choose the cake*) with regards to the target object (a cake) in the visual scene. In addition to the target object, the visual scene contained three distractor objects categorised as food, animals, furniture, or clothes. After having heard the verb, participants anticipated the target object (the cake) more than the distractor objects. Furthermore, eye-gaze towards target objects was shifted earlier in the supportive condition compared to the neutral condition. Children seem to be aware of the verb's restrictions and can readily anticipate an upcoming referent in the visual world (Nation et al., 2003).

In a later study, Mani and Huettig (2012) suggested that even younger children can anticipate upcoming targets. Children at the age of two were presented with sentences containing either a semantically constraining verb (e.g., *The boy eats the big cake*) or a neutral verb (e.g., *The boy sees the big cake*) while inspecting two adjacent images on a computer screen (e.g., a bird and a cake). In sentences that restricted the target noun, two-year-olds anticipated the target object (vs. distractor) more than in sentences that did not restrict the target noun. However, this was only the case for skilled (vs. unskilled) language producers. Overall, this behaviour resembles the findings on adult language processing. Bear in mind that adults' visual attention was directed towards an object, on the basis of the verb's restrictions, before the target word was mentioned (Altmann & Kamide, 1999).

4.1.2 *Grammatical Cues*

Furthermore, children are able to rapidly recruit grammatical cues such as gender marking during real-time language comprehension (Arnold et al., 2007; Lew-Williams & Fernald, 2007). In Lew-Williams and Fernald (2007), two- to three-year-old Spanish-learning children listened to a sentence followed by a question (e.g., *Encuentra la pelota. ¿La vez?* - 'Find the ball. Do you see it?'; *Encuentra la galleta. ¿Te gusta?* - 'Find the cookie. Do you like it?') while inspecting two adjacent computer screens displaying two objects. The two objects either had the same gender (same-gender trials) or different genders (different-gender trials). The results revealed that children inspected the target object earlier in different-gender trials than in same-gender trials. The authors claimed that children rapidly recruit gender marking for the interpretation of noun-phrases.

Arnold et al. (2007) found similar results for three- to five-year-old English-learning children. Participants listened to a puppet telling a short story in which the characters of the experiment (identified as clearly male or female in the familiarisation of the trial) were introduced and performed a reciprocal action. All the characters were placed on a table. Children then heard a sentence in which one of the characters uttered a wish (e.g., *Puppy is having lunch with Froggy. He wants some milk*) and responded to questions about the target character (e.g., *Can you show me who wants the milk in Elmo's story?*). Across all ages, children used the gender cue for correct pronoun interpretation, however, the order-of-mention cue only when the target object could also be identified via the gender cue. The authors saw this as evidence for a developing sensitivity of the order-of-mention cue. However, in isolation, the order-of-mention cue did not support children's pronoun interpretation. Children rapidly integrated gender marking to identify an object in the visual world. Similarly, grammatical knowledge (gender marking) mediated visual attention in French speaking adults. Participants were instructed to *Cliquez sur le bouton* ('Click on the button$_{\text{masculine}}$') whilst looking at four objects. One of the objects was similar in word onset (e.g., bottle - *bouteille* - feminine). When gender marking was neutral (plural form for both - *les*), adults inspected the related object (bottle) more than the distractor objects. When the two objects were clearly gender marked (*le* vs. *la*), no such preference to look at the related object was observed (Dahan et al., 2002).

4.1.3 *Syntactic Cues*

Children's visual attention is not only influenced by their grammatical knowledge but also by their syntactic knowledge. Gertner et al. (2006) investigated whether children, much like adults, are able to integrate syntactic cues (e.g., word order) during language comprehension. In English, transitive sentences are typically structured subject-verb-object. Gertner et al. (2006) conducted experiments with children of different age groups (21- and 25-month-olds). Character identification and practice trials preceded the test trials. While participants listened five times to the test sentence which included a novel verb[1] (e.g., *Hey look! The duck's gonna gorp the bunny! The duck is gorping the bunny! The duck is gorping the bunny! See? The duck is gorping the bunny! Find gorping. Find gorping!*), they watched two video clips in which a bunny and a duck appeared in reversed agent and patient roles and performed novel actions. The results for both age groups revealed that children inspected the target video (the video in which the subject was the agent and the object the patient) longer than the reversed role video after having heard the test sentence once. The authors concluded that children are able to use (English) word order to correctly interpret sentences with novel verbs. They thus argued for an early abstraction account (e.g., Fisher & Gleitman, 2002; Wexler, 1999; see Gertner et al., 2006 for a discussion) and against a lexical account[2] because the meaning of the verb was unknown but children used their (abstract) knowledge about word order to assign thematic roles.

In a similar study, Dittmar, Abbot-Smith, Lieven, and Tomasello (2008b) investigated whether the results reported by Gertner et al. (2006) can be replicated with 21-month-old German children and whether the practice phase might have had an influence on the results. The materials were similar to those by Gertner et al. (2006), however, a no-practice (vs. practice) condition was added to the design. In the practice trials, children listened to transitive sentences with familiar words (e.g., *The frog is washing the monkey*). In the

1 Novel verbs are invented verbs to which a language specific features can be applied (gorp - gorping - gorps- gorped). They are used in studies with children to avoid lexical biases or semantic restrictions.

2 Lexical accounts assume that children interpret thematic roles via verb-specific knowledge (gradually via verbs that have similar meanings in different contexts; e.g., Cameron-Faulkner, Lieven, & Tomasello, 2003; Lieven, Pine, & Baldwin, 1997; see Gertner et al., 2006 for a discussion).

no-practice trials, the familiar verbs occurred only in citation form (e.g., *This is called washing*). Children in the no-practice condition did not perform above chance in looks at the target video whereas children in the practice condition performed above chance. The authors argued that 21-month-old German children are also sensitive to syntactic cues. It seems, however, that very young children do require some sort of practice to facilitate above chance performance.

Overall, child language comprehension seems to resemble adult language comprehension in some ways. Children, like adults, shift their (visual) attention towards objects in the visual world even before the end of the target word. They readily anticipate upcoming thematic roles and objects on the basis of a verb's restrictions. Furthermore, grammatical and syntactic cues guide their (visual) attention. Although the literature reviewed thus far provided some experimental evidence for similarities between children and adults, we have to keep in mind that no direct comparison between children and adults was provided (rather a comparison to other studies) and the authors used different paradigms (Preferential Looking Paradigm[3] vs. Eye-tracking and the visual-world paradigm). Further experimental evidence suggests that there are substantial differences in language comprehension between children and adults.

4.2 DIFFERENCES

In the preceding section (section 4.1), we reviewed studies which provided evidence for similarities between child and adult language comprehension. This section focuses on studies which suggest processing differences between the two age groups. Although language acquisition is a rapid process, some aspects of language processing may not be present from the beginning. In this section, the focus is placed on different cues (case-marking and prosody) which revealed processing differences between children and adults. We further discuss how children, unlike adults, failed to use visual context in some but not all instances and highlight developmental differences in visual attention. Nevertheless, it seems that some effects of visual context on language comprehension are present from a very early age.

3 The Preferential Looking Paradigm (Fernald, Perfors, & Marchman, 2006): Two images are shown and a camera records which of the two images the child looks at on a frame by frame basis. This measure thus allows to study online language comprehension. But the rating of where the child looks is determined by an unbiased rater after the recording.

4.2.1 *Case-Marking*

The acquisition of case-marking is a process that continues to develop over time. A long line of research investigated German children's grammar acquisition between the ages of four and nine (e.g., Kany & Schöler, 2007; Motsch & Becker, 2014; Motsch & Rietz, 2016; Ulrich, Penke, Berg, Lüdtke, & Motsch, 2016. One of the main goals of this research was to generate data from typically developing children in order to create a reference frame of comparison for children suffering from language impairments (Motsch & Becker, 2014).

Some researchers argued that the acquisition of case-marking follows a sequential order: nominative - accusative- dative (e.g., Clahsen, 1984; Milles, 1985; Tracy, 1986). More recent studies, however, suggest that the acquisition of case does not follow the previously mentioned order. The use of nominative, accusative, and dative case-marking was observed simultaneously in children aged four (Szagun, 2004; Ulrich et al., 2016). The age at which children have a fully developed case-marking system for language production and comprehension is yet highly debated. Result from production and comprehension studies suggest that children's case-marking is underdeveloped until the age of seven (e.g., Dittmar et al., 2008a; Kany & Schöler, 2007; Maiworm, 2008; Motsch & Rietz, 2016; Popella, 2005; Schipke et al., 2011). Other researchers, however, provided evidence for an adult-like use of (nominative and accusative) case-marking during language production and comprehension in four- to six-year-old children (e.g., Brandt, Lieven, & Tomasello, 2016; Eisenbeiss, Bartke, & Clahsen, 2006; Özge, Küntay, & Snedeker, 2019; Özge et al., 2016; Szagun, 2004).

Although children can use syntactic cues such as word order, they struggled to use case-marking for correct thematic role assignment (Dittmar et al., 2008a). In an act-out task (Experiment 1: 2.7- and 4.10-year-olds) and a video-pointing task (Experiment 2: 2.7-, 4.10-, and 7.3-year-olds). Dittmar et al. (2008a) investigated German children's interpretation of transitive sentences. During the experiments, children listened to sentences containing novel verbs. The sentences were unambiguously case-marked SVO sentences (e.g., *Der Hund wieft den Löwen* - 'The (nominative) dog is weefing the (accusative) lion'), ambiguously case-marked SVO sentences (e.g., *Die Katze wieft die Ziege* - 'The cat is weefing the goat'), unambiguously case-marked OVS sentences (e.g., *Den Bären wieft der Tiger* - 'The (accusative) bear is weefing the (nominative) tiger'), and unambiguously case-marked SVO sentences with a familiar verb (e.g., *Der Hund schubst den Tiger.* - 'The (nominative) dog is pushing the (accusative) tiger'). In the act-out task, children at the age of two only correctly interpreted SVO sentences with familiar verbs whereas four-year-olds also interpreted SVO novel verb sentences correctly. However, for OVS sentences correct responses were at chance level. Dittmar and colleagues (2008a) proposed that children at the age of four are able to use the syntactic cue of word order (SVO word order) but can not yet differentiate case-marking from word order. The results from the video-pointing task, which was supposedly cognitively less demanding, revealed similar results for two- and four-year-olds. Only at the age of seven were children able to correctly interpret 'who does what to whom' in clearly case-marked OVS sentences.

Further evidence supporting the findings of young children's difficulty of using case-marking as a cue for thematic role assignment in German SVO and OVS sentences, was recently provided in an ERP study by Schipke et al. (2011). Using a violation paradigm, three- to six-year-olds listened to sentences that were correctly case-marked (e.g., SVO: *Der Tiger küsst den Frosch* - 'The (nominative) tiger kisses the (accusative) frog'; OVS: *Den Frosch küsst der Tiger* - 'The (accusative) frog kisses the (nominative) tiger'), or that contained incorrect case-marking - double nominative case-marking for SVO sentences (e.g., *Der Tiger küsst der Frosch* - 'The (nominative) tiger kisses the (nominative) frog') and double accusative case-marking for OVS sentences (e.g., *Den Frosch küsst den Tiger* - 'The (accusative) frog kisses the (accusative) Tiger'). Unlike adults who showed biphasic N400 and P600 responses for both violation types - representing thematic semantic and syntactic processes - children of all age groups only showed adult-like responses in double nominative constructions. In double-accusative constructions, responses were different from adults. Schipke et al. (2011) interpreted these findings as an indicator for the development of understanding accusative case after the age of six.

However, results from a study similar to Dittmar et al. (2008a) suggest that children at the age of six are able to use case-marking. Brandt et al. (2016) investigated German children's (three, four, and six years) use of case-marking and word order. Crucially, the stimuli contained not only simple transitive subject-object and object-subject but also transitive relative clauses. Similar to Dittmar et al. (2008a), in Experiment 2, children listened to unambiguously case-marked simple transitives and transitive relative clauses containing either lexical NPs or demonstrative pronouns (e.g., transitive SO: *Der Hase schubst jetzt den Vogel* - 'the-NOM rabbit pushes now the-ACC bird', relative SO: *Der*

Hase, der den Vogel schubst - 'the-NOM rabbit who NOM the-ACC bird pushes', transitive Subject Pronoun: *Der Hase schubst den jetzt mal* - 'the-NOM rabbit pushes him now', relative Subject Pronoun: *Der Hase, der den jetzt schubst.* - 'the-NOM rabbit who-NOM him now pushes', transitive OS: *Den Vogel schubst jetzt der Hase* - 'the-ACC bird pushes now the-NOM rabbit', relative OS: *Der Vogel, den der Hase schubst* - 'the-NOM bird who the-ACC the-NOM rabbit pushes', transitive Object Pronoun: *Den Vogel schubst der jetzt mal* - 'the-ACC bird pushes he now', relative Object Pronoun: *Der Vogel, den der jetzt schubst* - 'the-NOM bird who-ACC he now pushes'). During the experiment, children saw two video-clips simultaneously with reversed agent/patient order. The video-clips were followed by static images of the scenes. Participants were instructed to point to the correct image (e.g., *Zeig mir mal das Bild: der Hase schubst jetzt den Vogel* - 'show me the picture: the-NOM rabbit pushes now the-ACC bird'). The results suggest that six-year-olds, unlike three- to four-year-olds, were able to use case-marking in simple transitive sentence with SO and OS ordering but not in transitive relative clauses. The authors, thus, argued that construction specificity may influence children's use of case-marking.

Evidence by Özge et al. (2016) suggests that German children at the age of four to five are able to rapidly recruit case-marking in unambiguously case-marked, verb-final SOV and OSV sentences. In an eye-tracking study, participants heard sentences (SOV: *Der Hase wird im nächsten Moment den Kohl aufspüren* - 'The rabbit will shortly find the cabbage'; OSV: *Den Hasen wird im nächsten Moment der Fuchs aufspüren* - 'The fox will shortly find the rabbit') while inspecting a visual scene. The visual scenes contained a referent for the first noun-phrase (*a rabbit*), a plausible agent (a *fox*) and a plausible patient (a *cabbage*). After the sentence a picture appeared on the screen which depicted an event that matched or did not match the sentence (e.g., a fox finding a rabbit as matching the OSV sentence, or a rabbit finding a cabbage as mismatching the OSV sentence). Children were instructed to tell the experimenter whether the previously heard sentence matched the picture. The results revealed a greater preference to look at the agent in OSV sentences (vs. SOV sentences). Thus, children used case-marking to correctly assign thematic roles, even in non-canonical sentence structures (OSV). The authors concluded that the findings are in line with accounts claiming that grammatical abstraction occurs at an early age. Therefore, the results are in line with early and not late abstraction accounts[4] because children were able to abstract that the agent is performing an action and the patient is being acted upon.

Özge et al. (2019) further investigated Turkish four- to five-year-old children's ability to use case-marking for thematic role assignment. Turkish, similar to German, has a relatively flexible word order with a preference for an agent first interpretation in ambiguous sentences. Thus, case-marking can also influence thematic role assignment. The materials were similar to those in Özge et al. (2016). The results suggest that Turkish children between the ages of four and five, like their German counterparts, are able to use case-marking for thematic role assignment during online language comprehension.

The results of the studies on children's use of case-marking during language comprehension reported here are somewhat contradictory. The findings by Dittmar et al. (2008a) and by Schipke et al. (2011) suggest that German children cannot use case-marking for thematic role assignment until the age of seven and do not process case-marking in an adult-like manner in non-canonical German OVS sentences. But, when case-marking correlates with the preferred word order (SVO), they correctly assigned thematic roles. It thus seems that the understanding of accusative case changes at some point after the age of six. Crucially, the findings by Özge et al. (2016), Brandt et al. (2016), and Özge et al. (2019) provided a completely different view: Children between the ages of four and six were already able to use case-marking for correct thematic role assignment. However, the design of the studies by Özge (2016, 2019) differed from the designs by Dittmar et al. (2008a), and Schipke et al. (2011). Özge et al. (2016) and Özge et al. (2019) provided visual scenes containing plausible agents and patients for the different sentence structures whereas in Dittmar et al. (2008a), children saw two video clips with reversed agent and patient roles and the characters performed novel actions. In Schipke et al. (2011) children listened to violating structures but no visual display supported the linguistic input. It seems a supportive context is necessary for young children to rapidly exploit case-marking for the interpretation of 'who does what to whom'.

The previously described findings, nevertheless, confront us with a puzzle. Why do German children not (or at least only in supportive visual contexts) use case-marking during comprehension but seem to be able to

4 In late abstraction accounts such an abstraction cannot be made. The interpretation is based on the verb (e.g., Ambridge & Lieven, 2011; Tomasello, 1992; see Özge et al., 2016 for a discussion).

produce case-marking correctly in some instances? Existing findings on children's production of case-marking revealed that German children produce case-marking largely error free. Although these studies primarily focussed on the differences between typically developing children and learning-impaired children or children with cochlear implants, the results from the typically-developing children suggest that children readily produce accusative and nominative case (Eisenbeiss et al., 2006; Szagun, 2004). Other studies, however, suggest that German children continue to produce case-marking errors until the age of seven (e.g., Kany & Schöler, 2007; Maiworm, 2008; Motsch & Rietz, 2016; Popella, 2005). Overall, the picture that emerges here seems manifold. In some instances German children around the age of five produced case-marking largely error free, whereas in others errors continue to occur until the age of 7 (Eisenbeiss et al., 2006; Szagun, 2004 but Kany & Schöler, 2007; Maiworm, 2008; Motsch & Rietz, 2016; Popella, 2005). Four- to six-year old children also used case-marking for thematic role assignment in some but not all comprehension studies (Brandt et al., 2016; Özge et al., 2019, 2016 but Dittmar et al., 2008a; Schipke et al., 2011).

4.2.2 *Prosody*

When case-marking is ambiguous, other cues can help to distinguish 'who does what to whom'. As a reminder, adults used prosody as a cue for thematic role assignment or reference resolution. Although children are sensitive to prosody from a very early age, some but not all aspects of integrating such cues during language comprehension are similar in children as compared to adults. Five-month-old children showed a sensitivity to prosody when responding differently to positive or negative stimuli. Using a looking-time and facial-affect measure[5], children listened to approval or prohibition vocalisations in infant- and adult-directed English, infant-directed nonsense English/languages other than their first language (German, Italian, Japanese. The results revealed that children had positive reactions to approvals and negative reactions to prohibitions in infant-directed speech in English, nonsense English, German, and Italian but not in Japanese. In adult-directed speech, infants showed no such reactions. The authors argued that young children can distinguish whether vocalisations are approving or prohibiting in infant-directed speech but not in adult-directed speech (Fernald, 1993).

5 In a looking-time measure the time which the infants spend on looking towards on side or the other is measured. Facial-effect measure takes record of the child's facial responses (e.g., a positive affect could be a smile)

Arnold (2008) investigated the influence of accent on children's reference comprehension. Following the design by Dahan et al. (2002), four- to five-year-old children listened to anaphoric and non-anaphoric instructions which were either accented or unaccented (e.g., anaphoric: *Put the bacon on the star - Now put the BACON/bacon on the square*, non-anaphoric: *Put the bacon on the star - Now put the BAGEL/bagel on the square*). One additional condition contained anaphoric pronominal instructions (e.g., *Put the bacon on the star - Now put it on the square*). The visual display contained four objects including two cohort competitors (e.g., bagel/bacon). Eye-movement results revealed more looks to the previously-mentioned object in the unaccented (vs. accented) condition. Children, unlike their adult counterparts in Dahan et al. (2002), did not prefer a non-anaphoric interpretation of accent. However, children behaved similar to the adult control group in this experiment. Arnold (2008) concluded that accentuation can influence children's reference resolution. It seems, however, that prosody can influence reference resolution similarly in some instances in adults and children but differently in others.

Crucially, German children at the age of five can use prosody to correctly interpret transitive sentences (Grünloh et al., 2011). In a first experiment, auditory stimuli were either unambiguously or ambiguously case-marked OVS sentences with or without a pitch accent on the first noun-phrase (e.g., *Den HUND/Hund wieft der Elefant* - 'The (accusative) dog is weefing the (nominative) elephant'; *Die KATZE/Katze mommelt die Ziege* - 'The (accusative/nominative) cat is mommeling the (accusative/nominative) goat'). Using a video-pointing task, the sentences were presented to children while they watched two video-clips which were then followed by static images of the scenes. The video-clips showed two animal characters performing novel-verb actions (e.g., for the verb *wiefen*, an animal was rocking another animal on something like a rocking chair). The only difference between the two videos was that agent and patient roles were reversed.

Children's correct responses were only above chance level for unambiguously case-marked OVS sentences with a contrastive intonation. For unambiguously case-marked OVS sentences without a contrastive intonation, performance was at chance level. In ambiguous OVS sentences, children's response accuracy was below chance level. Thus, case-marking and prosody together function as a stronger cue than case-marking alone (i.e., the combination of cues affects language comprehension more than the single cues). However, prosody alone did not evoke more correct responses. When case-marking was ambiguous, (SVO) word order influenced thematic role assignment. The authors suggested that the results are in line with the findings by Dittmar et al. (2008a) in which word order and case-marking supported the interpretation of transitive sentences. Different types of (redundant) cues can thus inform child language comprehension (Grünloh et al., 2011).

In a second experiment, Grünloh and colleagues (2011) investigated whether the results would look different if the stimuli were presented in a discourse context. They used the materials from Experiment 1 but with one critical difference: The sentences occurred in a context in which an incorrect description of the scene in SVO sentence structure preceded the experimental sentences (e.g., *Der Löwe wieft den Frosch - Nicht den Frosch wieft der Löwe, sondern den Hund wieft der Löwe* - 'The (nominative) lion is weefing the (accusative) frog - Not the (accusative) frog is weefing the lion but the (accusative) dog is weefing the (nominative) lion'). Children's responses were above chance level for both neutral and contrastive unambiguous OVS sentences. For ambiguously case-marked sentences children's responses were only at chance level in combination with a contrastive intonation. When intonation was neutral, correct responses were below chance level.

The findings of this study support the idea that several cues (prosody, case-marking, and discourse) in combination facilitate child language comprehension (i.e., their interpretation of transitive sentences). Overall, prosody helped children to overcome their strong SVO word order bias but only in combination with case-marking. When case-marking was ambiguous, they fell back on their most reliable cue: Word order (Grünloh et al., 2011). Similar findings were reported for Russian five to six-year-old children in a study by Sekerina and Trueswell (2012). Children used pitch accents to identify target objects only in sentences with a canonical word order. Interestingly, the adult control group in the Grünloh et al. (2011) study did not use prosodic cues like children: They used case-marking when unambiguous and word order when case-marking was ambiguous. These findings are also similar to the results by Arnold et al. (2007). Children only used the order-of-mention cue in addition to the gender cue. In isolation the order-of-mention cue had no beneficial effects whereas the gender cue did.

However, during online language comprehension, prosody seems to affect children's visual attention similarly to adults if they are given more time. Ito, Jincho, Minai, Yamane, and Mazuka (2012) examined the effects of prosody on children's and adults' reference resolution. Six-year-old Japanese children inspected scenes with multiple animals, three of which belonged to the same category but differed in colour. They listened to two sets of questions. The first question referred to one of the animals (e.g., translation: 'Where is the pink cat/rabbit?'). The second question either contained an animal of the same category (e.g., translation: 'Where is the green cat?') or an animal of a different category (e.g., translation: 'Where is the orange monkey?'). The colour adjective either carried a pitch accent or not. Eye-movement results suggested that children inspected the cat more when the adjective was accented (e.g., GREEN cat) in comparison to when it was deaccented (e.g., green cat) but with some delay in comparison to adults. When the second question contained an animal from a different category, adults were garden-pathed by the pitch accent on the adjective (looks to the rabbit) whereas children only showed a similar gaze behaviour when the interval between the two questions was longer. The authors suggested that six-year-old children can rapidly integrate prosodic cues, given more time to establish discourse.

Taken together, these studies demonstrated that, although children are sensitive to prosody form a very early age, the way in which they exploit prosody during language comprehension seems to differ from adults. Not only do they need more time, they sometimes interpret accent differently. Thus far, we have seen that some aspects of child and adult language comprehension differ (e.g., using morpho-syntactic and supra-sentential cues to correctly interpret spoken utterances). Evidence for real-time differences comes from studies examining the effects of visual context on child language processing.

4.2.3 *Visual Referential Context*

One of the much-debated questions has been whether children are able to rapidly exploit non-linguistic cues such as referential context (Kidd & Bavin, 2007; Meroni & Crain, 2003; Snedeker & Trueswell, 2004; Trueswell et al., 1999; Weighall, 2008). For adults, non-linguistic information (e.g. referential context, contrast between objects, or depicted action events) can inform language comprehension in real-time (Chambers et al., 2004; Knoeferle et al., 2005; Sedivy et al., 1999; Tanenhaus et al., 1995). A number of studies have reported that children, unlike adults, are not able to rapidly integrate visual referential context to resolve prepositional-phrase-attachment ambiguities (Snedeker & Trueswell, 2004; Trueswell et al., 1999) until the age of eight (Weighall, 2008) but some argued that children behave similarly to adults (Kidd & Bavin, 2007; Meroni & Crain, 2003).

The very first visual world eye-tracking study to highlight the differences between adults and children in the use of referential context was reported by Trueswell et al. (1999). Four- to five-year-old children listened to ambiguous (1a) and unambiguous (1b) instructions.

(1) a. Put the frog on the napkin in the box.
b. Put the frog that's on the napkin in the box.
(Trueswell et al., 1999)

The visual context contained either two or one possible referent(s) for the frog (one- or two-referent context). The set-up of the two-referent context contained two frogs, one on a napkin and one on the table, an empty napkin (incorrect destination), and an empty box (correct destination). The set-up of the one-referent context was similar to the set-up of the two-referent context but the frog sitting on the table was exchanged by a horse. The two-referent context supports a modifier interpretation (*the frog on the napkin*) and the one-referent a destination interpretation (*put the frog on the napkin*). Without any visual context, listeners prefer a verb-phrase attachment of *on the towel*. The results revealed more looks and more movement actions to the incorrect destination, in both the one- and the two-referent context. Five-year-olds did not revise their preference to interpret *on the towel* as the destination rather than the location of *the frog*, independent of the number of referents in the visual context.

Trueswell and colleagues (1999) suggested that children failed to revise their initial parsing preferences as a result of their limited processing capacities: The destination interpretation was preferred early and could not be revised later in the sentence. The authors further argued that another possible explanation may be that children generally prefer a verb-phrase attachment (destination interpretation), possibly because the verb *put* is used more often for a modifier interpretation in child directed speech. This explanation seems plausible since previous research also reported effects of statistical knowledge of words on child language processing (e.g., Mitchell & Cuetos, 1991; Saffran, 2003; Saffran, Aslin, & Newport, 1996; Wells, Christiansen, Race, Acheson, & MacDonald, 2009). Another possible reason for why children failed to exploit the visual referential context is that they struggled with pragmatic inferences. Arguably, children have to infer from the visual context (two frogs) that the prepositional-phrase *on the napkin* distinguishes which of the two frogs is the intended referent (see Knoeferle, 2015 for a discussion). Taken together, Trueswell et al. (1999) provided the first evidence for a child's incremental processing system which systematically differs from an adults'. Consequently, the results encouraged other researchers to investigate this issue further.

Meroni and Crain (2003), for example, investigated why children did not use referential context for structural disambiguation. They conducted an act-out study with a similar procedure as the one in Trueswell et al. (1999) but made two changes: 1) in the two-referent context, two frogs were placed on two differently coloured napkins (one on a red napkin and one on a blue napkin) and the set-up contained an additional empty red napkin (pragmatic-block condition), 2) Children were instructed to inspect the scene and then turn away and listen to instructions such as *Put the frog on the red napkin in the box*. Only after listening to the instructions were they allowed to inspect the scene again and perform an action (eyes-closed condition). The findings revealed that children performed correct actions in 92% of the trials. Based on their results, the authors argued that children failed to use the visual referential context in the study by Trueswell et al. (1999) because they made pragmatic inferences about the *frog* when hearing *Put the frog on the napkin...* (i.e., the frog that was not on the napkin). Meroni and Crain (2003) further suggested that the execution of actions in correspondence to linguistic input is difficult for children because they do not have enough time to revise their initial parsing preference. When they had enough time to revise these preferences (by looking away while listening to the linguistic input) their performance improved. The results thus support the idea of limited processing capacities in children. Given more time, they can overcome their initial parsing preferences.

In contrast, Weighall (2008) suggested that the findings by Meroni and Crain (2003) resulted from either the pragmatic-block or the eyes-closed condition. Since the authors did not provide a direct comparison between these two conditions, Weighall (2008) directly compared Trueswell et al.'s study (1999) with Meroni & Crain's study (2003) with four- to eleven-year-old children. Using an act-out task, this study implemented both, an eyes-closed condition, and an open-eye condition plus a pragmatic-block vs. no pragmatic-block condition in addition to the one- and two-referent context. In contrast to Meroni and Crain (2003), the findings of this experiment suggest that children at the age of five prefer a destination interpretation of ambiguous prepositional-phrases, regardless of the experimental set-up. Young children interpreted *on the napkin* more often as a destination than a modifier of *the frog* in the two-referent context compared to the one-referent context. By the age of eight, however, children reached adult-like performance in resolving PP-attachment ambiguities.

Snedeker and Trueswell (2004) examined whether it was a lexical bias that influenced the results in Trueswell et al. (1999). The verb *put* occurs more frequently with a modifier interpretation and perhaps this was the reason why children interpreted the prepositional-phrase *on the napkin* as the destination of *the frog* regardless of the number of referents in the visual context. Snedeker and Trueswell (2004) used a similar design to Trueswell et al. (1999) but chose verbs which biased different structural preferences. Child participants listened to instructions containing PP-attachment ambiguities (2a,2b,2c) while inspecting one- or two-referent contexts.

(2) a. Choose the cow with the stick. (Modifier Bias)
b. Feel the frog with the feather. (Equi Bias)
c. Tickle the pig with the fan. (Instrument Bias)
(Snedeker & Trueswell, 2004)

Although the eye-movement data revealed a marginal effect of context, the overall results were consistent with those by Trueswell et al. (1999): Children did not exploit the referential context for their final interpretation of the ambiguous sentence. They did, however, rely on the lexical bias of the verbs. In sentences with an instrument biased verb, five-year-olds performed actions on the target instrument. Modifier-biased sentences resulted in actions on the target animal. Snedeker and Trueswell (2004) argued that lexical biases, rather than VP-attachment preferences (Trueswell et al., 1999), inform ambiguity resolution. Adults, however, were not only influenced by the verb's lexical bias but also by the number of referents in the visual context. The findings suggest that more reliable cues such as lexical cues develop earlier than less reliable cues (Snedeker & Trueswell, 2004).

Evidence from another study replicated the effects of lexical bias on PP-attachment ambiguities but Kidd and Bavin (2007) suggested that children resolve syntactic ambiguities similar to adults. The authors investigated whether children rely on verb semantics when they encounter sentences with PP-attachment ambiguities. Participants listened to sentences with definite or indefinite NPs that were either NP-attached (e.g., *The girl looked at the/a boy with the/a basketball because she wanted to play*), or VP-attached (e.g., *The girl looked at the/a boy with much excitement because she wanted to play*). Verbs were either action verbs or psychological verbs/verbs of perception. The sentences were divided into separate parts. Via button press, the next part of the sentence was initiated. After having heard the whole sentence, seven- to nine-year-olds responded to post-sentence questions such as *Did the girl want to play?*, or *Did the boy have a basketball?*.

The data suggests that children, just like the adult control group, rely on verb semantics: Difference time scores between button presses were significantly longer in NP-attached sentences than in VP-attached sentences containing action verbs. Such a difference was not observed for sentences containing psychological predicates or verbs of perception. The authors concluded that child language parsing of PP-attachment ambiguities is similar to that of adults. The results of this study, although in a different paradigm than Snedeker and Trueswell (2004), supported the idea that verb bias plays a role in resolving PP-attachment ambiguities. The authors argued that a verb's statistical regularities affect language comprehension. Effects of referential context were, however, absent in both children and adults.

Choi and Trueswell (2010) proposed a different solution for children's inability to resolve syntactic ambiguities with the support of non-linguistic information. In a study similar in design to Snedeker et al.'s (2004), they tested five-year-old Korean and English children. Korean is a verb-final language and thus does not provide lexical biases until the end of a sentence. Similar to English children, Korean children failed to overcome their initial destination interpretation despite the disambiguating verb. The authors argued that children cannot rapidly exploit the visual referential context because their cognitive control abilities are not yet fully developed and not because of the verb's lexical bias. In both languages, English and Korean, children preferred early-arriving cues (VP-attachment preference) over late-arriving cues (referential context).

Another example for differences between child and adult language processing comes from a study conducted by Huang and Snedeker (2009b). The authors explored whether children's (in)ability to generate pragmatic inferences is responsible for the limited effects of referential context using scalar implicatures (SIs). During a preferential-looking study, five-year-old children were instructed to *Point to the girl that has some of the socks* after they heard a story which introduced the characters (a boy and a girl) and the distribution of the objects (socks and soccer-balls). In one condition, one girl had some of the socks and another girl had all of the soccer-balls (scalar implicature consistent trials - SI-consistent). In another condition, one girl had all of the socks and another girl had some of the soccer-balls (scalar implicature violating trials - SI-violating). Unlike adults, children inspected the target more in SI-violating trials than in SI-consistent trials. Across conditions, fixations on the girl with all of the soccer-balls were greater than fixations on the girl with some of the soccer-balls. The authors argued that children, unlike adults, cannot distinguish between the two SI-contexts and thus concluded that scalar quantifiers do not inform language comprehension for children at the age of five. Note that adults were able to make pragmatic inferences although these occurred after the semantic interpretation (Huang & Snedeker, 2009a). Thus, it seems plausible that one main difference between child and adult language processing lies in their (in)ability to generate pragmatic inferences.

In sum, visual context effects on language comprehension in children seem to be limited, at least in the case of PP-attachment ambiguities. Clear differences between the two age groups emerged in their use of visual referential context. It remains unclear, however, why these differences between children and adults exist. Evidence for child-adult processing differences comes from studies that suggest underdeveloped cognitive control abilities, an influence of lexical biases, and difficulties in computing pragmatic inferences online. Children may prefer early over late-arriving cues because their cognitive control abilities are not yet fully developed (Choi & Trueswell, 2010). Another factor which may have resulted in limited visual context effects, is the verbs' lexical bias or semantics such that children rely more on the verb than the visual context (Kidd & Bavin, 2007; Snedeker & Trueswell, 2004). Alternatively, children struggle to infer from the visual context that the prepositional-phrase can disambiguate reference (Trueswell et al., 1999). Adults computed pragmatic inferences online whereas children did not until the age of eight (Huang & Snedeker, 2009a, 2009b). It seems that between the ages of five and eight some development of the parser takes place at least in situations in which they have to overcome their initial parsing preferences to correctly interpret ambiguous sentences. In the following section, we discuss more evidence on developmental differences in children.

4.2.4 *Developmental Differences*

Existing evidence suggests a gradual development in the close coordination of language comprehension and visual attention. Fernald, Thorpe, and Marchman (2010) investigated how children interpret familiar adjectives during language comprehension. Adult listeners interpreted colour adjectives incrementally within a few hundred milliseconds and even before the following noun when the context provided additional information about target referents (Eberhard et al., 1995; Sedivy et al., 1999). 30-month- and 36-month-old English children heard questions (e.g., *'Can you find the blue car?'*) while inspecting scenes that either depicted different objects with the same colour (e.g., a blue car and a blue house - the adjective is uninformative), different objects with different colours (e.g., a blue car and a red house - the adjective is informative), or the same objects in different colours (e.g., a red and a blue car - the adjective is informative). The results suggest a developmental difference between 30- and 36-month-old children. 30-month-olds shifted their gaze in more than half of the trials to the incorrect referent when two objects were depicted that only differed in colour (a blue car and a red car) whereas 36-month-olds interpreted adjectives incrementally, resulting in gaze-shifts to the target object in both conditions. The authors thus argued that children's ability to interpret adjectives incrementally and rapidly is a gradual process which develops towards an adult-like performance over the third year of their lives.

However, visual attention is not only mediated differently depending on a comprehenders' age. Borovsky, Elman, and Fernald (2012) explored the relationship between vocabulary knowledge and incremental sentence comprehension. Eye-movements of three- to ten-year-old children were recorded while they listened to spoken utterances such as *The pirate hides the treasure* and inspected a visual display with four single images: A target image (treasure), an agent-related image (ship), an action-related image (bone), and an unrelated image (cat). Children were instructed to point to the picture that fits the sentence. Indeed, children with a larger vocabulary launched more fixations to the target objects than children with smaller vocabulary. Furthermore, older children and children with a larger vocabulary anticipated the target object earlier than younger children and children with a smaller vocabulary. Thus, Borovsky et al. (2012) suggested that a relationship between vocabulary knowledge and incremental sentence processing exists.

Weighall and Altmann (2011) also addressed the question of developmental differences on child language processing. They were particularly interested in how working memory capacities might influence the use of contextual information during language comprehension. During the experiment, six- to eight-year-old children heard center-embedded (e.g., *The bear that bumped the bear will hug the cow*) or right-branching (e.g., *The cow will hug the cat that bumped the bear*) sentences and responded to either a main clause or a relative clause question (e.g., main clause question for a center-embedded clause: *What* *sort of animal will hug the cow?*, or relative clause question: *What sort of animal bumped the bear?*). They saw an image that either depicted or did not depict the action described in the following sentence which was accompanied by the target image. Children with a higher and a lower verbal working memory span (ascertained via a Listening Span Task) used contextual information (previously depicted actions) to respond to the questions. However, only higher span children were able to use action contexts for the two different sentence structures. Lower span children used action contexts only when the question directly related to the action (relative clause questions). The authors suggested that pragmatically appropriate action contexts can facilitate comprehension for more complex sentence structures, even in low span children.

Taken together, it seems that a child's language processor experiences important developmental changes over the first years of life. The rapid interplay between language comprehension and visual attention seems to develop gradually (Fernald et al., 2010). Perhaps limited processing capacities are one explanation for why children process language differently to adults. However, more recent evidence suggests that, during real-time language comprehension, rapid visual context effects can also be similar in both age groups.

4.2.5 *Visual Context Effects*

Recent studies have reported rapid visual context effects that are similar in child and adult language processing. Richardson and Kirkham (2004) reported that children, much like adults, can keep track of moving objects in a visual display. In a visual world eye-tracking study, six-month-old children inspected a visual display containing two boxes. Each of the boxes contained an object (e.g., a toy cat and a toy dog) which was accompanied by a sound (*boing* or *bring*). The objects were presented to children one after the other such that the first display contained a toy cat in the right box accompanied by the sound *boing* and the second display contained a toy dog in the left box accompanied by an *bring* sound. The second box in the display was always a blank box. The objects then disappeared from the boxes and one of the previously-heard sounds was played.

When adults listened to statements which referred to one of two empty boxes in the visual display (a spinning cross indicated whether the statement referred to the left box or the right box), they inspected the correct location of the then-empty box upon hearing a related question, even when the two boxes moved around in the display. Eye-movement results for children indicated that they correctly identified the location of the previously-seen object upon hearing the related sound (i.e., they looked at the empty box on the left upon hearing *bring* and at the empty box on the right upon hearing *boing*). Thus, very young children, much like adults, can correctly identify a location of objects, even when these are no longer co-present (empty boxes).

Rapid visual context effects in children and adults also emerged in studies by Zhang and Knoeferle (2012) and Münster (2016). In a visual world eye-tracking study, Zhang and Knoeferle (2012) investigated children's (and adults') ability to integrate depicted action events for thematic role assignment. Participants listened to structurally unambiguous SVO (e.g., *Der Bär schubst sogleich den Stier* - 'The bear (subj) pushes immediately the bull (obj)') and OVS sentences (e.g., *Den Bär malt sogleich der Wurm* - 'The bear(obj) paints immediately the worm (subj)') while inspecting a scene depicting a bear, a worm, and a bull. Participants further responded to post-sentence comprehension questions about 'who does what to whom'. Actions referring to the verb were either depicted (e.g., a bear pushing a bull or a bear being painted by a worm) or not depicted. Children, much like adults, inspected the target character (the bull patient in SVO sentences and the worm agent in OVS sentences) more when actions were depicted compared to when they were not. However, adults started to inspect the correct target character earlier (upon hearing the verb) than children (upon hearing the adverb). Depicted action events further improved children's interpretation of the otherwise difficult non-canonical OVS sentences. The results suggest that non-linguistic visual context can inform language comprehension in both children and adults.

Recap that Münster (2016) observed similar results: Children were able to use depicted actions to correctly interpret non-canonical OVS sentences. It seems that visual context can influence child language processing similar to adult language processing but only if the context does not require pragmatic inferences. Nonetheless, the time course of when children integrate these visual cues differed from adults. Both, Zhang and Knoeferle (2012) and Münster (2016) reported a delay in the time course of eye-movements by one word region (adults verb region, children adverb region).

It seems that children are able to rapidly integrate some, but not all, non-linguistic information. When the visual context required children to compute pragmatic inferences, they failed to exploit the visual referential context and could not overcome their initial parsing preferences but when depicted actions were mediated by the verb and identified thematic role relations, five-year-olds overcame their initial parsing preference (SVO interpretation) and correctly interpreted unambiguous OVS sentences.

4.3 Summary

In sum, the literature reviewed in this chapter revealed that children, much like adults, are able to link the linguistic input to the visual world (Fernald et al., 1998; Johnson & Huettig, 2011; Johnson et al., 2011), even in an anticipatory manner (Mani & Huettig, 2012). Children's visual attention is further guided by syntactic and grammatical cues (Arnold et al., 2007; Dittmar et al., 2008b; Gertner et al., 2006; Lew-Williams & Fernald, 2007). However, we have also seen developmental differences between child and adult language processing (e.g., Fernald et al., 2010) which seem to be especially pronounced in the pragmatic domain (Snedeker & Trueswell, 2004; Trueswell et al., 1999).

We should keep in mind that the studies reviewed in this chapter used different measures to evaluate child language comprehension: The auditory moving window technique, act-out tasks (Dittmar et al., 2008a; Meroni & Crain, 2003; Weighall, 2008), video-pointing tasks (Brandt et al., 2016; Dittmar et al., 2008a; Grünloh et al., 2011), auditory sentence comprehension tasks (Weighall & Altmann, 2011), self-paced listening tasks (Kidd & Bavin, 2007), visual world eye-tracking studies (Borovsky et al., 2012; Nation et al., 2003), preferential looking studies (Dittmar et al., 2008b; Fernald, 1993; Fernald et al., 1998; Gertner et al., 2006; Huang & Snedeker, 2009a; Johnson & Huettig, 2011; Johnson et al., 2011; Lew-Williams & Fernald, 2007; Mani & Huettig, 2012; Snedeker & Trueswell, 2004), facial effects response measures (Fernald, 1993), or a combination of two (Arnold, 2008; Münster, 2016; Trueswell et al., 1999; Zhang & Knoeferle, 2012). Certainly, differences between these various types of measures and tasks exist. Act-out tasks, for example, demand sentence interpretation as well as the execution of an action. Dittmar et al. (2008a) argued that an act-out task has high demands on working memory and executive functions whereas a video-pointing task might not. Furthermore, during act-out tasks, children have to execute their action responses on the fly (Meroni & Crain, 2003). Eye-tracking studies on the other hand only require automatised mechanisms (i.e., eye-movements; Mishra, Olivers, & Huettig, 2013).

Although all the reviewed studies add to the understanding of child language comprehension, the differences between the measures used has to be taken into account, for example, when conducting follow-up studies (which might result in different findings if a different measure is used - offline studies can only account for the outcome of the comprehension process whereas online studies additionally illustrate the time course of processing). Although some studies provide evidence for differences between child and adult language processing, using different measures and materials leaves a couple of questions unanswered: Is the time course of language processing different in children as compared to adults? Do children struggle to use non-linguistic context because they cannot compute pragmatic inferences yet (as suggested by Snedeker & Trueswell, 2004; Trueswell et al., 1999)?

The insights from this chapter are especially interesting for accounts of visually situated language processing. Existing findings suggest that child and adult language processing are similar in some aspects (e.g., relating words to objects) but different in others (e.g., rapid visual context effects). In the next chapter, we provide an overview of such an account which models the coordinated interplay of visual attention and utterance comprehension.

5

THE COORDINATED INTERPLAY ACCOUNT

The reviewed literature on adult language comprehension has provided an overview of visual world eye-tracking studies which have shown aspects of how language is processed in real-time. During comprehension listeners can draw upon different sources of non-linguistic information for reference resolution, thematic role assignment, and structural disambiguation (Altmann & Kamide, 2007; Chambers et al., 2004; Huettig & Altmann, 2005; Kamide, Altmann, & Haywood, 2003; Knoeferle & Crocker, 2006; Knoeferle et al., 2005; Sedivy et al., 1999; Tanenhaus et al., 1995). What remains unclear up to this point, however, is how linguistic and non-linguistic information interact online. This issue has been addressed by Knoeferle and Crocker (2006) and Knoeferle and Crocker (2007). The account was motivated by evidence from visual world eye-tracking studies which suggest that non-linguistic visual information (visual context) interacts with incremental language processing.

5.1 THE CIA

Knoeferle and Crocker (2006) put forward an account of situated language processing which describes the close temporal coordination of utterance and scene processing in combination with other sources of information dubbed *the Coordinated Interplay Account (CIA)*. The CIA contains three stages of processing which are informationally and temporally dependent. In the first stage, *Sentence interpretation* (step i), comprehenders draw on incremental, anticipatory, and integrative processing mechanisms. Upon hearing a word ($word_i$), a current interpretation (int_i) of the word is constructed based on linguistic constraints of the word. Additionally, expectations (ant_i) based on linguistic and long-term knowledge are formed. In the second stage, *Utterance mediated attention* (step i′), the interpretation and expectations of the word (from step i) can mediate attention towards referents in the visual scene or representations of these referents in working memory (referential or anticipatory search). The information from the scene ($scene_{i'}$) is merged with representations of the scene ($scene_{i''-1}$). Objects and events that are no longer present in the scene decay. In the third stage, *scene integration* (step i"), the interpretation of the word$_i$ (int$_i$) and the expectations (ant$_i$) are being reconciled with the scene information (scene$_{i'}$) by co-indexing and revision of the scene events. The updated interpretation can then further influence the interpretations and expectations of the following word$_{i+1}$. Each stage accommodates a working memory component for the interpretation, the expectations, and the representations of the scene (Knoeferle & Crocker, 2007; Knoeferle et al., 2014).

In sum, the CIA proposes that visual context affects incremental language comprehension. Some aspects were, however, underspecified in the 2007 version of the CIA. What remained unclear was whether there are differences in visual context effects (see Knoeferle et al., 2014 for a discussion). Visual context effects can differ such that incongruencies between linguistic and visual input are processed differently compared to when the linguistic context matches the visual context. Thus, Knoeferle et al. (2014) further extended the CIA (Figure 5.1) on the basis of ERP findings which can reflect visual context effects in combination with verification responses (Knoeferle, Urbach, & Kutas, 2011). Knoeferle et al. (2014) observed different ERP responses for role mismatches than for action mismatches. Therefore, a truth value index was added to the interpretation (int$_{\text{truth value}}$) to account for discarded or mismatching representations. The authors further argued that verification is a part of language processing and should, thus, be included in the CIA (Knoeferle et al., 2014).

5.2 THE SOCIAL COORDINATED INTERPLAY ACCOUNT (SCIA)

Münster (2016) suggested further extensions to the CIA version of 2007 and Münster and Knoeferle (2018) included these extensions into the revised Coordinated Interplay Account from 2014 (Figure 5.2). The extensions of the CIA accommodate social-contextual aspects of language comprehension. Recall that in her experiments, Münster (2016) observed time course differences between child and adult language processing: Children's eye-movements towards the target character performing an action (vs. no action) were delayed by one word region. Thus, the author concluded that the age of the comprehender needs to be addressed in the CIA because it can influence the interpretation, expectation, and also working memory. Another finding was that the depiction of a positive prime face resulted in facilitation effects in younger adults but not in children. The author thus argued that social factors play a role in language processing which, so far, have been largely ignored by the CIA. She suggested to extend the account by a component which

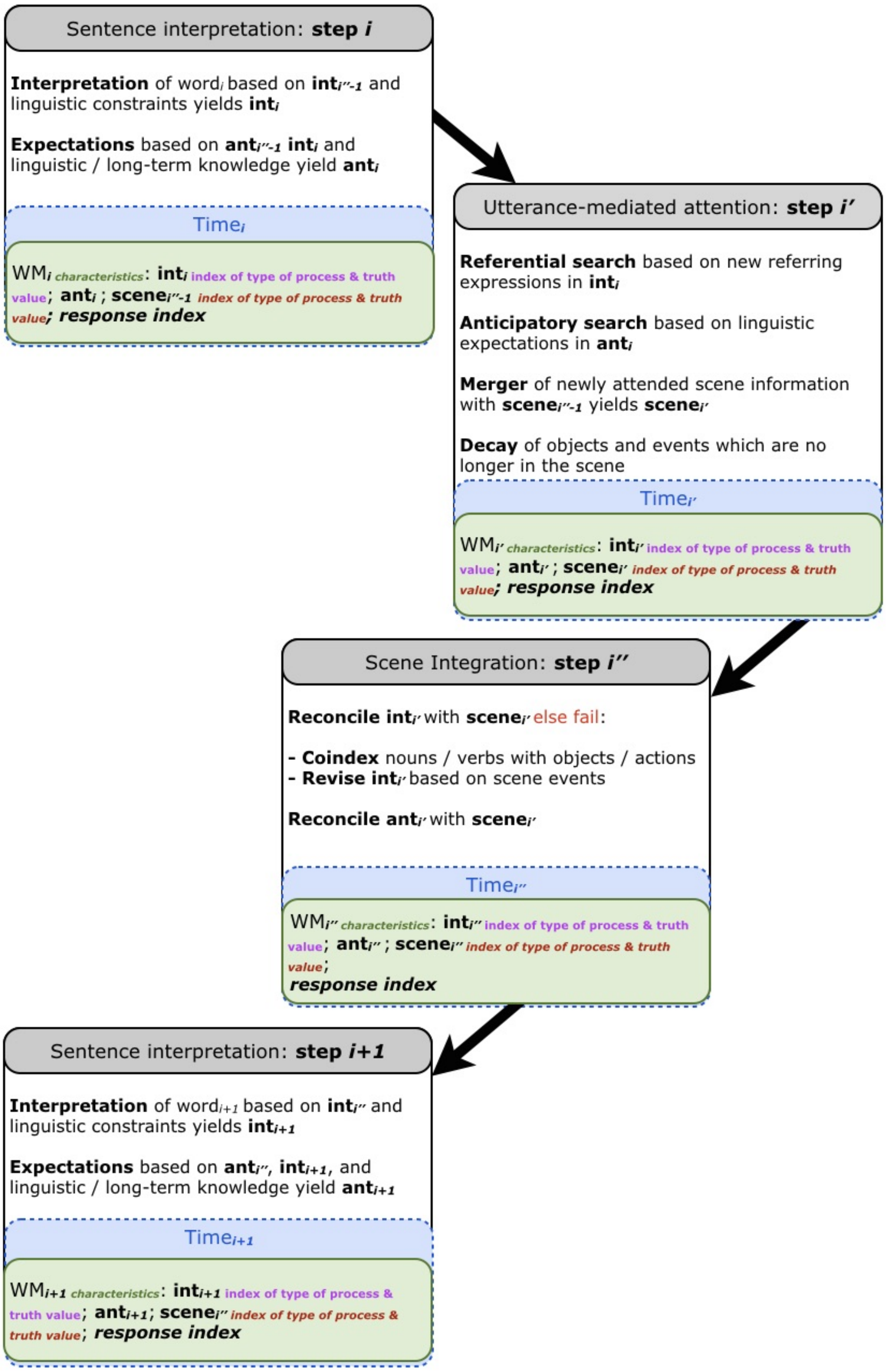


Fig. 5.1: *The revised Coordinated Interplay Account (Knoeferle et al., 2014).*

includes the internal properties of the comprehender (ProCom). ProCom includes biological (e.g., age) and experiential (e.g., literacy) characteristics of the comprehender (Münster & Knoeferle, 2018) and influences all stages of the CIA - sentence interpretation, utterance mediated attention, and scene integration. She further suggested to extend the expectations parameter (ant_i) to ant_s to incorporate social factors (social knowledge, social extra-linguistic expectations and social stereotypes). The account could further benefit from the results of the experiments presented in this thesis. Since our studies directly compare children and adults, the findings might support the idea of integrating properties of the comprehender (age), especially on a pragmatic level of sentence processing.

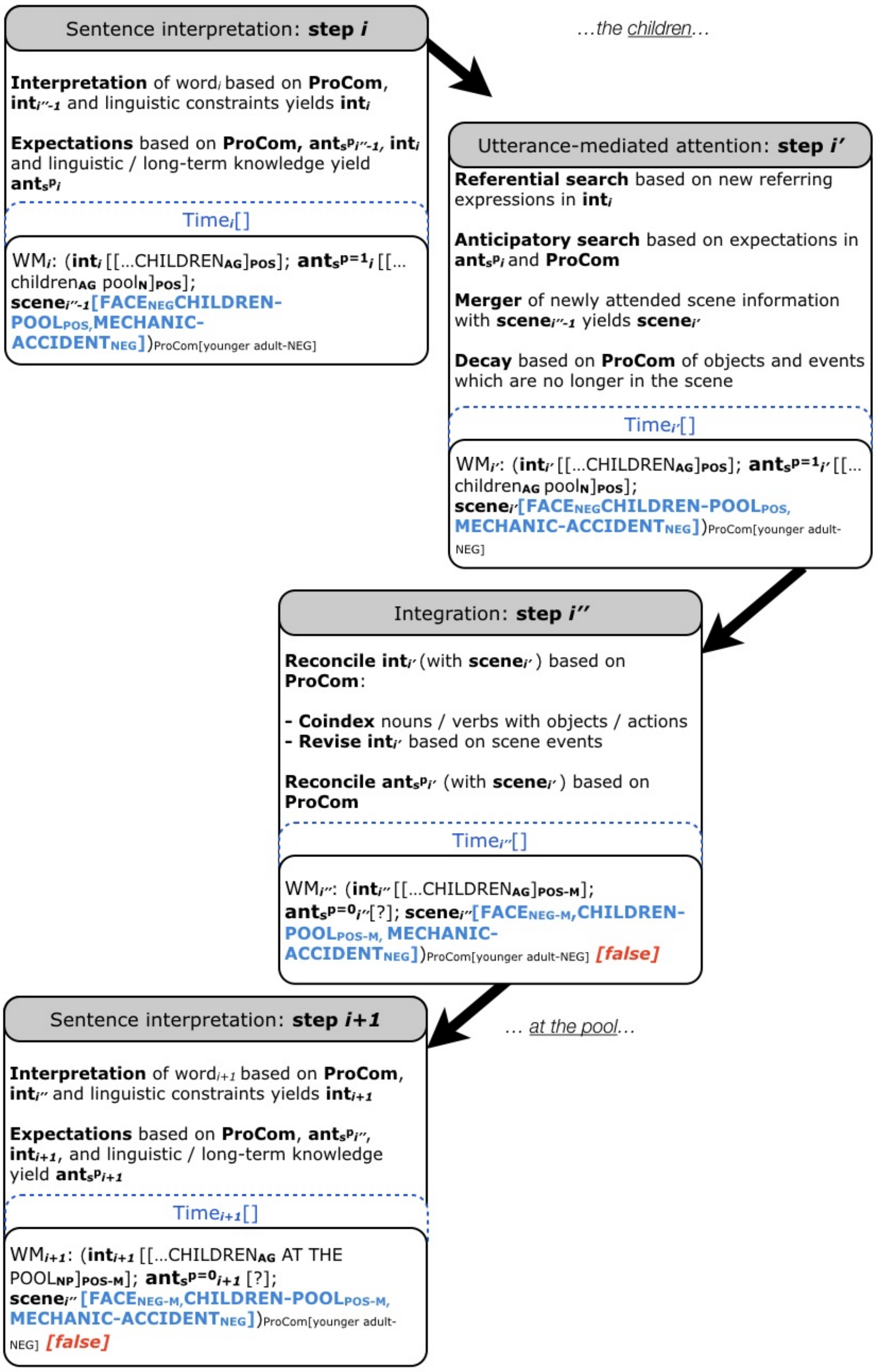


Fig. 5.2: *The Social Coordinated Interplay Account (Münster & Knoeferle, 2018).*

6

SUMMARY

In the previous five chapters, we reviewed studies on situated language comprehension. We started with the first existing findings which provided evidence for the close time-lock of visual attention and utterance comprehension. Linguistic, as well as non-linguistic information, can rapidly inform comprehension processes (e.g., semantic interpretation, syntactic disambiguation, and thematic role assignment). As a result, we highlighted the fact that the relation between spoken utterances and visual attention is bi-directional such that spoken utterances mediate visual attention towards objects in the scene. In turn, once these objects are visually attended, they can influence utterance comprehension.

Results from further studies suggested a tension between semantic interpretation and pragmatic inferencing (i.e., a temporal delay in the online computation of pragmatic inferences). Whether the existing findings extend to processes other than syntactic disambiguation (e.g., thematic role disambiguation) remains an open question. We have seen that adults are able to rapidly exploit case-marking, prosody, and non-linguistic visual information for incremental thematic role assignment.

Another question that remains unanswered is whether previous findings on adult language comprehension extend to other age groups (e.g., children). Although many researchers have investigated child language comprehension, much of the evidence is from offline studies. Language mediated visual attention is somewhat delayed in younger children but even young infants can relate spoken words to objects in the visual display. Real-time processing differences between children and adults emerged in the pragmatic domain. Unlike adults, children failed to rapidly exploit the visual referential context, likely because they could not compute pragmatic inferences online. However, children used other visual information (e.g., depicted actions) for thematic role assignment in unambiguous SVO and OVS sentences. Children also used prosody to an extent (when case-marking was unambiguous) and case-marking in supportive visual contexts to determine 'who does what to whom'.

Overall, findings from prior research suggested some differences between children's and adults' incremental thematic role assignment. Adults rapidly exploited case-marking, prosody, and depicted actions during online language comprehension. Children, however, rapidly recruited depicted actions and case-marking in supportive visual contexts but used prosody only when case-marking was unambiguous. Therefore, the evidence suggested some differences in children's and adults' incremental thematic role assignment.

Eye-tracking as a measure for language processing in combination with comprehension questions allows us to not only exactly distinguish real-time processing differences and/or similarities between children and adults, it also provides insights into the outcome of the comprehension process.

Part II

CASE-MARKING AND/OR PROSODY? EFFECTS ON REAL-TIME THEMATIC ROLE ASSIGNMENT

7 EXPERIMENT 1

Experiment 1[1] investigated whether five-year-old children and adults are able to use case-marking and/or prosody to assign thematic roles in unambiguous German SVO and OVS sentences. Case-marking is believed to be a strong cue to thematic role assignment and has been found effective for that purpose with adult participants (Kamide, Scheepers, & Altmann, 2003; Knoeferle, 2007; Matzke et al., 2002). Contradictory results have, however, been reported for the use of case-marking in five-year-old children. Existing findings suggest that children at the age of five struggle to exploit case-marking for thematic role assignment (Dittmar et al., 2008a). For ambiguous and unambiguous German SVO and OVS sentences, the results of an act-out task revealed that children preferred to interpret the sentences via (SVO) word order instead of case-marking. The first noun-phrase of the sentence was interpreted as the subject/agent (SVO bias), even if it was case-marked as the object/patient of the sentence. It seems that the understanding of accusative case-marking is difficult for children up to the age of 6 (Schipke et al., 2011). Children's ERP responses suggest that they are already able to resolve case-marking violations for SVO sentences at the age of three but such responses indexed no case violation processing for OVS sentences even at the age of six (Schipke et al., 2011).

Evidence from an offline study corroborated the results presented by Dittmar et al., 2008a. During a video-pointing task, children listened to transitive subject-object and object-subject sentences which were unambiguously case-marked. The results suggest that children at the age of six used case-marking in simple transitive sentences (Brandt et al., 2016), similar to the seven-year-olds in Dittmar et al. (2008a). Recent evidence from eye-tracking also suggests that already between the ages of four to five German children are able to use case-marking for correct thematic role assignment in unambiguous German SOV and OSV sentences (e.g., Özge et al., 2016). However, information from the visual display likely supported the interpretation of the linguistic input. The scenes provided plausible role relations (a hare, a fox,

1 A six-page conference paper reporting the results from this Experiment has been published in the Proceedings of the 39th Annual Meeting of the Cognitive Science Society: Kröger, Münster, and Knoeferle (2017).

and a cabbage: Hares eat cabbages and foxes eat hares) but they did not disambiguate role relations (case-marking did). Scenes contained world knowledge about who is the most likely agent and the most likely patient. Evidence by Zhang and Knoeferle (2012) and Münster (2016) suggests even clearer visual context effects. Depicted actions disambiguated role relations (only one role filler performed the action mentioned in the linguistic input). When the actions were absent, unambiguous case-marking alone was insufficient to disambiguate role relations, perhaps because they were not stereotypical (e.g., a bear painting a worm). Thus, the visual context seems to matter but it remains unclear to what extent children use case-marking in scenes that do not support thematic role assignment by either world knowledge associated with the characters or disambiguating action depictions.

Another useful cue that can help to establish a link between the linguistic input and the visual world is supra-sentential information. Prosody assigns, among others, focus to certain constituents of a sentence (e.g., via accentuation) and can help to disambiguate syntactic structures. Existing evidence from eye-tracking suggests that adults use prosody to overcome their SVO preference in ambiguous OVS sentences when the sentence was assigned an OVS-biasing prosodic contour (Weber et al., 2006). We discuss this study in more detail in the introduction of Experiment 2. Children also exploited prosody for thematic role assignment but in an act-out task (Grünloh et al., 2011). Thematic roles were presented in a reversed order in two short video clips followed by a static image of the scenes. The videos showed both agent->action->patient and patient->action->agent events, thus enabling participants to directly contrast these two types of opposing role relations. However, prosodic effects (of an accentuated first noun-phrase in OVS sentences) only emerged when case-marking was also present. When case-marking was absent (ambiguous OVS sentences), children used their (SVO) word order bias instead of prosody and interpreted the OVS sentences as agent-first sentences.

7.1 MATERIALS AND DESIGN

PARTICIPANTS. 24 young adults (mean age = 25.5) and 24 five-year-old children (age range 4.5 - 5.8, mean = 4.7, SD = .56) participated in the experiments. They were all monolingual native speakers of German[2] with normal or corrected vision and hearing. Young adults were students from Biele-

2 For all our studies, monolingual native speakers were native speakers of German who did not acquire a second language before the age of six.

feld University who were paid for participation. Children were recruited from kindergartens in and around Bielefeld. Children received a toy and a certificate as a reward. Adult participants gave consent before taking part in the experiment. For children, parents signed a consent form. Children gave consent in oral form. We asked them if they want to play a game on the computer. The Bielefeld University ethics committee approved the study.

speech stimuli. We instructed a linguistically trained female native speaker of German to record 24 unambiguous transitive German subject-verb-object (SVO) and object-verb-subject (OVS) sentences with biasing or neutral prosodic contours. We emulated the biasing prosodic contours by Weber et al. (2006)[3]. In SVO sentences, the biasing prosodic contour contained an L*+H accent on the first noun-phrase and an additional H* accent on the verb. OVS sentences were assigned an L+H* accent on the first noun-phrase. If a main stress falls on the verb, the main stress does not fall on the first noun phrase (indicating an patient/object first interpretation) and can thus be interpreted as a signal for an agent/subject first interpretation. Additionally, these prosodic contours allow a distinction early in the sentence. For the neutral prosodic contour, we instructed our speaker to pronounce the sentences with the smallest amount of intonation possible. Overall, the experiment consisted of four conditions: a) SVO-biasing prosody, b) SVO-neutral prosody, c) OVS-biasing prosody, and d) OVS-neutral prosody (Table 7.1). We did not manipulate the auditory materials with regards to intonation patterns. What we did, however, was insert 500 ms in between each word, starting with the SVO-biasing prosody sentences consecutively. We annotated the on- and offsets for NP1, verb, adverb, and NP2 for each sentence (Appendix A.1). Following from the onsets of the SVO-biasing prosody sentences, we adjusted the onsets within one item accordingly (SVO-neutral, OVS-biasing, OVS-neutral). Thus, all onsets within one item were identical. Transitive verbs were chosen on the basis of possible depiction (clear links from verb to depicted actions; e.g., for the verb *filmen* ('film') we used a camera to depict the action) and child friendliness (i.e., avoiding verbs that can be related to violence such as the verb 'shoot').

3 In Weber et al. (2006), the sentences were recorded by a female native speaker of German who was phonetically trained. She was instructed to distinguish SVO and OVS sentences prosodically in light of the local structural ambiguity in the visual scene.

Table 7.1: Conditions Experiment 1

Condition	Sentence Structure	Prosodic contour	Visualisation
a	SVO	L*+H(NP1), H*(verb)	
b	OVS	L+H*(NP1)	
c	SVO	neutral	
d	OVS	neutral	

VISUAL STIMULI. We created 24 scenes containing three clipart animal characters each (e.g., Figure 7.1). The middle character and one of the adjacent characters were depicted as performing identical actions. These actions were depicted in the form of objects or images. For example, for the verb *filmen* ('film') the two characters were depicted holding a camera (object) and for the verb *küssen* ('kiss') images of a red mouth with three hearts next to it were positioned on the character's mouth. The third character did not perform an action. The scene thus provided a context for both sentence structures but did not permit comprehenders to unambiguously identify the correct thematic role relations upon hearing the verb. The middle character was always role ambiguous. He could either be the agent or the patient of the scene: He was performing an action and was being acted upon. We dubbed the other character depicted as performing an action the *agent* role filler of the scene, the middle character the *ambiguous* role filler of the scene, and the character not performing an action the *patient* of the scene.

We used the database COSMAS II (Corpus Search, Management and Analysis System; Baayen, Piepenbrock, & Gulikers, 1995) to obtain frequencies per million words for all the animal characters we used as agents and patients. We then analysed the mean frequencies of agents and patients, obtained from the database, using a paired samples t-test. The results revealed no significant difference between the frequencies of the two (agent and patient; $t(1,23) = .001$, $p = .999$). Across all agent and patient pairs, one did not occur significantly more often than the other. We adjusted the size of the characters. They all appeared in similar height and width. We avoided stereotypicality in the scenes with the idea that participants could not identify role relations via world-knowledge (eg., *Der Eisär filmt als nächstes den Papagei*, 'the polarbear films next the parrot'; it is not stereotypical that a polarbear films a parrot).

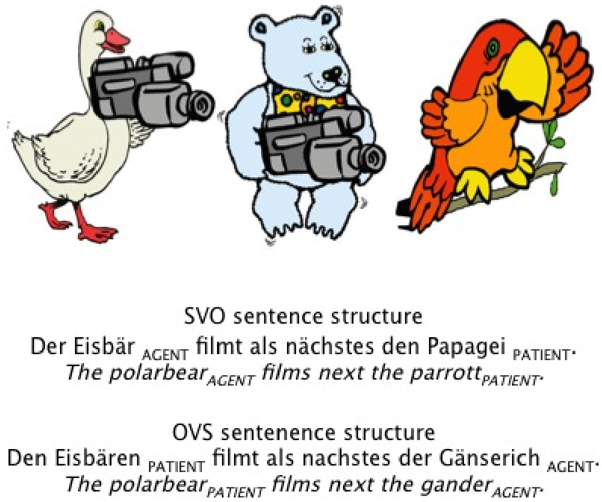


Fig. 7.1: *Example of a scene.*

pretest. All scenes were pretested with five-year-old children (N=19) who were not participants of the main experiment. Children were instructed to point at one animal character within one scene (e.g., *Zeig mal auf den Eisbären*; 'Point at the polarbear'). We counterbalanced scenes for left and right (left-right-left-right) and presented each of the scenes four times to the children: Three times to point at each of the animals and once to verify the depicted actions. For these we asked questions containing the target verb and one of the other verbs used in other items (eg., *Filmt der Gänserich den Eisbären oder schubst der Gänserich den Eisbären?*; 'Does the gander film the polar bear or does the gander push the polar bear?'). The results revealed a very high response accuracy for each animal, in many cases 100%. Across all agent, ambiguous, and patient role fillers, correct responses were very high (Figure 7.2). For the verbs, percentages were slightly lower. One of the reasons for this was that for the verb *interviewen* ('interview'), children's correct responses were below chance level (32%). For this verb, the animal characters in the scene were holding a microphone in their hands. We decided to keep the visual scene and exchange the verb *interviewen* ('interview') with the non-English word *befragen* ('question'). Before each experimental trial, we showed the scene to the children and asked them to answer the following question: *Befragt der Hund den Grashüpfer oder besprüht der Hund den Grashüpfer?* 'Does the dog question the grasshopper or does the dog spray the grasshopper?'. Children correctly responded to this question in 58% of all 24 testing sessions.

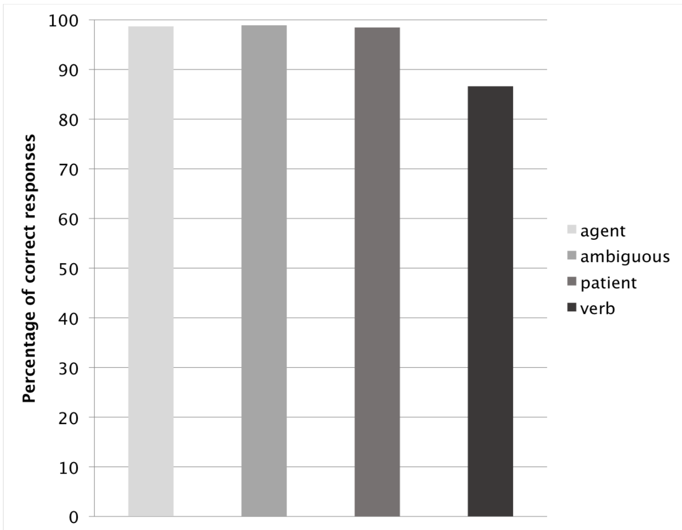


Fig. 7.2: *Percentage of correct responses in the Pretest per role filler type (agent, patient, ambiguous, and action).*

DESIGN. Each animal character filled both roles (that of the agent and the patient). To counterbalance the role an animal portrayed, we created additional scenes. For example, in one scene, the gander is depicted as filming (agent). In a counterbalancing scene, the gander is the patient - not holding an object (Figure 7.3). The counterbalanced animal characters were not identical to the animal characters in the other scenes. They depicted the same animal character but were not identical in shape. We did this to avoid any effects of recognition (i.e., each animal occurred both in agent and in patient role).

We created eight lists following a Latin square design. Two prosodic contours and two sentence structures resulted in four lists. In order two avoid a left or right bias in the visual display, we added left and right balancing to the Latin square design, resulting in eight lists. Hence, each scene appeared once in each direction, either with all characters facing left or right. We implemented the left/right counterbalancing by mirroring each image horizontally. We pseudo-randomised all lists resulting in 24 different lists.

Within each list, we allowed for each condition to occur twice in a row but not more than twice. Eight filler trials were added to the experiments. We kept the number of fillers constant to increase the similarity of the experiment between the two age groups. The filler scenes contained either three role fillers (similar to the experimental items) or two role fillers. The sentences were structured SVO but two filler sentences contained an adverb at the beginning, five had coordinated NPs, and one in which only one animal character was mentioned. Filler trials were distributed randomly but did not follow one another.

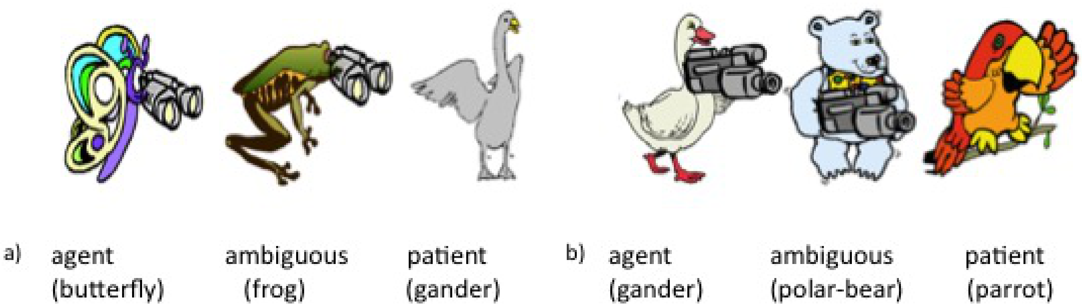


Fig. 7.3: *Counterbalancing scenes: The gander is the patient in scene (a) and the agent in scene (b).*

PREDICTIONS. Existing findings suggest that the combination of eye-movements (or ERP responses) and verification tasks (in our case post-sentence comprehension questions) can inform online language comprehension (Knoeferle et al., 2011). Responses to post-sentence comprehension questions should thus be in line with the eye-gaze patterns such that looks to the patient (vs. agent) should result in an SVO interpretation and looks to the agent (vs. patient) in an OVS interpretation. Additionally, children's cognitive resources (K-ABC data) may influence visual attention (Zhang & Knoeferle, 2012).

Prior research has reported rapid effects of case-marking on adults' thematic role assignment (Kamide, Scheepers, & Altmann, 2003; Matzke et al., 2002). Although our scenes depicted actions ambiguously, we expected to replicate effects of case-marking on eye-movements: Adults should anticipate the correct target role filler in SVO and OVS sentences, depending on case-marking (i.e., anticipate the patient (vs. agent) in SVO sentences and the agent (vs. patient) in OVS sentences). Prosody influenced adults' grammatical function assignment in locally structurally ambiguous sentences (Weber et al., 2006). If adults rapidly exploit prosody in addition to case-marking, there should be a clear difference in the onset of looks towards the target role filler in the biasing prosody conditions compared to the neutral prosody conditions. Additionally, we might find more looks towards the patient (vs. agent) in SVO sentences and the agent (vs. patient) in OVS sentences in the biasing prosody conditions compared to the neutral prosody conditions. Overall, prosody should influence the number and timing of looks towards the target role filler but not adults' offline performance if adults use case-marking to determine 'who does what to whom' in ambiguous action scenes. For the comprehension-data, we expected a high number of correct responses in all conditions in the absence of clear differences between the two prosodic conditions.

To which extent children are able to rapidly exploit case-marking for thematic role assignment like adults depends on whether they can exploit the event relations even though the scenes in Experiment 1 depicted actions ambiguously. Existing findings regarding children's ability to use case-marking have revealed contradictory results (Dittmar et al., 2008a but Özge et al., 2016). If children need a more supportive context to exploit case-marking, we should not observe any clear effects of case-marking on incremental thematic role assignment (i.e., more looks to the patient vs. agent in both SVO and OVS sentences). Alternatively, depicted event relations may still be helpful to integrate object case-marking, resulting in more looks to the agent (vs. patient) in OVS sentences (e.g., seeing a polar bear as the patient of a filming event may help children to interpret the sentence as OVS) and more looks to the patient (vs. agent) in SVO sentences.

In an offline study, children used prosody when case-marking was unambiguous (Grünloh et al., 2011). However, children seemed to struggle with pragmatic inferences during online language comprehension (Huang & Snedeker, 2009b; Trueswell et al., 1999). If children do not use case-marking but do use prosody, we should observe more looks to the patient (vs. agent) in the SVO-biasing compared to the SVO-neutral prosody condition. For OVS sentences, we should observe more looks to the agent (vs. patient) in the biasing prosody condition and more looks to the patient (vs. agent) in the neutral prosody condition. Otherwise, we might observe the effect of an SVO bias (more looks towards the patient versus agent), independent of case-marking and prosody. Additionally, if children's cognitive resources influence their gaze-patterns, we may see looks to the target character in children with higher (vs. lower) cognitive abilities.

Prosody could, in principle, influence children's comprehension data (if they do not use case-marking). For the post-sentence comprehension questions, we expected more correct responses in SVO sentences than in OVS sentences and more in the biasing prosody than in the neutral prosody conditions, provided that children are able to use prosody. If children cannot yet use prosody (and do not use case-marking) for correct thematic role assignment, we should find a high number of correct responses in SVO sentences, whereas children's correct responses for OVS sentences should be at chance level (similar to the findings by Dittmar et al., 2008a). Response accuracies may also be higher in children with higher cognitive abilities (vs. lower). Zhang and Knoeferle (2012), for example, suggested that children's cognitive abilities influence effects of visual context.

PROCEDURE. Participants' eye-movements were monitored with an Eyelink 1000 eye tracker with a sampling rate of 500 Hz monocular and an average accuracy of 0.5° in the remote set-up using a 16mm lens. The estimated head distance was about 500 Hz using a target sticker on participants' foreheads. We used the remote set-up to conduct the experiment with children in kindergartens. Images were presented on a DELL laptop with a screen resolution of 1920x1080 pixels. The laptop was placed on a home-made laptop stand and the eye tracker was placed in front of the stand. This way, participants looked straight at the middle of the laptop screen without the eye tracker blocking the view. Each testing session lasted approximately 20 minutes. Before starting the experiment, a calibration was conducted manually. We used a five-dot calibration scheme. A drift-correct-dot separated the trials to ensure accurate calibration of the eye tracker and the same gaze starting point for each scene. The experiment started with a written instruction for adults and a smiley and an oral instruction for children.

Prior to the experimental items, three practice items familiarised participants with the set-up and the task at hand. The design of the practice items was identical to the design of the experimental items. Sentences were either structured SVO or OVS and prosody was biasing or neutral. After the practice trials, another calibration followed before starting the experiment. During each trial, the target sentence followed 2000 ms after target display. We added 500 ms after the offset of NP2 to capture late effects. 1500 ms after the sentence the comprehension question followed (Figure 7.4) which was either in active or in passive voice (e.g., *Wer filmt hier?* 'Who films here?' or *Wer wird hier gefilmt?* 'Who is being filmed here?'). Across all lists, we balanced the type of question asked to avoid priming effects of sentence structure (active voice rather priming SVO sentence structure and passive voice rather OVS sentence structure). Across the first four lists, four active questions were followed by four passive questions whereas in the last four lists four passive questions were followed by four active questions. Therefore, each item occurred once with an active and once with a passive question.

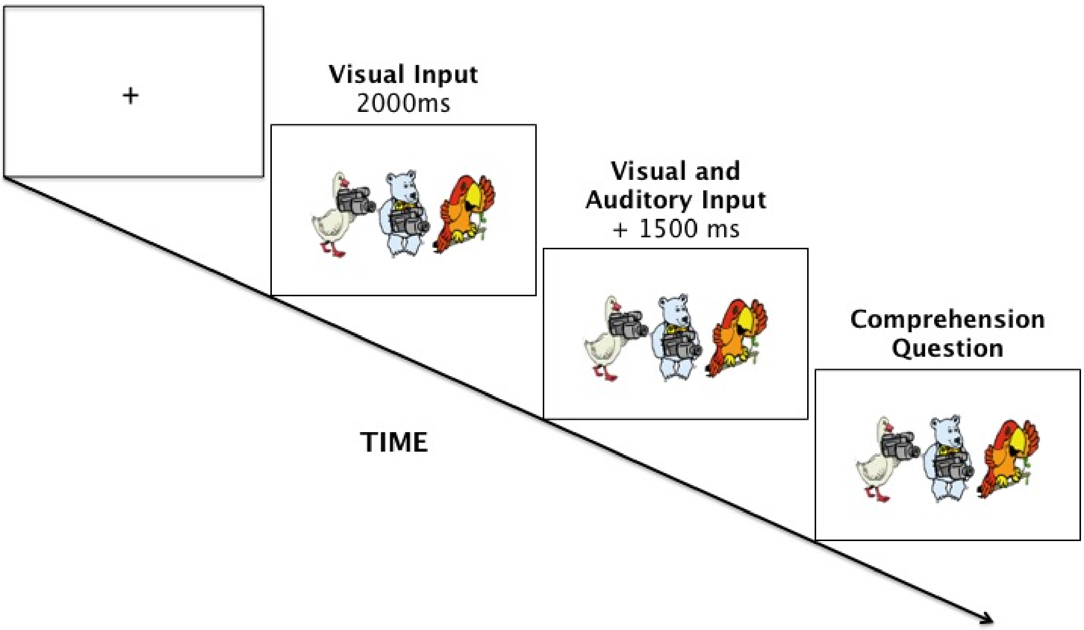


Fig. 7.4: *Visualisation of the procedure for one trial.*

ANALYSES. For the analyses, we first predefined three areas of interest (AOIs) in the visual scenes: Left, middle, and right. We created the AOIs after we had run the first experiment within the program Data Viewer (SR Research Ltd., Canada). In order to avoid a bias for the size of the AOIs, we turned off the visualisation of the fixations before we created the areas of interest free hand for one direction (e.g., either agent->ambiguous-> patient or patient->ambiguous-> agent). We then mirrored the AOIs, using a JAVA conversion tool. We verified that the interest areas were successfully mirrored by calculating the coordinates manually in Excel (screen resolution minus coordinates of original interest area: e.g., coordinate 1 original = 831; 1920 − 831 = 1089; new coordinate 1 = 1089). The results revealed that the AOIs were successfully mirrored.

The middle AOI always contained the ambiguous role filler, whereas the left and right AOIs could either contain agent or patient role fillers, depending on sentence structure and balancing. Regardless of condition, the ambiguous role filler occurred in NP1 position. We thus excluded the middle AOI (ambiguous role filler) from our analyses. To verify that both children and adults inspected the ambiguous role filler during the first noun-phrase, we present mean fixation counts to agent, ambiguous, and patient role fillers in the age group comparisons (subsection 7.2.3, Figure 7.14). One of the main reasons for analysing only left and right AOIs was that we were interested in anticipatory eye-movements (towards the agent or the patient in the scene). We merged fixations with shorter durations than 80 ms - they do not reflect information processing (Rayner, Sereno, Morris, Schmauder, & Clifton Jr, 1989; Sturt, Keller, & Dubey, 2010). For the analyses, we report marginally significant results when p values were between .05 and .10. We report these effects in detail to capture all possible influences of our manipulations. We do not report all the details of non-significant results to keep the results report concise (only F and p values).

Time Course Graphs. We created Time Course Graphs to get an overview of the visual bias to the agent or the patient in the scene throughout the time course of the sentence, assuming that looks to the patient represent an SVO interpretation and looks to the agent an OVS interpretation (Knoeferle, Crocker, Scheepers, & Pickering, 2003; Knoeferle et al., 2005; Scheepers & Crocker, 2004). First of all, we deleted all fixations with durations longer than 800 ms - the eye tracker might have momentarily lost the eye (Rayner et al., 1989; Stites & Federmeier, 2015; Sturt et al., 2010). Afterwards, we deleted all cases which included the middle interest area (ambiguous role filler) because we were interested in anticipatory eye-movements towards agent or patient. A new variable was created including agent and patient role which was dependent on the balancing (left/right), and the AOI (left/right): Variable 'role' (e.g., when the balancing is left - agent -> ambiguous -> patient - and the AOI is left, the role is agent).

We created time bins of 20 ms which contained fixations at each point in time. The data was then aggregated by role, word order, and prosody, with the aggregated variable containing the sum of fixations from each time bin. The outcome of this included the number of fixations by condition for each 20 ms time frame. In order to calculate log-ratios, we added 0.5 to each fixation sum (log-ratios of 0 are undefined, division by 0 is also undefined) and then computed log-ratios of agent over patient (Arai, Van Gompel, & Scheepers, 2007; Carminati & Knoeferle, 2013). Log-ratios are a relative measure that represent the bias of looks towards one character relative to the other (ln(SVOagent)/(SVOpatient); ln(OVSagent)/(OVSpatient)). Using these log-ratios, we plotted looks towards the patient (negative numbers) over looks towards the agent (positive numbers). When the log-ratio is 0, no preference to inspect one character over the other exists (equal number of looks to agent and patient; Arai et al. (2007)). We calculated mean onsets of each word region to illustrate that participants looked more towards the agent (vs. patient) or the patient (vs. agent) of the scene. For the NP2 region, we used the longest NP2 offset as the offset of NP2 (Table 7.2).

Table 7.2: Mean Onsets and Longest Word Durations Experiment 1 (in ms)

	NP1	Verb	Adverb	NP2
Mean onsets	0	1437	2726	4012+500
Longest duration	1870	1592	1686	1608 (1108 +500)

Word Regions. For the word region analyses, we predefined two word regions of interest: Verb and adverb region of the sentence (for each trial: Beginning of verb onset to adverb onset for the verb and beginning of adverb onset to NP2 onset for the adverb; see Appendix A.1)[4]. These two regions were defined on the basis of the prosodic structure of the sentences. We focused on the verb region because towards the end of the verb the two prosodic structures can be distinguished from one another. Whenever there is nuclear stress on the first noun-phrase, the verb experiences a fall in stress. When there is no main stress on the first noun-phrase, the main stress is on the verb (Weber et al., 2006). We were further interested in the adverb region to capture postverbal eye-movements. Additionally, we analysed the NP1 and NP2 region for early and late effects of prosody and case-marking. However, disambiguation effects during NP2 may overlap with effects of word mention on eye-movements. The results the NP1 word region are only reported if we observed significant effects, the results of NP2 only when the data revealed additional effects which did not occur in the verb or adverb region.

Just as for the time course graphs, we deleted all fixations with a duration longer than 800 ms. For the verb and adverb region, we added new variables (each in a separate document), including verb and adverb duration (verb region: Verb onset to adverb onset; adverb region: Adverb onset to NP2 onset).

4 Corrections for the number of time windows were not made. If we adjusted the p-values after Bonferroni (two analyses regions, adjusted $p = .025$), the key results and also the conclusions would still hold (see also Knoeferle et al., 2014, page 136).

Using the SPSS syntax window, we then calculated fixations within 20 ms time bins for the longest duration of the verb and adverb region across all sentences (Table 7.2) and created a new variable that included the number of fixations within the respective time region. We then aggregated the data by subject, role, word order, and prosody for the by participants analysis (F_1) and by item, role, word order, and prosody for the by items analysis (F_2). Following this, we restructured the data from cases into variables which then consisted of the fixation sum per subject/item, role, word order, and prosody. Finally, we calculated log-ratios of agent over patient for each condition. On the basis of the log-ratios, we conducted Analyses of Variance (ANOVAs) following a two (word order) by two (prosody) design by participants and by item for the verb and adverb region.

ANOVAs assume normality of the sampling distribution of means but are robust against violations of this assumption when they are mild or moderate (Keppel & Zedeck, 1982, see Luck, 2014, page 317 for a discussion). Eye-tracking data may not be normally distributed but log-transformation can help to correct towards a normal distribution (Tabachnick & Fidell, 2007, page 94). In order to find out whether the data are normally distributed, the Kolmogorov-Smirnov test can be applied which is a normality test used for large data sets. However, the results of the test have to be interpreted with caution since they can be significant in large data sets even though the data is only slightly different from a normal distribution. Therefore, it has been suggested to additionally consider P-P or Q-Q plots for interpretation (Field, 2009, page 148). Within the data used in the experiments, violations of normality were only mild in some sets of data based on visual observation of Q-Q plots of log-transformed data. Additionally, ANOVAs are expected to be robust with 20 degrees of freedom for error (Tabachnick & Fidell, 2007, page 246).

To compare both age groups, we conducted a two by two ANOVA (word order and prosody) including age as a between-participants factor. We present results in graphs showing the mean log-ratios of looks of agent over patient. We mainly display the interaction graphs since they provide a good overview of all conditions. However, if main effects occurred in the comparisons between children and adults, we also depict graphs showing the main effects only to visualise the data clearly. All positive numbers represent a preference of looking at the agent (vs. patient) in the scene and all negative numbers a preference of looking at the patient (vs. agent). Interaction effects were analysed further by means of paired sample t-tests. We manually applied a Bonferroni correction to the α level (.05/6) to safeguard against Type 1 errors: New $\alpha = .008333$ (Field, 2009).

Accuracy. During each testing session, we recorded participants' responses to post-sentence comprehension questions manually for each trial (1 for correct responses and 0 for incorrect responses). For the analyses, we did not exclude cases in which participants did not give an answer but annotated those as incorrect responses (no response = incorrect response). For each condition, we calculated the sum of all correctly answered comprehension questions then calculated the results in percent (number of correct responses by condition divided by number of possible correct responses per condition). We analysed the accuracy data using Generalised Linear Mixed Effects Models (Lme4 R package) which have been found to be effective for binomial data (D. Bates, Mächler, Bolker, & Walker, 2014). We defined two factors following from the manipulations of the experiment (word order and prosody) and included them as fixed factors.

Using the scale function in R[5], we first centred the factors before we ran the models. We started with the most complex model: Interaction of fixed effects plus the interaction of fixed effects in the random slopes including random intercepts for participants and items[6] (Barr, Levy, Scheepers, & Tily, 2013) and followed with a stepwise reduction for model comparison (Jaeger, 2011). If the most complex model did not converge, we simplified it by removing the interaction in the random slopes and replaced it by main effects in the random slopes. If this model did not converge either, we simplified it further by removing first factor1 and then factor2 from the random slopes, leaving random intercepts for participants and items only. If neither of those two models converged, we kept only the two main effects and removed the interaction. Once one of the models converged and the following simplified model also converged, we ran an ANOVA for model comparison. If the two models were significantly different, we kept the output of the more complex model (i.e., the maximal converging model; see Appendix H for all maximal converging models).

Additionally, we separated active and passive voice comprehension data and analysed the subsets using Generalised Linear Mixed Effect Models. The rationale behind this method is that existing findings on child language com-

5 $accuracydata\$factor = scale(as.numeric(accuracydata\$factor1))$ $accuracydata\$factor = scale(as.numeric(accuracydata\$factor2))$

6 e.g., Experiment 1: $acc.glmer = glmer(accuracy \sim wordorderC * prosodyC + (1 + wordorderC * prosodyC|subject) + (1 + wordorderC * prosodyC|item), data = accuracychildren, family = binomial)$

prehension data have revealed effects of voice and additional effects within the passive voice comprehension data (Münster, 2016).

The accuracy data might furthermore interact with the eye-tracking data. We thus computed the percentages of correct responses by participants and divided the data in high vs. low accuracy, using the median of these percentages. We used the median instead of the mean because the median is free from the influence of very high and very low values. We then ran a two (word oder) by two (prosody) repeated-measures ANOVA including accuracy (high vs. low) as a between-participants factor.

K-ABC. The Kaufmann Assessment Battery for Children (K-ABC) is used to assess children's intelligence and skill sets. It is mainly applied in an educational or clinical setting to account for differentiated learning in disabled children. However, the test battery has also been found to be useful for research on children's cognitive abilities (Melchers & Preuß, 1994). For our experiments with children, we used three subtests from the assessment battery to determine possible cognitive variability between child participants. A high variability could, in principle, influence the eye-tracking and comprehension data (e.g., Zhang & Knoeferle, 2012). Children were evaluated on three subtests: *Word Order*[7], *Spatial Memory*, and *Number Recall*. During *Word Order*, children were asked to name the objects they were presented with: A house, a key, a bird, and a cup.

The second task was to point to the images in the same order the experimenter had mentioned them (eg., house - key: Point towards the house first and then towards the key). This test probes, among others, for children's ability to connect a given auditory input with their motor responses (Melchers & Preuß, 1994). In the *Spatial Memory* subtest, the experimenter showed five images including objects or animals. Children were asked to remember the position of the objects. Approximately five seconds after exposure, the experimenter covered the pictures with another page that showed an empty grid. Children were asked to point at the position in the grid where the previously-seen object had been. This subtest investigates, among others, perceptual organisation and visual short-term memory (Melchers & Preuß, 1994).

In the last subtest (*Number Recall*), participants were asked to repeat numbers in the same order the experimenter mentioned them in (starting with two numbers in a row and going up to 6 numbers). This subtest evaluates chil-

7 Note that Word Order in this context stands for a series of different words.

dren's attention span and auditory-linguistic memory performance (Melchers & Preuß, 1994). Each testing session lasted approximately five minutes. We recorded the number of correct responses manually (1 for a correct and 0 for a false response).

Following from these tests, we conducted a two-tailed Spearman's rank-order correlation to account for a possible relationship between the accuracy data and the K-ABC data. We used the two-tailed test because we could not predict whether cognitive abilities improved or reduced post-trial accuracies (Field, 2009). For each participant, we computed the percentages for accuracy (number of possible correct answers 24) and K-ABC (number of possible correct answers 22).

Furthermore, children's cognitive abilities could, in principle, influence the eye-tracking responses (Zhang & Knoeferle, 2012). Therefore, we divided the K-ABC data into high and low results (via a median split). We then ran two (word order) by two (prosody) repeated-measures ANOVAs for verb and adverb region, including K-ABC (high vs. low) as a between-participants factor.

7.2 RESULTS AND DISCUSSION

7.2.1 *Adults*

TIME COURSE GRAPHS. At the beginning of the NP1 region, participants looked more at the agent (vs. patient) in SVO sentences but shifted their gaze towards the patient (vs. agent) towards the middle of the word region. In OVS sentences no such preference was observed. Participants inspected both characters (agent and patient) equally. However, the overall number of looks towards the agent (vs. patient) or patient (vs. agent) during this time region was very low, presumably because participants looked at the ambiguous character being mentioned (Figure 7.14). Participants might prefer to look at the agent (vs. patient) slightly more because the role filler is holding the same object as the ambiguous character and is thus possibly more salient than the patient role filler who is not holding an object.

From the beginning of the verb region onwards, we observed more looks towards the patient (vs. agent) in SVO sentences and more looks to the agent (vs. patient) in OVS sentences in the absence of a clear difference between the two prosodic conditions. At the beginning of the adverb region, we observed that the preference to look at the patient decreased in SVO sentences (in both prosodic conditions). Participants looked equally much at the agent and the patient at the beginning of the adverb region. After the verb, participants might have inspected the other role filler performing an action (the agent) simply because the action is mediated by the verb. Since the scene depicts two characters performing an action, participants may have inspected the agent more than the patient (not performing an action). Another explanation for this behaviour might lie in the set-up of the visual scene. After having heard the verb, participants may want to ensure they interpreted the sentence correctly, since the alternative interpretation is also available in the scene. In OVS sentences, this kind of behaviour did not occur. Following from the previous possible reasoning, participants inspected both characters holding an object to interpret the linguistic input and thus do not inspect the patient.

Towards the end of the adverb region, this pattern again transformed into more looks to the patient (vs. agent). No such pattern occurred in OVS sentences. Participants preferred to inspect the agent (vs. patient) throughout the whole adverb region. Again, no clear differences between the two prosodic conditions were observed. During the NP2 region participants again looked more at the patient than the agent in SVO sentences. In OVS sentences, participants continued to look more to the agent (vs. patient). Interestingly, it seems that the overall target preference was slightly higher for OVS sentences compared to SVO sentences. Participants inspected the agent (vs. patient) more in OVS sentences than the patient (vs. agent) in SVO sentences. Throughout the time course of the whole sentence, clear differences between the two prosodic conditions were absent.

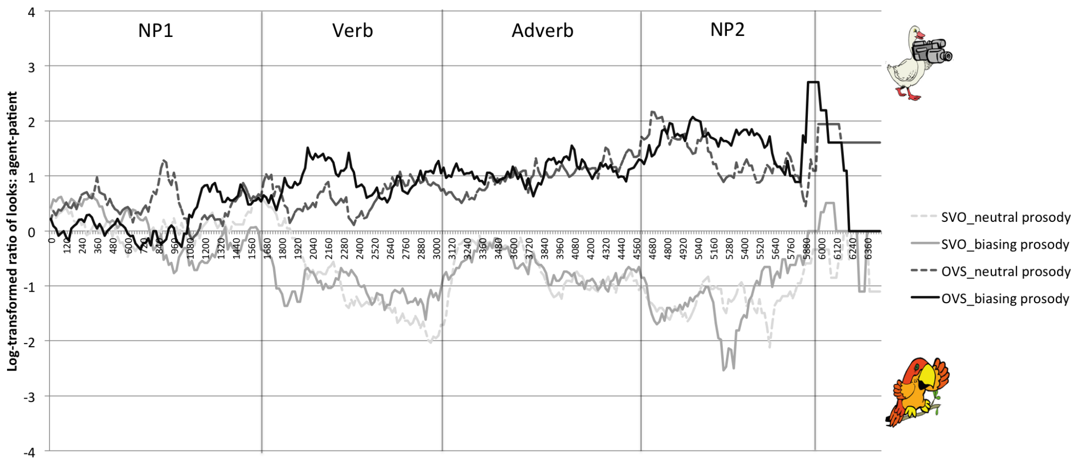


Fig. 7.5: *Experiment 1 adults: Time course of eye-movements including the visual bias (y-axis) towards agent (positive values) relative to patient (negative values) in all conditions (ln(agent/patient)) over time in ms (x-axis). The vertical lines represent mean onsets for each word region.*

WORD REGIONS. In the verb region (Figure 7.6a) we observed no reliable effects of prosody, $F_1(1, 23) = .037$, $p = .849$, $F_2(1, 23) = 1.030$, $p = .321$. However, the data revealed a significant effect of word order, $F_1(1, 23) = 23.178$, $MSE = 4.605$, $p < .001$, $\eta^2 = .502$, $F_2(1, 23) = 49.709$, $MSE = 1.558$, $p < .001$, $\eta^2 = .684$. Participants looked more at the patient (vs. agent) in SVO sentences and at the agent (vs. patient) in OVS sentences. There was no interaction of word order and prosody, $F_1(1, 23) = 1.080$, $p = .310$, $F_2(1, 23) = .995$, $p = .329$. Descriptively, the target preference was slightly higher in the biasing prosody conditions compared to the neutral prosody conditions, indicating some sensitivity to prosody (Figures 7.6 a) and b)).

In the adverb region (Figure 7.6b), the data showed no reliable effects of prosody, $F_1(1, 23) = .054$, $p = .818$, $F_2(1, 23) = .231$, $p = .635$. The effect of word order was again significant $F_1(1, 23) = 64.212$, $MSE = 2.408$, $p < .001$, $\eta^2 = .736$, $F_2(1, 23) = 69.549$, $MSE = 2.803$, $p < .001$, $\eta^2 = .751$ but the interaction of word order and prosody was not reliable, $F_1(1, 23) = .598$, $p = .447$, $F_2(1, 23) = .000$, $p = .993$. Participants directed looks to the target role filler depending on case-marking on the first noun-phrase (indicating SVO or OVS word order interpretations).

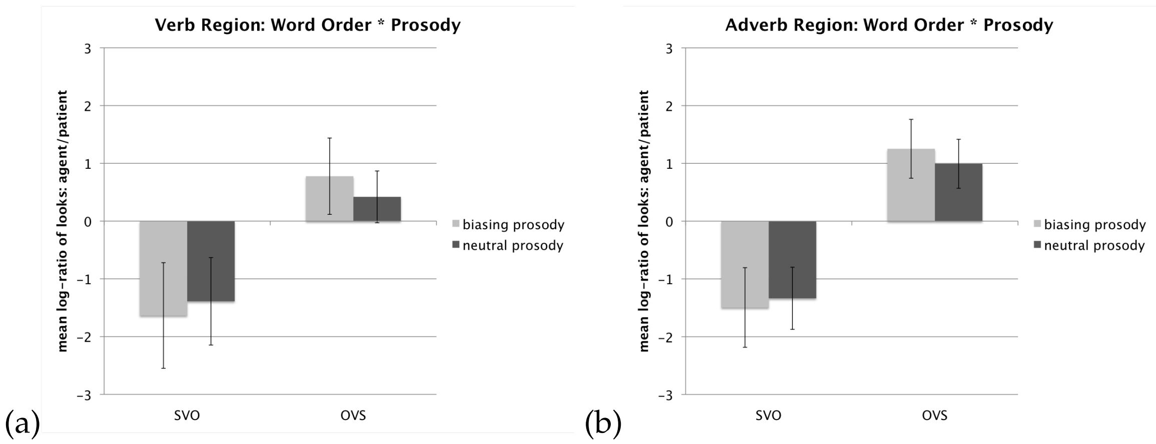


Fig. 7.6: *Experiment 1 adults: Mean log-ratios of looks towards the agent (positive values) relative to the patient (negative values) during the verb (a) and adverb (b) region in all four conditions. The graphs represent the interaction of word order and prosody. Error Bars reflect 95% Confidence Intervals.*

ACCURACY. Response accuracy was overall very high (99%) in the absence of a clear difference between conditions. For SVO sentences, in both prosodic conditions, the percentage of correct responses was 100%. For OVS sentences with a biasing prosody, response accuracy was 98.6%. For OVS sentences with a neutral prosody 99.3% (Figure 7.7). The difference in the number of correct responses between SVO and OVS sentence structures likely resulted from the mere fact that SVO is, in general, preferred over OVS sentence structure in German (Dittmar et al., 2008a; Hemforth, 1993; Hemforth & Konieczny, 2013). Mixed Effects Models (Maximal Converging Model: accuracy ~ wordorderC + prosodyC + (1 | subject)) revealed no reliable effects of word order or prosody and no interaction. Given that response accuracy was very high in all conditions, null effects of case-marking or prosody were plausible.

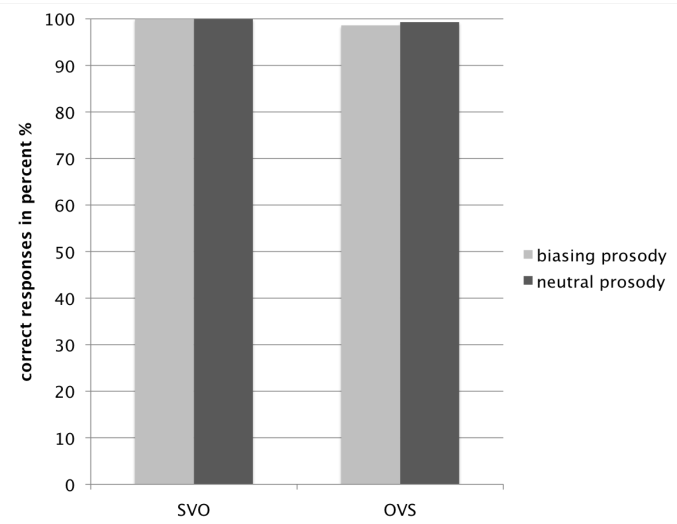


Fig. 7.7: *Experiment 1 adults: Correct responses to post-sentence comprehension questions by condition (x-axis) in percent (y-axis).*

7.2.2 *Children*

TIME COURSE GRAPHS. We observed a mild tendency of looks towards the agent (vs. patient) during the NP1 region for SVO sentences but not for OVS sentences. From the beginning of the verb region until the end of the sentence, participants looked more at the patient (vs. agent) in all conditions in the absence of clear effects of prosody. At the beginning of the adverb region, children looked even more at the patient (vs. agent) in the OVS-biasing prosody condition compared to the OVS-neutral prosody condition. From the beginning of the NP2 region (the mentioning of the target character), children looked more at the patient (vs. agent) in SVO sentences and more at the agent (vs. patient) in OVS sentences.

Similar to the adult data, children's preference to look at the patient (vs. agent) decreased slightly at the beginning of the adverb region, especially in the SVO-neutral prosody condition. This effect seems to be persistent across both age groups. Perhaps both, adults and children, made inferences from the linguistic input (the verb) and thus inspected animals performing actions in the scene (the ambiguous and agent role filler but not the patient). For adults this was only true for SVO sentences. For children, this gaze-pattern occurred in both sentences structures.

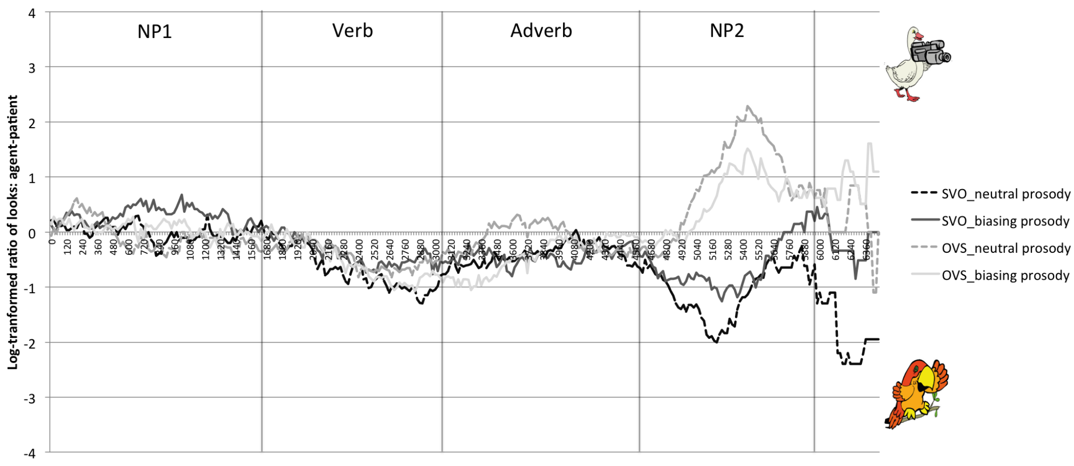


Fig. 7.8: *Experiment 1 children: Time course of eye-movements including the visual bias (y-axis) towards agent (positive values) relative to patient (negative values) in all conditions (ln(agent/patient)) over time in ms (x-axis). The vertical lines represent mean onsets for each word region.*

WORD REGIONS. The results of the verb region (Figure 7.9a) revealed no significant effects of word order, prosody or an interaction of both (word order: $F_1(1,23) = .483$, $p = .494$, $F_2(1,23) = .533$, $p = .624$; prosody: $F_1(1,23) = .623$, $p = .438$, $F_2(1,23) = 1.265$, $p = .272$; interaction word order and prosody: $F_1(1,23) = 1.252$, $p = .275$, $F_2(1,23) = .255$, $p = .618$).

In the adverb region (Figure 7.9b), there were no significant effects of word order or prosody and no significant interaction of word order and prosody (word order: $F_1(1,23) = .640$, $p = .432$, $F_2(1,23) = .247$, $p = .624$; prosody: $F_1(1,23) = 2.029$, $p = .168$, $F_2(1,23) = .016$, $p = .901$; interaction word order and prosody: $F_1(1,23) = .017$, $p = .898$, $F_2(1,23) = .081$, $p = .778$). Participants looked more at the patient (vs. agent) in all four conditions. The grand mean was overall negative and significantly different from zero, for both analyses by participants and by item, intercept $p < .001$.

ACCURACY. The data revealed an overall response accuracy of 71%. In SVO sentences, 89.6% of the questions were correctly answered within the SVO-biasing prosody condition and 93.1% within the SVO-neutral prosody condition. In OVS sentences response accuracy was below chance level: 45.8%

(a) (b)

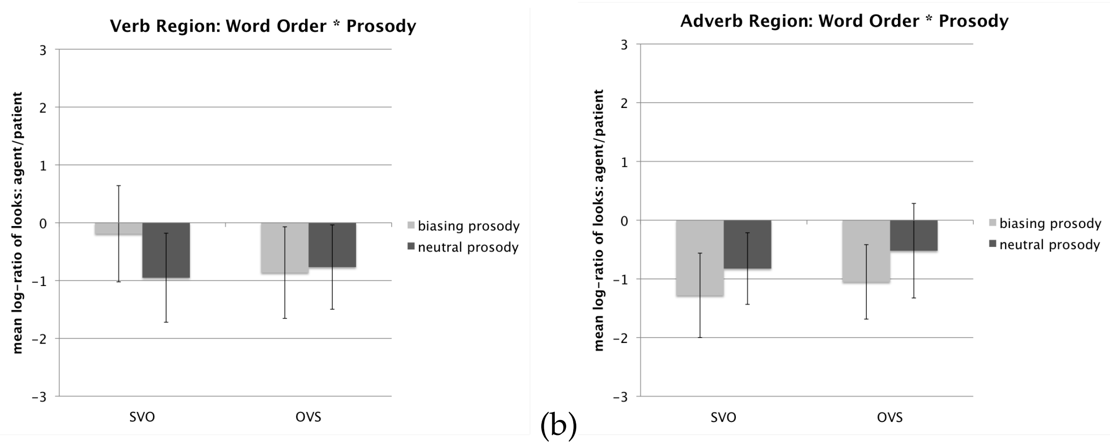


Fig. 7.9: *Experiment 1 children: Mean log-ratios of looks towards the agent (positive numbers) relative to the patient (negative numbers) during the verb (a) and adverb (b) region in all four conditions. The graphs represent the interaction of word order and prosody. Error Bars reflect 95% Confidence Intervals.*

in the biasing prosody condition and 40.3% in the neutral prosody condition (Figure 7.10).

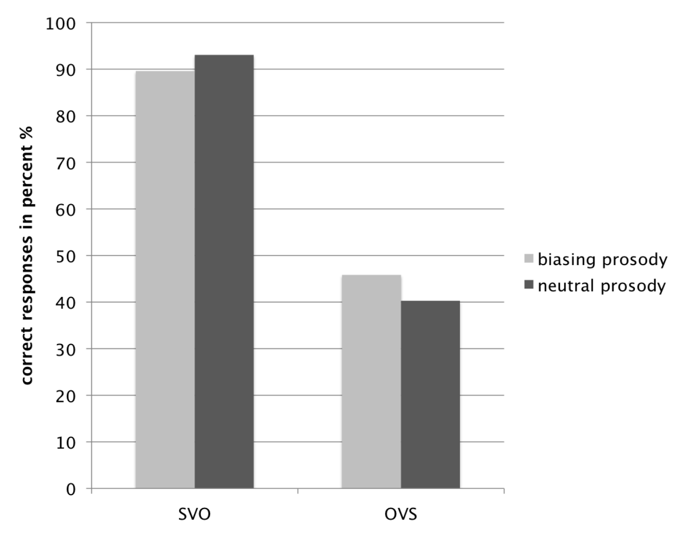


Fig. 7.10: *Experiment 1 children: Correct responses to post-sentence comprehension questions by condition (x-axis) in percent (y-axis).*

Overall, we did not observe a clear difference between the two prosodic conditions. However, Mixed Effects Models results revealed a clear effect of word order (Table 7.3). Response accuracy was higher for SVO than for OVS sentences. We did not observe an effect of voice. The absence of clear effects of voice of the comprehension question in the accuracy data likely resulted from the high number of correct responses in SVO sentences. We thus looked at the descriptive statistics to account for possible effects in OVS (and SVO) sentences. The results again showed very high percentages of correct responses for SVO sentences in both active (90.3%) and passive (92.4%) comprehension questions. The results for OVS sentences, however, revealed slight differences between active (49.3%) and passive (36.8%) voice (Figure 7.11). Children did not seem to struggle to respond correctly to active or passive voice comprehension questions when the sentence structure is SVO (preferred over OVS). When the sentence structure is less preferred (OVS), children struggled more to respond correctly to comprehension questions in passive voice than in active voice.

Table 7.3: Generalised Linear Mixed Effects Model Results: Accuracy Experiment 1 Children. Maximal Converging Model: accuracy ~ wordorderC + prosodyC + (1 | subject) + (1 | item).

	Coeffiecient	**Standard Error**	**z-value**	**p**
Intercept	-.3497	0.2433	-1.437	0.151
word order	2.9369	0.2702	10.871	0.000
prosody	0.0725	0.2201	0.329	0.742

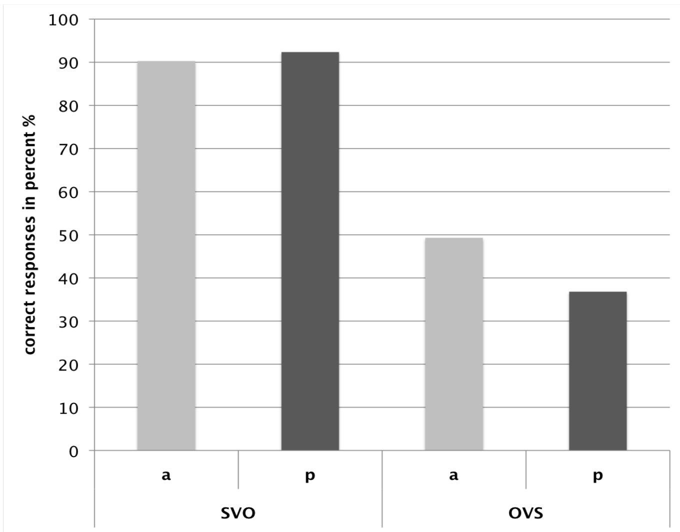


Fig. 7.11: *Experiment 1 children: Correct responses to post-sentence comprehension questions by sentence structure and voice (x-axis) in percent (y-axis). a - active voice comprehension question. p - passive voice comprehension question.*

ACCURACY AND EYE-MOVEMENTS. For the verb and adverb region, we ran two (word order) by two (prosody) repeated-measures ANOVAs including accuracy (high vs. low) as a between-participants factor. The results for both word regions showed no significant interaction of accuracy and word order, accuracy and prosody, or all three factors (verb region - interaction word order and accuracy: $F_1(1, 23) = .204$, $p = .656$; interaction prosody and accuracy: $F_1(1, 23) = .000$, $p = .985$; interaction word order, prosody, and accuracy: $F_1(1, 23) = .005$, $p = .944$; adverb region - interaction word order and accuracy: $F_1(1, 23) = 1.101$, $p = .305$; interaction prosody and accuracy: $F_1(1, 23) = 1.203$, $p = .285$; interaction word order, prosody, and accuracy: $F_1(1, 23) = .882$, $p = .358$. Therefore, differences in response accuracy had no effects on eye-movements.

K-ABC. Children in this study responded more often correctly in the *Word Order* subtask (89.2%) than in the *Spatial Memory* (75.8%) and in the *Number Recall* subtest (75.1%; Figure 7.12; see Appendix G, Table G.1 for single subjects results). To probe for a possible relation between children's cognitive

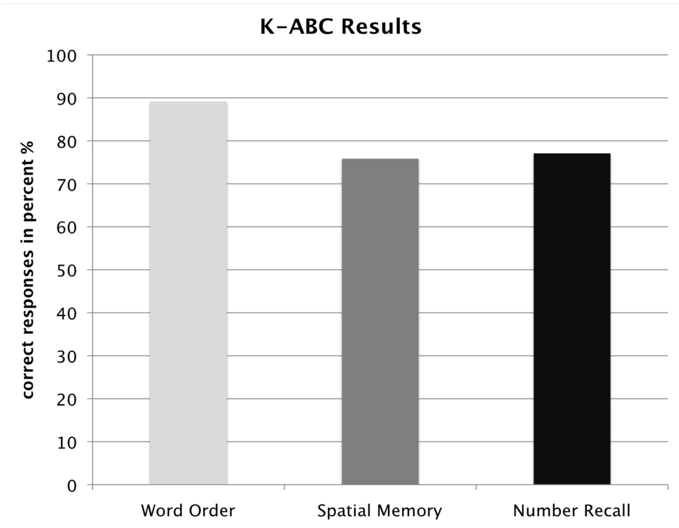


Fig. 7.12: *Experiment 1 children: Results of K-ABC cognitive abilities tests. Overall number of correct responses in percent divided by subtests.*

abilities and their performance in the post-sentence questions, we conducted a Spearman's rank-order correlation (the K-ABC data was not normally distributed). We did not find a correlation between the accuracy data and the K-ABC results (Table 7.4). Following from the fact that there was no correlation between the two data sets, it is very likely that children's cognitive abilities did not influence response accuracies. However, one of the reasons for this might be the low variability in cognitive abilities between participants.

Table 7.4: Experiment 1 children: Spearman's rho correlation: K-ABC - accuracy data

		Accuracy	K-ABC
Accuracy	Spearman's rho	1	.212
	Sig. (2-tailed)		.321
	N	24	24
K-ABC	Spearman's rho	.212	1
	Sig. (2-tailed)	.321	
	N	24	24

Descriptively speaking, we did, however, observe some differences between children with lower vs. higher cognitive abilities. Our findings indicated less correct responses for SVO sentences in children with lower (vs. higher) cognitive abilities (Figure 7.13). Furthermore, it seemed that children with lower cognitive abilities used prosody to an extent in non-canonical OVS sentences (above chance in OVS sentences with a biasing prosody vs. below chance with a neutral prosody).

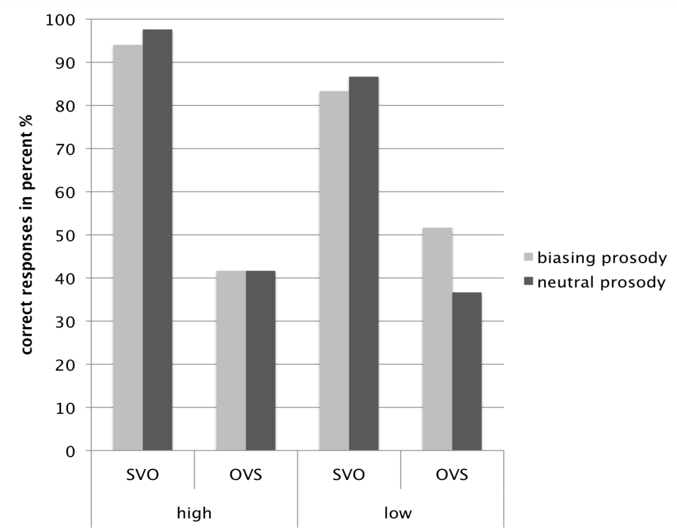


Fig. 7.13: *Experiment 1 children: Correct responses to post-sentence comprehension questions by condition and high/low cognitive abilities (x-axis) in percent (y-axis).*

K-ABC AND EYE-MOVEMENTS. We further ran two (word order) by two (prosody) repeated-measures ANOVAs using K-ABC (high vs. low) as a between-participants factor for verb and adverb region. The data revealed no reliable interaction of children's cognitive abilities (K-ABC scores) and word order, K-ABC scores and prosody, or K-ABC scores, word order, and prosody (verb region - interaction word order and K-ABC: $F_1(1, 23) = 1.205$, $p = .284$, interaction prosody and K-ABC: $F_1(1,23) = .255$, $p = .618$, interaction word order, prosody, and K-ABC: $F_1(1,23) = .475$, $p = .498$; adverb region: interaction word order and K-ABC: $F_1(1,23) = 2.869$, $p = .104$, interaction prosody and K-ABC: $F_1(1,23) = .001$, $p = .977$, interaction word order, prosody, and K-ABC: $F_1(1,23) = .002$, $p = .964$). In other words, children's cognitive abilities did not influence the eye-tracking responses.

7.2.3 *Adults and Children*

We excluded the ambiguous role filler from our analysis because we were interested in anticipatory eye-movements. The ambiguous role filler was always mentioned at the beginning of each sentence. We thus expected participants to inspect the ambiguous role filler during the first noun-phrase. Figure 7.14 shows that both adults and children mostly looked at the ambiguous role filler during NP1 (mean fixation sum is higher than for the other characters). They also inspected the agent and patient in the scene but the overall numbers were very low. In the following, we report analyses that treat participant age as a between-participant variable.

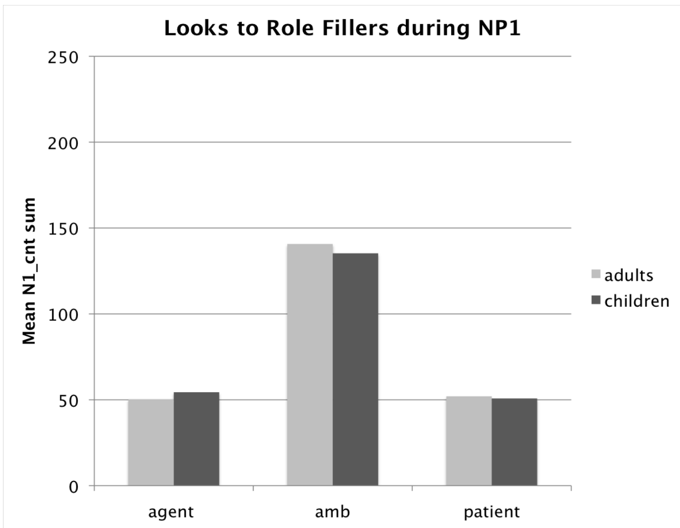


Fig. 7.14: *Experiment 1 adults and children: Mean fixations sum on the three role fillers (agent, ambiguous, patient) during the NP1 region.*

WORD REGIONS. In the verb region (Figure 7.15a), the interaction of word order and age was reliable, $F_1(1,47) = 17.592$, $MSE = 3.775$, $p < .001$, $\eta^2 = .277$, $F_2(1,47) = 25.829$, $MSE = 1.894$, $p < .001$, $\eta^2 = 360$ but the data revealed no interaction of prosody and age, $F_1(1,47) = .296$, $p = .598$, $F_2(1,47) = 2.164$, $p = .148$, or word order, prosody, and age, $F_1(1,47) = 2.309$, $p = .136$, $F_2(1,47) = 1.221$, $p = .275$.

In the adverb region (Figure 7.15b), the interaction of word order and age was also significant, $F_1(1,47) = 20.933$, $MSE = 2.744$, $p < .001$, $\eta^2 = .313$, $F_2(1,47) = 17.935$, $MSE = 3.814$, $p < 001$, $\eta^2 = .281$. We previously reported that adults looked more at the patient (vs. agent) in SVO sentences and more at the agent (vs. patient) in OVS sentences whereas children looked more at the patient (vs. agent) in both SVO and OVS sentences. Thus, the interaction effect of word order and age during verb and adverb region is plausible because children and adults anticipated the agent and patient differently (word order effect in adults and no word order effect in children; Figure 7.15). Further interactions were not significant (interaction prosody and age: $F_1(1,47) = 1.766$, $p = .190$, $F_2(1,47) = .146$, $p = .704$; interaction word order, prosody, and age: $F_1(1,47) = .387$, $p = .537$, $F_2(1,47) = .052$, $p = .821$).

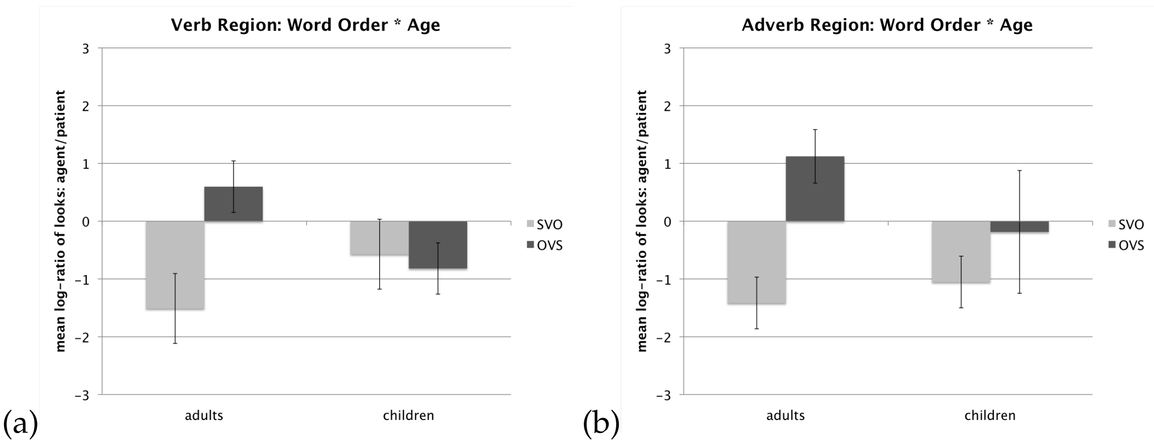


Fig. 7.15: *Experiment 1 adults and children: Mean log-ratios of looks towards the agent (positive values) relative to the patient (negative values) during the verb (a) and adverb (b) region. The graphs represent the interaction of age and word order. Error bars reflect 95% confidence intervals.*

ACCURACY. The results of a Generalised Linear Mixed Effects Model including both age groups corroborated the word-region analyses of the eye-movement responses. The effect of age was significant such that adults' response accuracy was significantly higher than children's which likely resulted from the low number of correct responses for OVS sentences in children (Table 7.5).

Table 7.5: Generalised Linear Mixed Effects Model results: Accuracy Experiment 1 adults vs. children. Maximal Converging Model: accuracy ~ wordorderC + prosodyC + ageC + (1 | subject) + (1 | item).

	Estimate	Standard Error	z-value	p
Intercept	4.92638	0.64548	7.632	0.001
word order	2.89007	0.26641	10.848	0.000
prosody	0.09295	0.21595	0.430	0.667
age	-5.27211	0.67106	-7.856	0.000

7.3 DISCUSSION

Experiment 1 investigated children's and adults' thematic role assignment in a) unambiguously case-marked SVO and OVS sentences and b) in ambiguous action scenes (two role fillers performed identical actions). The scenes did not disambiguate role relations per se. The middle character (NP1) was always role ambiguous and could thus be interpreted as the agent or the patient in the scene. However, depicted role relations could, in principle, permit early thematic role assignment if case-marking and/or prosody help to disambiguate the visual scene. We monitored participants' eye-movements while they inspected clipart scenes and listened to related German SVO and OVS sentences (Table 1). Existing findings suggest that adults rapidly exploit case-marking for thematic role assignment (Kamide, Scheepers, & Altmann, 2003; Knoeferle, 2007; Matzke et al., 2002). Evidence for children's use of case-marking, however, revealed conflicting results, suggesting the importance of visual information (Brandt et al., 2016; Dittmar et al., 2008a; Münster, 2016; Özge et al., 2016; Zhang & Knoeferle, 2012).

Eye-Movements. In line with previous findings, adult participants rapidly exploited case-marking for thematic role assignment. However, unlike in Zhang and Knoeferle (2012) - when no actions were depicted but case-marking was unambiguous - participants started to inspect the target character from the beginning of the verb region (instead of the adverb region in Zhang & Knoeferle, 2012), suggesting that they used the visual scene - although the actions were depicted ambiguously - for early thematic role assignment. We observed more anticipatory looks towards the patient (vs. agent) in SVO sentences and towards the agent (vs. patient) in OVS sentence.

By contrast, children did not rapidly exploit case-marking for such visual anticipation. From the middle of the verb region onwards, they directed more anticipatory looks towards the patient (vs. agent) of the scene during SVO and OVS sentences. It seems that children interpreted OVS sentences as agent-first sentences, despite unambiguous case-marking on the object noun-phrase. Previous research on children's use of case-marking suggests the importance of additional visual information for thematic role assignment (Münster, 2016; Özge et al., 2016; Zhang & Knoeferle, 2012). In these studies, additional information (beyond sentential case-marking) likely provided a supportive background to determine the thematic roles of the linguistic input. Our scenes, however, did not constrain thematic role relations by means of world knowledge (Özge et al., 2016) or action depiction that – once the verb became available – permitted children to distinguish 'who does what to whom' (Münster, 2016; Zhang & Knoeferle, 2012). It is possible that the lack of additional contextual information resulted in children failing to use case-marking in real-time.

Throughout their first years of life, children's comprehension mechanisms develop rapidly. They learn a lot from their immediate environment, among others, by observing who interacts with whom. Perhaps to be able to rapidly exploit case-marking, children need very clear (unambiguous) information from the visual display. Unless depicted unambiguously, they fail to use this information to correctly interpret more demanding structures (e.g., non-canonical OVS sentences) and fall back on more common structures (e.g., SVO sentences). Future research could examine when children start to abstract from the visual display and when they are able to use case-marking for syntactic structuring in an adult-like manner.

Prior research reported evidence for effects of prosody on real-time thematic role assignment in adults (Weber et al., 2006) and offline thematic role assignment in children (Grünloh et al., 2011). Unlike these previous findings, we failed to observe clear effects of prosody in both children and adults. One of the reasons for the null effect of prosody in adults might be that case-marking is used as a stronger/more reliable cue to thematic role assignment than prosody (i.e., case-marking is a cue that affects visual attention/language comprehension more than prosody). When both types of information are available, adults might rely more on case-marking than on prosody. Sedivy et al. (1999) made similar arguments for prosodic marking and object colour contrast. The authors argued that colour contrast enables a strong contrastive interpretation which in turn eliminates further effects of a contrastive intonation. The same argument might hold for case-marking and prosody such that case-marking fully disambiguated role relations without any additional beneficial effects of prosody.

Unlike prior research (Grünloh et al., 2011), we did not observe clear effects of prosody in children. One reason for why our results differed from those by Grünloh and colleagues (2011) might be that we used different measures. They used an act-out task but we used an online measure (eye-movements) in combination with post-sentence comprehension questions. In Grünloh et al. (2011), children saw two video clips followed by static images of the scene which presented thematic roles in a reversed order. Thus, children could directly contrast the opposing role relations. Our scenes, however, depicted actions ambiguously, including both opposing role relations within one scene. This may have added a degree of difficulty to the task, resulting in no clear effects of prosody on thematic role assignment.

We did, however, observe some influence of prosody on eye-movements in both children and adults. Although the results were not reliable, it seems that participants of both age groups are somewhat sensitive to prosody. We observed slightly higher preferences of agent/patient inspection in the basing compared to the neutral prosody conditions.

Accuracy. For adults the response accuracy corroborated the eye-movement data. Anticipatory looks to the patient in SVO sentences resulted in correct responses for SVO sentences and anticipatory looks to the agent in OVS sentences in correct responses for OVS sentences. The overall number of correct responses was very high for both sentence structures.

Responses to post-sentence comprehension questions in children were below chance level (42%), corroborating existing findings by Zhang and Knoeferle (2012) - when no actions were depicted. However, in Dittmar et al. (2008a), correct responses were at chance level in the act-out task (52%) but also below chance level (35%) in the video-pointing task. Taken together, it seems that the understanding of accusative case (i.e., OVS sentence interpretation) is difficult for children at the age of five.

We have to keep in mind, however, that half of the OVS sentences were case-marked on the determiner and via suffixation on the first noun (e.g., *Den Bär-EN*). This additional suffixation resulted in slightly more correct responses than the determiner-only case-marking (38.9% vs. 45.9%). It seems that the more information children have (visual or linguistic), the better can they use case-marking for thematic role assignment.

7.4 SUMMARY AND CONCLUSION

Overall, we did not observe significant effects of prosody in either adults or children. Adults used case-marking rather than prosody for rapid thematic role assignment. At the beginning of the verb region, adults started to inspect the patient in SVO sentences and the agent in OVS sentences, regardless of a biasing or neutral prosody (Figure 7.6a). Just like we predicted, adults correctly responded to post-sentence comprehension questions in almost 100% of all questions (Figure 7.7). Not surprisingly, children did not show the same pattern in the behavioural data. Upon hearing the verb, children inspected the patient (vs. agent) in the scene in all four conditions (Figure 7.9a). Prosody and case-marking could not overwrite their strong SVO word order bias. Perhaps children have a strong visual bias. When looking at the ambiguous character, its posture may have directed children's attention to the patient (vs. agent). In other words, the visual bias may have supported garden-pathing in OVS sentences in children.

8

EXPERIMENT 2

Experiment 2a followed suit and examined whether case-marking is, indeed, a stronger/more reliable cue for thematic role assignment/has more pronounced effects on visual attention/language comprehension than prosody. Since children's ability to use case-marking seems to be limited to visual contexts that provide world knowledge (Özge et al., 2016 but Dittmar et al., 2008a; Schipke et al., 2011; and our results of Experiment 1), participants in Experiment 2 were exclusively adults. In Experiment 1, we could not answer whether the absence of prosodic effects generalises to a stronger contrast in prosody between SVO and OVS sentences. In Experiment 1, we mentioned that Sedivy et al. (1999) argued that colour contrast evoked a strong contrastive interpretation which led to no further effects of contrastive prosody. Our findings from Experiment 1 and Sedivy and colleagues' (1999) argument support the idea that other cues such as lexical and morpho-syntactical cues are used before prosodic cues. One question, however, remains: Do adults use prosody when morpho-syntactic cues are not available?

Experiment 2b was thus designed to investigate whether adult participants use prosody when case-marking is ambiguous in ambiguous action scenes that are free from world knowledge. We aimed to replicate the findings from Experiment 2a (unambiguous OVS sentences) in a first condition pair. A second pair tested ambiguous OVS sentences. Existing findings suggest that prosody can rapidly be exploited to disambiguate syntactic structure (Weber et al., 2006). In a visual world eye-tracking study, participants rapidly exploited prosody to identify grammatical functions when the scene depicted role fillers such as a cat, a bird, and a dog, for which world knowledge implicated thematic relations (e.g., cats chase birds and dogs chase cats). But the scenes did otherwise not disambiguate the upcoming thematic role relations. Feminine case-marking (identical in nominative and accusative case) on the determiner of the first noun-phrase created locally structurally ambiguous sentences: *Die Katze (amb.) jagt womöglich den Vogel (acc/obj)/der Hund (nom/-subj)*– 'The cat (amb.) chases possibly the bird (obj/patient)/the dog (subj/agent)'. Listeners made more anticipatory eye-movements towards the patient (vs. agent) in the scene for ambiguous SVO sentences. In ambiguous OVS sentences, participants did not inspect the agent significantly more than the patient. Prosody could, however, overwrite participants' strong SVO word order bias. The results for ambiguous OVS sentences revealed a similar number of looks towards the agent as towards the patient. Prosody was the only available information to correctly assign thematic roles prior to the disambiguating case-marking on the second noun-phrase. At the same time, world knowledge associated with the scene may have provided a supportive background.

8.1 MATERIALS AND DESIGN

PARTICIPANTS. 48 monolingual native speakers of German with normal or corrected vision and hearing participated in the experiments. 24 young adults (mean age = 24.3, SD = 3.1) took part in Experiment 2a and 24 (mean age = 23.1, SD = 3.62) in Experiment 2b. They were all students from the Bielefeld University who were paid for participation and gave written informed consent before taking part.

SPEECH STIMULI. The same linguistically trained female native speaker of German recorded unambiguous SVO and OVS sentences for Experiment 2a. For Experiment 2b she additionally recorded OVS sentences that were ambiguously case-marked (feminine or neuter gender) on the first noun-phrase. She was instructed to use the same prosodic contours described in Experiment 1, with one difference: The neutral prosodic condition was exchanged by SVO- or OVS-biasing prosodic contours (*non-biasing*). Thus, all sentences in this experiment were assigned either an SVO-biasing or and OVS-biasing prosodic contour resulting in four conditions for each experiment. Experiment 2a consisted of conditions a) SVO-SVO biasing prosody, b) OVS-OVS biasing prosody, c) SVO-OVS biasing prosody, and d) OVS-SVO biasing prosody. Experiment 2b included conditions a) OVS-OVS biasing prosody, b) ambiguous OVS-OVS biasing prosody, c) OVS-SVO biasing prosody, and d) ambiguous OVS-SVO biasing prosody (Table 8.1). We did not use the speech stimuli recorded for Experiment 1 because the conditions were different in Experiment 2 (Experiment 1: biasing vs. neutral; Experiment 2: biasing vs. non-biasing) and the realisation of the prosodic structures/the quality of the recording should be as similar as possible (i.e., if we used the recordings from Experiment 1, loudness, for example, might have been different between the two recordings).

Table 8.1: Conditions Experiment 2

Experiment	Condition	Sentence Structure	Prosodic contour	Bias
2a	a	SVO	L*(NP1)+H, H*(verb)	SVO
	b	OVS	L+H*(NP1)	OVS
	c	SVO	L+H*(NP1)	OVS
	d	OVS	L*(NP1)+H, H*(verb)	SVO
2b	a	OVS	L+H*(NP1)	OVS
	b	ambOVS	L+H*(NP1)	OVS
	c	OVS	L*(NP1)+H, H*(verb)	SVO
	d	ambOVS	L*(NP1)+H, H*(verb)	SVO

visual stimuli. Since Experiment 2 is a follow up study of Experiment 1 (adults), we used the images described in Experiment 1 but exchanged the middle character of 12 scenes and 12 sentences respectively (items 1, 3, 5 8, 10, 12, 15, 16, 17, 21, 22, 24). For these 12 unambiguous OVS sentences, accusative case was not only marked on the determiner of the first noun-phrase but additionally marked via suffixation (-en) at the end of the first noun (eg. *der Elefant - den Elefanten*). To avoid any effects of this double case-marking (on the determiner and via suffixation on the noun), which could in theory influence the eye-tracking and comprehension data, we excluded the cases in which double case-marking occurred. Overall, we exchanged six animals in the scenes (*Bär* 'bear', *Affe* 'monkey', *Hase* 'hare', *Eisbär* 'polarbear', *Elefant* 'elephant', *Löwe* 'lion' - each animal occurred twice) with animals that did not prompt additional suffixation in accusative case (*Käfer* 'bug', *Biber* 'beaver', *Storch* 'stork', *Panda* 'panda, *Seehund* 'sealion ', *Kraken* 'kraken') but could still be transformed to feminine gender.

design. The overall design was similar to the design of Experiment 1. However, we added 72 new filler trials to the experiment. Among these, we varied the number of role fillers mentioned in the scene (one, two, or three), the word order (SVOO: *Die Eule strickt dem Arzt einen Schal* - 'The owl knits the doctor a scarf.'; directOVSO: *Der Ballerina saugt der Widder das Wohnzimmer* - 'The ballerina hoovers the ram the living room.'; indirectOVSO: *Eine Blume übergibt der Handwerker dem Flamingo* - 'A flower delivers the craftsman to the flamingo'), and the type of adverb (manner, frequency, place). All variations contained eight filler items each. Within this group of eight, four filler items were assigned an SVO-biasing prosody and four an OVS-biasing prosody. For each experiment, we pseudo-randomised 24 lists allowing for each condition to occur twice sequentially and inserting at least two filler trials in between experimental trials. After the practice trials, at least three filler trials were inserted before the first experimental trial. For Experiment 2b, we changed the item balancing such that each NP1 animal character (which occurred twice across items) occurred once with an unambiguous and once with an ambiguous OVS sentence, once with an SVO- and once with an OVS-biasing prosody, once with a left and once with a right balancing, and once with an active and once with a passive voice comprehension question.

PREDICTIONS. Since we know from previous studies (Kamide, Scheepers, & Altmann, 2003; Matzke et al., 2002) and from Experiment 1 that adults can rapidly recruit case-marking for thematic role assignment, we expected participants to exploit case-marking in Experiment 2a. Thus, the results should reveal more looks towards the patient (vs. agent) in SVO sentences and more looks towards the agent (vs. patient) in OVS sentences. If they additionally exploit prosody, then we should see a difference between the two prosodic conditions in the amount and timing of looks towards the target role filler. Prosody was not expected to influence adults' offline performance in the presence of unambiguous case-marking.

For Experiment 2b, we expected participants to rapidly exploit case-marking in unambiguous OVS sentences. If they do not rapidly exploit prosody in addition to case-marking, we should see more looks towards the agent (vs. patient), in the absence of a clear difference between the two prosodic conditions. If participants do, however, rapidly exploit case-marking and prosody, we may see higher and possibly earlier agent preferences in the biasing compared to the non-biasing prosody conditions. For OVS sentences with ambiguous case-marking, we expected to find more and earlier looks towards the agent (vs. patient) in the biasing prosody condition and more looks to the patient (vs. agent) in the non-biasing prosody condition (SVO interpretation). That is, if adults rapidly exploit prosody when case-marking is ambiguous. If they do not rapidly exploit prosody for thematic role assignment, we should observe more looks to the patient (vs. agent) in both prosodic conditions. We did not expect prosody to influence the number of correct post-trial comprehension responses because case-marking was either available on the first noun-phrase or on the second noun-phrase.

We expected to find effects of case-marking and/or prosody at the end of the verb, beginning of the adverb region. This is the point in time that distinguishes the two prosodic structures: The additional stress on the verb can be distinguished from the main stress on the first noun-phrase (see Weber et al., 2006). However, case-marking effects may emerge even earlier, since the first noun-phrase is marked for case and may influence participants looks towards the target character from the beginning of the verb region.

PROCEDURE AND ANALYSIS. The procedure and analyses were identical to Experiment 1. The only differences were the duration of the experiment, the on-and offsets (Appendix A.2, A.3), the mean on- and offsets for the time course graphs and the longest durations of each word region (Table 8.2, Table 8.3). Each testing session lasted about 40 minutes.

Table 8.2: Mean Onsets and Longest Word Durations Experiment 2a (in ms)

	NP1	Verb	Adverb	NP2
Mean onsets	0	1421	2658	3944
Longest duration	1630	1469	1625	1643 (1143 +500)

Table 8.3: Mean Onsets and Longest Word Durations Experiment 2b (in ms)

	NP1	Verb	Adverb	NP2
Mean onsets	0	1458	2696	3982
Longest duration	1673	1469	1625	1515 (1015+500)

8.2 RESULTS AND DISCUSSION

8.2.1 *Experiment 2a*

TIME COURSE GRAPHS. During the NP1 region, we observed a numerical trend in the looks towards the agent (vs. patient) in all four conditions. In SVO sentences, we observed a difference in the amount of looks towards the agent (vs. patient) for the SVO-SVO biasing prosody condition compared to the SVO-OVS biasing prosody condition. Participants looked more towards the agent (vs. patient) when there was no nuclear stress on the first noun-phrase compared to when the first noun-phrase was assigned a nuclear stress. In unambiguous OVS sentences, no difference between the two prosodic conditions emerged (OVS-OVS biasing prosody and OVS-SVO biasing prosody). Towards the middle of the verb region, participants started to inspect the patient (vs. agent) more in SVO sentences and the agent (vs. patient) more in

OVS sentences in the absence of a clear difference between the two prosodic conditions. During the adverb region, the preference to look at the target character was similar to the verb region (looks to the patient in SVO sentences and the agent in OVS sentences).

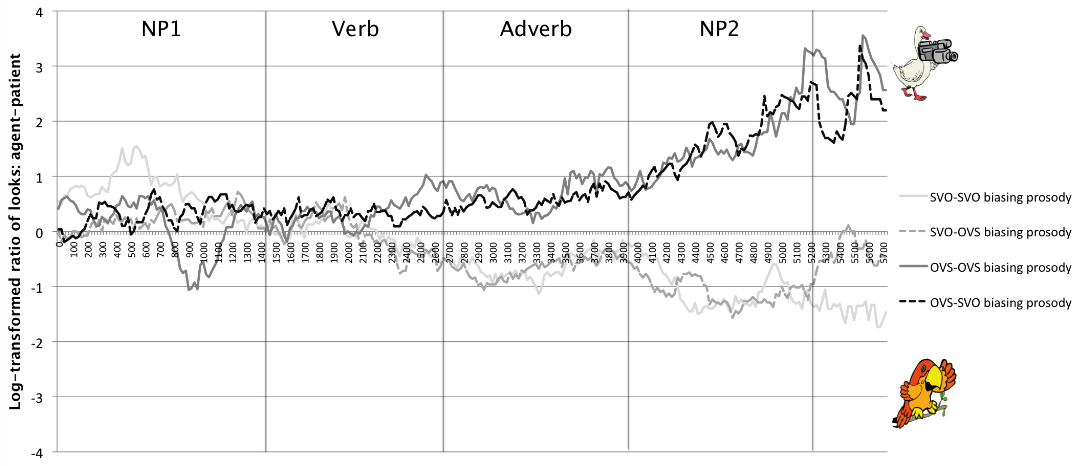


Fig. 8.1: *Experiment 2a: Time course of eye-movements including the visual bias (y-axis) towards agent (positive values) relative to patient (negative values) in all conditions (ln(agent/patient)) over time in ms (x-axis). The vertical lines represent mean onsets for each word region.*

Towards the end of the adverb region this preference decreased for SVO sentences. The bias to look at the patient (vs. agent) was very small. We argued in Experiment 1 that this might result from the materials. The agent and the ambiguous character were both depicted as performing an action. This action is mediated by the verb and might influence participants' looks towards the other character performing an action - the agent. We further suggested that participants might verify their interpretation since both SVO and OVS interpretations are possible within one scene. During the NP2 region, we observed more looks towards the patient in SVO sentences and the agent in OVS sentences. Clear differences between the two prosodic conditions were absent.

However, the number of target inspections was overall higher in OVS sentences compared to SVO sentences - even more than in the adult data of Experiment 1. Participants may have inspected the agent (vs. patient) more in OVS sentences because OVS is generally less preferred than SVO (Dittmar et al., 2008a; Hemforth, 1993; Hemforth & Konieczny, 2013). Perhaps participants wanted to verify their interpretation resulting in more looks towards the agent. Another possible explanation is that the agent performed an action whereas the patient did not. Maybe more inspections fell on the agent simply because he performed an action and is thus slightly more salient than the patient.

WORD REGIONS. As early as the NP1 region (Figure 8.2a), the effect of word order was significant in the by participants analysis, $F_1(1,23) = 4.416$, $MSE = 2.515$, $p < .05$, $\eta^2 = .161$ but not in the by item analysis $F_2(1,23) = .530$, $p = .474$. Additionally, the data revealed a marginal effect of prosody in the by participants analysis but not in the by items analysis, $F_1(1,23) = 3.221$, $MSE = 1.157$, $p = .086$, $\eta^2 = .123$, $F_2(1,23) = 1.982$, $p = .173$. The interaction of prosody and word order was not reliable $F_1(1,23) = .145$, $MSE = 2.106$, $p = .707$, $\eta^2 = .006$, $F_2(1,23) = 1.140$, $p = .297$.

The marginal effect of prosody likely resulted from the number of looks to the patient (vs. agent) in condition b) OVS-OVS biasing prosody and the agent (vs. patient) in condition a) SVO-SVO biasing prosody (see Figure 8.1). Several factors might come into play here: a) the absence of a nuclear stress in SVO sentences during the NP1 region might have triggered looks towards the agent (vs. patient), whereas the presence of a nuclear stress in OVS sentences might have influenced looks to the patient (vs. agent). It is also possible that the overall low number of looks at the agent and patient during NP1 (Figure 9.16) and a possible high variability in the use of prosody between participants might have led to the described pattern.

In the verb region (Figure 8.2b), the data revealed no significant effects of prosody, $F_1(1,23) = .068$, $p = .797$, $F_2(1,23) = .057$, $p = .813$. However, the effect of word order was significant, $F_1(1,23) = 9.404$, $MSE = 3.234$, $p < .05$, $\eta^2 = .290$, $F_2(1,23) = 5.990$, $MSE = 1.961$, $p < .05$, $\eta^2 = .207$ (Figure 8.2). Participants looked more at the patient (vs. agent) in SVO sentences and more at the agent (vs. patient) in OVS sentences. The interaction between word order and prosody was not significant, $F_1(1,23) = .144$, $p = .708$, $F_2(1,23) = .155$, $p = .698$.

In the adverb region, effects of prosody were not reliable. There was no interaction of word order and prosody (prosody: $F_1(1,23) = .001$, $p = .981$, $F_2(1,23) = .078$, $p = .783$; interaction word order and prosody: $F_1(1,23) = .065$, $p = .801$, $F_2(1,23) = .237$, $p = .631$. However, the effect of word order was significant, $F_1(1,23) = 20.740$, $MSE = 3.498$, $p < .001$, $\eta^2 = .474$, $F_2(1,23) = 19.684$, $MSE = 2.836$, $p < .001$, $\eta^2 = .461$. Again, participants inspected the target role filler depending on case-marking.

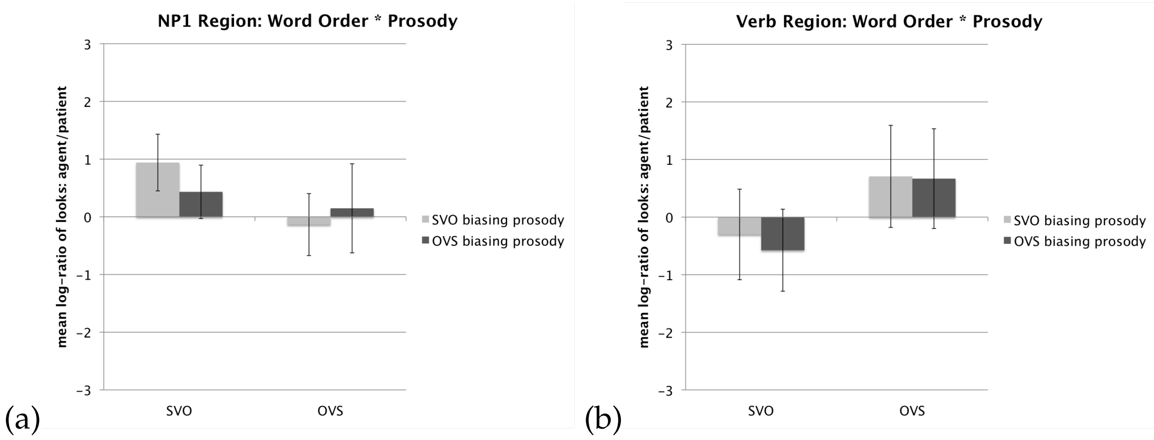


Fig. 8.2: *Experiment 2a: Mean log-ratios of looks towards the agent (positive numbers) relative to the patient (negative numbers) during the verb (a) and adverb (b) region in all four conditions. The graphs represent the interaction of word order and prosody. Error bars reflect 95% confidence intervals.*

We described earlier that using the two biasing prosodic contours created a very clear contrast between the two. Using these two contours respectively, might, however, also have led to false expectations during the course of the experiment: the SVO-biasing prosody did not only occur with SVO sentences but also with OVS sentences in the same way that the OVS-biasing prosody occurred with OVS and SVO sentences. Participants might have recognised that the prosodic structure sometimes mismatched the sentence structure. If this was the case, then the recognition may have influenced the eye-tracking data. Therefore, we ran a block-wise two (word order) by two (prosody) ANOVA including time (first half vs. second half) as a within participants factor. The results revealed no interaction effects of block and word order, block and prosody, or block, word order, and prosody during the verb region (interaction word order and block: $F_1(1,23) = .494$, $p = .489$, $F_2(1,23) = .002$, $p = .968$; interaction prosody and block: $F_1(1,23) = .999$, $p = .328$, $F_2(1,23) = 1.016$, $p = .324$; interaction word order, prosody, and block: $F_1(1,23) = 1.959$, $p = .175$, $F_2(1,23) = .109$, $p = .744$.

During the adverb region, however, the interaction of word order, prosody and block was marginally significant in the by participants analysis but not in the by items analysis, $F_1(1,23) = 4.114$, $MSE = 5.667$, $p = .054$, $\eta^2 = .152$ $F_2(1,23) = .050$, $p = .825$ (Figure 8.3). It seems that in the first block participants inspected the respectively appropriate target character (patient in SVO sentences and agent in OVS sentences) more in sentences with an OVS-biasing prosody whereas in the second block participants inspected the respectively appropriate target character more in the SVO-biasing prosody conditions. However, block was not part of our design. Therefore, this subtle interaction which only emerged in the by participants analysis may have been influenced by an uneven distribution of conditions between the first and the second half of the experiment (due to randomisation). Further interactions were not reliable (interaction word order and block: $F_1(1, 23) = .787$, $p = .384$, $F_2(1, 23) = .117$, $p = .735$; prosody and block: $F_1(1, 23) = .205$, $p = .655$, $F_2(1, 23) = .400$, $p = .534$).

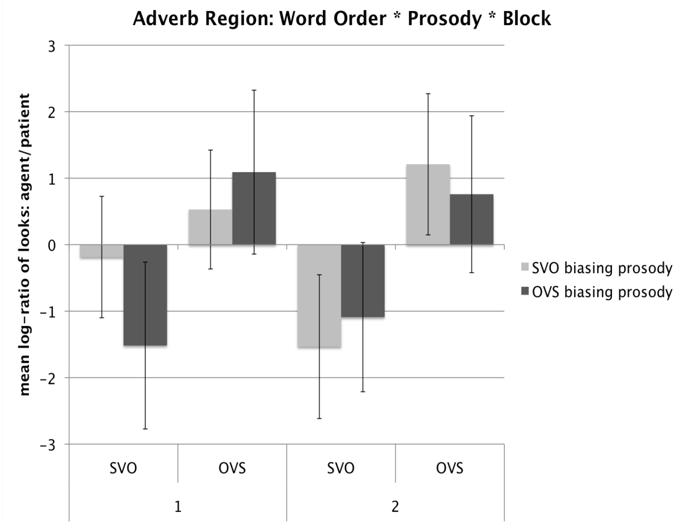


Fig. 8.3: *Experiment 2a: Mean log-ratios of looks towards the agent (positive numbers) relative to the patient (negative numbers) during the adverb region in all four conditions divided by block (1 - first half, 2 - second half). The graph represents the interaction of word order, prosody, and block. Error bars reflect 95% confidence intervals.*

The data revealed no clear effects of prosody although we increased the contrast between the two prosodic conditions but we nevertheless observed some differences between the two prosodic conditions. Thus, we zoomed in on some participants to account for possible individual differences which may have cancelled out any clear effects of prosody. In Figure 8.4, we can see that participant 1, for example, showed a high preference to look at the patient (vs. agent) in SVO sentences with an OVS-biasing prosody, and no preference to look at the agent or patient in SVO sentences with an SVO-biasing prosody. Participant 7, however, looked more at the patient (vs. agent) in SVO sentences with an SVO-biasing prosody and more at the agent (vs. patient) with an OVS-biasing prosody. It is possible that these individual differences in target preference influenced our results.

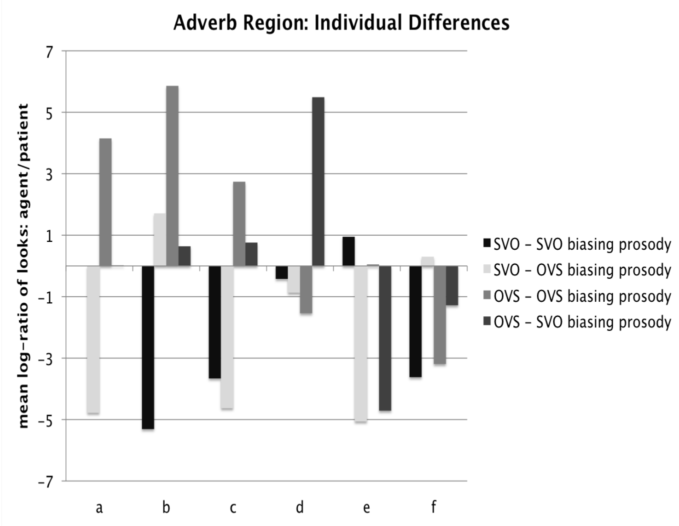


Fig. 8.4: *Experiment 2a: Individual differences in the mean log-ratios of looks towards the agent (positive numbers) relative to the patient (negative numbers) during the adverb region in all four conditions. Alphabetic letters represent one participant each.*

ACCURACY. The data revealed an overall response accuracy of 94.6%. In condition a) SVO-SVO biasing prosody, participants correctly responded to all comprehension questions (100%). Condition c) SVO-OVS biasing prosody revealed 97.9% correct answers. In condition b) OVS-OVS biasing prosody, the percentage of correct responses was 91.7% and in condition d) OVS-SVO biasing prosody 88.9% (Figure 8.5). Overall, the number of correct responses in SVO sentences was higher than in OVS sentences, although OVS sentences were unambiguously case-marked.

Compared to the adults group in Experiment 1, the number of correct responses in OVS sentences was slightly lower in this experiment, especially in OVS sentences with an SVO-biasing prosody (Figures 7.7 and 8.5). In Experiment 1, this condition did not exist but we used a neutral prosody as a baseline. Perhaps the non-biasing prosody (SVO-biasing in OVS sentences) influenced the number of correct responses more than the neutral prosody because it biased a different sentence structure (SVO instead of OVS). For OVS sentences, we observed a difference of 7.1% between neutral-prosody and SVO-biasing prosody in the descriptive data for OVS sentences. However, we cannot exclude the possibility that this difference simply resulted from the differences in the materials. In Experiment 2a, the number of fillers was 72 whereas in Experiment 1 it was 8. Additionally, many filler sentences in Experiment 2a had an SVO sentence structure which may have influenced the number of correct responses for OVS sentences.

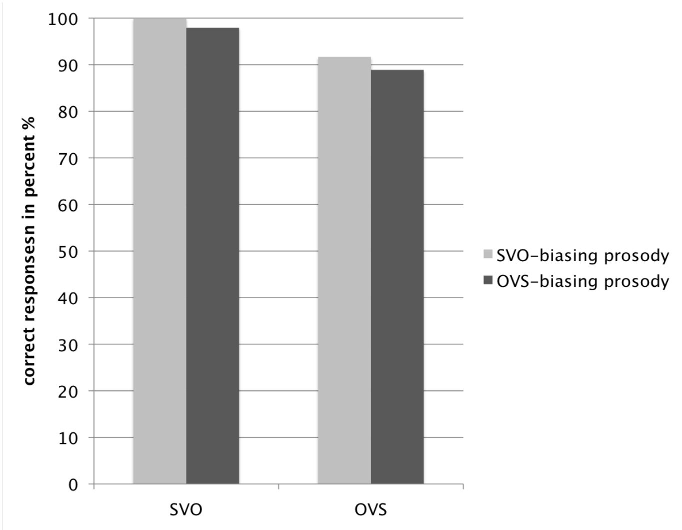


Fig. 8.5: *Experiment 2a: Correct responses to post-sentence comprehension questions by condition (x-axis) in percent (y-axis).*

The results of the Generalised Linear Mixed Effects Model (Table 8.4) revealed a significant effect of word order such that SVO sentence structure resulted in higher accuracies compared to OVS sentence structure. Prosody had no reliable effects on accuracies.

Table 8.4: Generalised Linear Mixed Effects Model Results: Accuracy Experiment 2a. Maximal Converging Model: accuracy ~ wordorderC * prosodyC + (1 | subject) + (1 | item).

	Coeffiecient	**Standard Error**	**z-value**	**p**
Intercept	2.6102	0.3720	7.017	0.000
word order	1.4882	0.6635	2.243	0.0249
prosody	-0.3338	0.4112	-0.812	0.4168
wordorder*prosody	18.4658	186.9560	0.099	0.9213

8.2.2 *Experiment 2b*

TIME COURSE GRAPHS. During the NP1 region, no preference to look towards the agent or the patient was observed, although in condition c) OVS-SVO biasing prosody participants looked slightly more towards the agent than the patient. Towards the end of the NP1 region and the beginning of the verb region, participants inspected the agent (vs. patient) more for sentences with an SVO-biasing prosodic contour (no main stress on NP1) compared to sentences with an OVS-biasing prosodic contour (main stress on NP1).

During the verb region, participants inspected the agent more than the patient when sentences were case-marked (OVS sentences) compared to when case-marking was ambiguous - more looks towards the patient (vs. agent) in conditions b) aOVS-OVS biasing prosody and d) aOVS-SVO biasing prosody. In the adverb region, participants anticipated the agent (vs. patient) in conditions a) and c) (unambiguous OVS) whereas in conditions b) and d) - ambiguous OVS - they anticipated the patient more than the agent, however, this preference was not prominent. Upon hearing the second noun-phrase, a clear preference to look at the agent (vs. patient) was observed in all conditions.

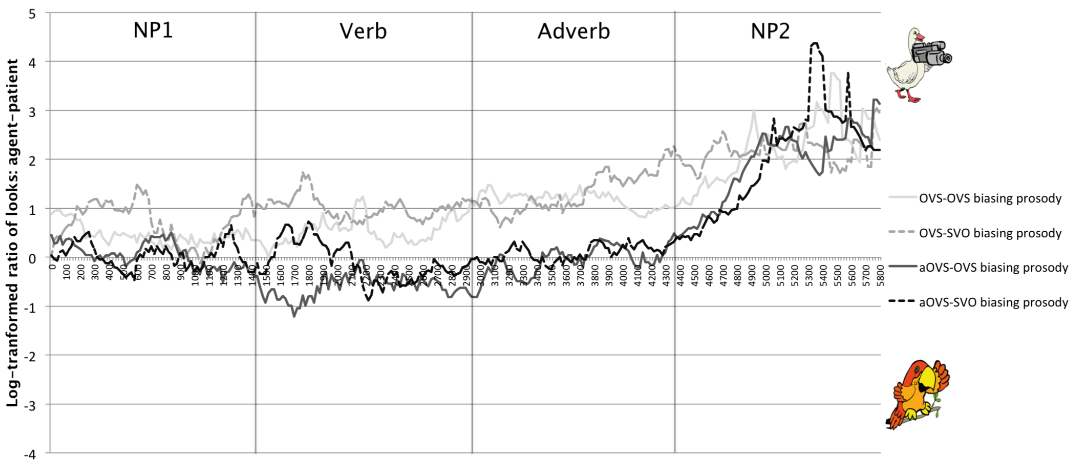


Fig. 8.6: *Experiment 2b: Time course of eye-movements including the visual bias (y-axis) towards agent (positive values) relative to patient (negative values) in all conditions (ln(agent/patient)) over time in ms (x-axis). The vertical lines represent mean onsets for each word region.*

WORD REGIONS. In the verb region (Figure 8.7a), the effect of word order was significant, $F_1(1,23) = 15.560$, MSE $= 4.798$, $p < .01$, $\eta^2 = .965$, $F_2(1,23) = 10.640$, MSE $= 6.069$, $p < .01$, $\eta^2 = .878$. Participants inspected the agent (vs. patient) more in unambiguous OVS sentences. In ambiguous OVS sentences, we observed more looks to the patient (vs. agent; Figure 8.7). Furthermore, the results revealed no significant effects of prosody in the by participants analysis, $F_1(1,23) = .903$, $p = 352$, however, the by items analysis revealed a significant effect of prosody, $F_2(1,23) = 10.774$, MSE $= 2.133$, $p < .01$, $\eta^2 = .882$. The interaction of word order and prosody was not reliable, $F_1(1,23) = .012$, $p = 913$, $F_2(1,23) = .708$, $p = 409$.

In the adverb region (Figure 8.7b), the effect of word order was also significant, $F_1(1,23) = 21.945$, MSE $= 3.602$, $p < .01$, $\eta^2 = .994$, $F_2(1,23) = 22.289$, MSE $= 2.626$, $p < .01$, $\eta^2 = .995$. We observed more looks to the agent (vs. patient) in unambiguous OVS sentences and more looks to the patient (vs. agent) in ambiguous OVS sentences. In ambiguous OVS sentences with an SVO-biasing prosody, however, participants exhibited a slight preference to look at the agent rather than the patient (Figure 8.7b). The data did not reveal reliable effects of prosody or an interaction of word order and prosody (prosody: $F_1(1,23) = .056$, $p = .759$, $F_2(1,23) = 1.474$, $p = .237$; interaction word order and prosody: $F_1(1,23) = .646$, $p = .430$, $F_2(1,23) = 1.555$, $p = .225$).

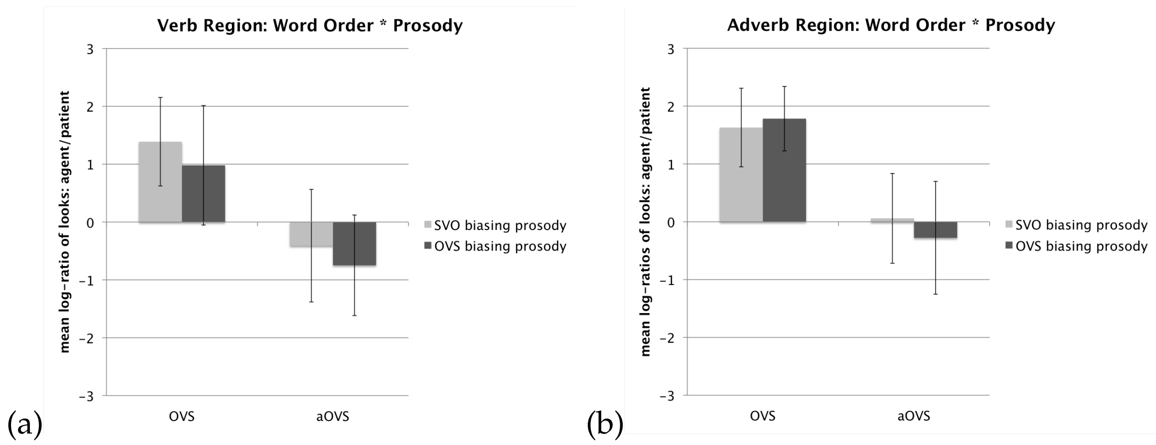


Fig. 8.7: *Experiment 2b: Mean log-ratios of looks towards the agent (positive numbers) relative to the patient (negative numbers) during the verb (a) and adverb (b) region in all four conditions. The graphs represent the interaction of word order and prosody. Error bars reflect 95% confidence intervals. OVS - unambiguously case-marked object-verb-subject sentence structure. aOVS - ambiguously case-marked object-verb-subject sentence structure.*

ACCURACY. Responses to post-sentence comprehension questions revealed an overall accuracy of 94.4%. However, some differences emerged between conditions: a) OVS-OVS biasing prosody 96.5%, c) OVS-SVO biasing prosody 96.5%, b) ambiguous OVS-OVS biasing prosody 93.1%, and d) ambiguous OVS-SVO biasing prosody 91.7% (Figure 8.8). Overall, we observed a slight difference in the number of correct responses between unambiguous and ambiguous OVS sentences. Participants' response accuracy was lower in ambiguous OVS sentences compared to unambiguous OVS sentences. Perhaps, this difference in the number of correct responses resulted from the fact that case-marking was available from the beginning of unambiguous OVS sentences, whereas case-marking disambiguated sentence structure at the end of (un)ambiguous OVS sentences (NP2). Additionally, the number of SVO sentences in the fillers may have led participants towards an SVO interpretation when case-marking was ambiguous.

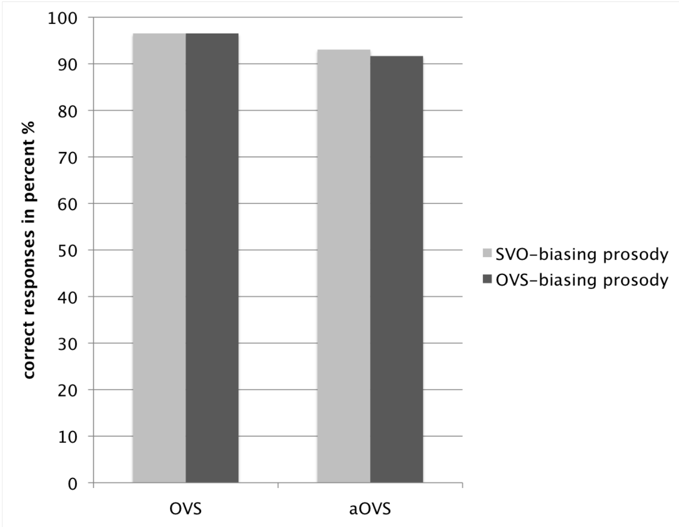


Fig. 8.8: *Experiment 2b: Correct responses to post-sentence comprehension questions by condition (x-axis) in percent (y-axis). OVS - object-verb-subject sentence structure. aOVS - ambiguously case-marked object-verb-subject sentence structure.*

Generalised Linear Mixed Effects Model results corroborated the observed difference in response accuracies between unambiguous and ambiguous OVS sentences: The effect of word order was significant (Table 8.5). Participants' response accuracy was higher in unambiguously case-marked than in ambiguously case-marked OVS sentences.

Table 8.5: Generalised Linear Mixed Effects Model Results: Accuracy Experiment 2b. Maximal Converging Model: accuracy ~ wordorderC + prosodyC + (1 | subject) + (1 | item).

	Coeffiecient	**Standard Error**	**z-value**	**p**
Intercept	4.6115	0.6845	6.737	0.000
word order	-1.0154	0.4346	-2.336	0.0195
prosody	-0.1650	0.4068	-0.406	0.6849

8.3 DISCUSSION

In Experiment 2b, we further investigated whether case-marking affects adult's visual attention/language comprehension more than prosody (2a) and whether adults are able to use prosody when case-marking is ambiguous (2b) in scenes that ambiguously depict role relations. Case-marking and prosody can help comprehenders to rapidly assign thematic roles (Kamide, Scheepers, & Altmann, 2003; Matzke et al., 2002; Weber et al., 2006).

Eye-Movements. Similar to the findings of Experiment 1, in which adults rapidly exploited case-marking rather than prosody for early target anticipation, we did not observe reliable effects of prosody. Therefore, these results strengthen the argument that morpho-syntactic information (case-marking) does have stronger links or is used as a stronger/more reliable cue for thematic role assignment than supra-sentential information (prosody) and thus the idea that case-marking influences visual attention/language comprehension more than prosody.

Previous research has shown that prosody can be used as a cue to resolve syntactic closure ambiguities (Kjelgaard & Speer, 1999) and more importantly grammatical function assignment ambiguities (Weber et al., 2006). The results of Experiment 2b (ambiguous OVS sentences) corroborated Weber et al.'s findings. Although we did not observe clear effects of prosody, the stress on the verb in ambiguous OVS sentences slightly influenced participants' agent anticipation. We found minimal evidence of anticipatory looks to the agent (vs. patient) in condition d) aOVS-SVO biasing prosody whereas participants directed more looks to the patient (vs. agent) when a main stress on the first noun-phrase was assigned.

In unambiguous OVS sentences, we did not observe a similar gaze-pattern as in ambiguous OVS sentences (the SVO-biasing prosody did not elicit looks to the agent). Case-marking likely influenced thematic role assignment: More looks to the agent (vs. patient). One question, however, remains: Why did we find only subtle trends in the use of prosody (eg., in the time course graphs, in the by items analyses)? Although this issue can probably be addressed for most eye-tracking studies, individual differences in the use of prosody for comprehension likely influenced the results. For example, existing findings suggest that individual differences in the production of contrastive focus exist (i.e., there is no one to one mapping of contrastive stress to pragmatic function; Grice, Ritter, Niemann, & Roettger, 2017). These differences in language production might also influence language comprehension.

Prior research reported that speakers realised stress placement differently: Among others, they found different realisations of stress for contrastive stress placement (no one-to one mapping of stress and focus type; Grice et al. (2017)). Weber et al. (2006) also argued that a main stress on the first noun-phrase is not exclusively related to OVS sentence intonation or OVS sentence interpretation. One of the reasons why we observed that the presumed SVO-biasing prosody (instead of the OVS-biasing prosody) influenced participants' slight SVO preference might again be related to the ambiguity of our visual scenes. Two animal characters always performed identical actions (eg., for the verb *filmen* 'film' they were holding a camera). Thus, both animal characters could, in principle, act as agents in the scene. Keeping in mind that the middle character (role ambiguous character) was always mentioned at the beginning of the sentence and thus pre-verbal, a main stress on the verb might have directed participants' eye-movements towards the other character holding an object (the agent). Thus, the ambiguity in the visual display likely influenced the function of the prosodic contour. Furthermore, the visual scenes in Weber et al. (2006) did not include depicted actions and thus the main stress on the verb provided new information for an SVO interpretation which was not the case in our experiments.

Accuracy. Accuracies were high in both unambiguous SVO and OVS and in ambiguous OVS sentences. For unambiguous sentences, anticipatory looks to the patient in SVO sentences and the agent in OVS sentences were in line with correct responses to post-sentence comprehension questions. For ambiguous OVS sentences, response accuracy was slightly lower than for unambiguous OVS sentences. Since SVO word order is preferred over OVS in German (Dittmar et al., 2008a; Hemforth, 1993; Hemforth & Konieczny, 2013), it is possible that participants initially interpreted ambiguous sentences as SVO sentences (corroborated by patient anticipation during verb) and had to restructure towards OVS at the end of the sentence. Perhaps they failed to restructure towards OVS in a few cases resulting in slightly lower response accuracy for ambiguous sentences.

8.4 SUMMARY AND CONCLUSION

To sum up, the results of Experiments 2a and 2b (unambiguous SVO and OVS sentences) corroborate the findings of Experiment 1. The absence of clear effects of prosody underlines the idea that morpho-syntactic information (case-marking) is used as a stronger cue for/ has stronger links to thematic role assignment than supra-sentential information (prosody). In other words, case-marking affects visual attention/language comprehension more than prosody. We observed anticipatory eye-movements to the patient in SVO sentences and the agent in OVS when case-marking was unambiguous (as early as the verb region, Figure 8.2a). When case-marking was ambiguous, the additional stress on the verb slightly eliminated participants SVO preference during the adverb region, eliminating the bias to inspect the patient, whereas the main stress on the first noun-phrase did not (Figure 8.7b). The ambiguity in the visual display likely influenced these results.

Part III

WIGGLE, DEPICTED ACTION, OR BOTH? DISTINCT AND COMBINED VISUAL CUES DURING REAL-TIME THEMATIC ROLE ASSIGNMENT

9

EXPERIMENT 3

In Experiment 1, we argued that children might need additional information from the visual display to use case-marking for correct thematic role assignment. We further suggested in Experiment 2 that the visual display slightly influenced adults' use of prosody. What remains unclear is whether visual information can help to disambiguate ambiguous sentence structures. Experiment 3[1] thus investigated the influence of visual information on children's and adults' thematic role assignment in ambiguous OVS sentences.

In rich contexts, adult comprehenders can exploit non-linguistic cues (among others) to incrementally interpret the linguistic input. Visual referential context, contrast between objects, depicted actions, or events can all rapidly influence the interpretation, syntactic structuring, and thematic role assignment of spoken utterances (e.g., Chambers et al., 2004; Knoeferle et al., 2005; Sedivy et al., 1999; Tanenhaus et al., 1995). Existing findings revealed that adults used referential context to disambiguate syntactic structures (e.g., *Put the apple on the towel in the box*). Either one or two referents were present in the visual display (one referent: an apple on the towel; two referents: an apple on a towel and another apple on a napkin). Participants' gaze-pattern suggested a destination interpretation of *the towel* in the one-referent context and a location interpretation in the two-referent context.

Further studies have found that listeners used adjectival information (small vs. tall) contrastively to incrementally establish reference to an object (Sedivy et al., 1999). Additionally, comprehenders used visual saliency for syntactic ambiguity resolution. In a visual world eye-tracking study, participants used visual saliency (colour, intensity, and orientation of objects) for correct prepositional-phrase attachment (e.g., *The girl will put the orange on the tray in the bowl*; Coco, 2015). Event relations depicted in the visual context helped to incrementally assign thematic roles. Participants listened to locally ambiguous German SVO and OVS sentences (transl: 'the princess$_{\text{agent/patient}}$ washes/paints apparently the pirate$_{\text{patient}}$/the fencer$_{\text{agent}}$'). Sentences did

1 A six-page conference paper reporting the results from this Experiment has been published in the Proceedings of the 40th Annual Meeting of the Cognitive Science Society: Kröger, Münster, Burigo, and Knoeferle (2018).

not provide information about who is the agent or the patient prior to the second noun-phrase. Scenes depicted the princess in agent and patient role (the fencer was acting upon the princess while the princess was acting upon a pirate). During the verb ('washes/paints'), participants successfully anticipated the agent or patient role filler. Thus, the ambiguity in the linguistic input was resolved via event information in the visual display (Knoeferle et al., 2005).

Prior research further investigated effects of different types of information on language comprehension such as speaker gaze vs. a dot cursor (Brennan, Chen, Dickinson, Neider, & Zelinsky, 2008), speaker gaze vs. an arrow (Staudte et al., 2014), depicted actions vs. a recent emotional speaker face (Münster et al., 2015), or speaker gaze shifts vs. depicted actions (Kreysa et al., 2014). Although these cues all rapidly influenced language comprehension, some revealed similar but others different effects on language comprehension. Speaker gaze shifts and depicted actions, for example, elicited similar gaze-patterns to the target character (although only when they were both employed as purely deictic cues; Kreysa et al., 2014). An emotional prime face, however, only seemed to influence participants' visual attention when actions were also depicted (Münster et al., 2015).

For children, not all aspects of visually situated language processing are present from the beginning. A study with 6-month-olds provided first evidence for the close temporal coordination of auditory input and visual attention (Richardson & Kirkham, 2004). When infants saw an object (e.g., a toy) and jointly heard a sound (e.g., *boing boing*) and listened to a spatially non-informative *boing* sound immediately after, they inspected the side of the screen more which was previously occupied by the associated toy than the other side of the screen. During utterance comprehension, ten- to eleven-year-olds (Nation et al., 2003) and two-year-olds (Mani & Huettig, 2012) rapidly anticipated upcoming referents (e.g., a cake) during *eats the* when they listened to *The boy eats the big cake* but not when they listened to *The boy sees the big cake*.

However, existing evidence from eye-movements also suggests some differences between children's and adults' visual attention and also some developmental differences between children of different age groups. Unlike 30-month-olds, 36-month-olds (much like adults) rapidly inspected a blue car (vs. red car) when they responded to questions such as *Can you find the blue car?* (Fernald et al., 2010). Further differences emerged in the use of referential context for structural disambiguation. When hearing *Put the frog on the napkin* *in the box*, five-year-olds interpreted *on the napkin* as the destination of *the frog* in both, the one- and the two-referent context, although the two-referent context supports a location interpretation (Trueswell et al., 1999). However, depicted action events influenced children's thematic role assignment in structurally unambiguous German SVO and OVS sentences (Münster, 2016; Zhang & Knoeferle, 2012). When actions were depicted, children inspected the patient (vs. agent) more in SVO sentences and the agent (vs. patient) more in OVS sentences.

Taken together, it remains an open question to which extent distinct visual cues can guide comprehenders' visual attention and influence language comprehension. Likewise, it remains unclear whether visual cues are processed in a similar fashion by children in comparison to adults to disambiguate locally structurally ambiguous OVS sentences. Crucially, the extent to which these cues influence thematic role assignment when they are present for only a short period of time, similar to some linguistic cues (e.g., prosody), requires further research. Experiment 3 thus investigated whether temporally limited depicted actions influence children's and adults' thematic role assignment. Additionally, we explored the influence of a wiggling motion of the target character which could function as a pragmatic/focusing cue (similar to prosody in Experiment 1 and 2) to the extent that comprehenders infer that a wiggling character is the agent which in turn disambiguates the sentence early.

We thus investigated whether a wiggling target character helps children to correctly assign thematic roles or whether they fail to draw pragmatic inferences, as suggested by Trueswell et al. (1999). In their study, children failed to infer that the napkin modified one of the frogs. A wiggling target character might, however, attract the comprehenders' attention through the abrupt motion (similar to the action) and convey thematic roles but not per pragmatic inferences. Since different types of visual cues elicited similar or combined effects on language comprehension, we further assessed whether one cue is stronger than the other (action or wiggle) and whether the combination of cues (action plus wiggle) has beneficial effects on thematic role assignment. Overall, the data from Experiments 1 and 2 suggests that prosody is a weak cue for thematic role assignment. We thus did not additionally manipulate prosody in Experiment 3. We used one prosodic structure (OVS-biasing prosody) for all sentences.

9.1 MATERIALS AND DESIGN

PARTICIPANTS. 24 young adult (mean age = 27.8, SD = 3.51) and 24 five-year-old (age range 4.5 - 5.8, mean = 4.6, SD = .49) monolingual speakers of German took part in the experiments. All had normal or corrected vision and hearing. Adults were paid and children received a toy and a certificate for participation. Written informed consent was given by adults themselves and by parents for their children. Children gave consent orally (we asked them if they would like to play a game on the computer). The experiment was approved by the ethics committee of the Bielefeld University. Children were recruited from kindergartens in and around Paderborn (a city close to Bielefeld).

SPEECH STIMULI. Auditory material included the ambiguous OVS sentences recorded for Experiment 2b, with an L*+H accent on the first noun-phrase. Since we did not observe clear effects of prosody in the first two experiments, we did not expect prosody to influence the results. Case-marking on the first noun-phrase was either feminine or neuter which are both identical in accusative and nominative case, creating local structural ambiguity. Case-marking on the second noun-phrase was unambiguous such that case-marking at the end of the sentence disambiguated 'who does what to whom'.

VISUAL STIMULI. We used the 24 scenes from Experiment 2 but manipulated the availability of visual cues. Condition a) *no action no wiggle* only contained the three clipart characters (no cue baseline). In condition b) *no action one wiggle*, the target character (the agent) wiggled up and down (first five and then ten pixels up and then five and ten pixels down). In condition c) *one action no wiggle*, the scene depicted the agent performing an action (e.g., for the verb *filmen* ('films') the agent held a camera). Condition d) *one action one wiggle* merged the two single cues: The agent performed an action and wiggled up and down at the same time (see Table 9.1 for an overview of the conditions).

Table 9.1: Conditions Experiment 3

Condition	Sentence Structure	Visual Cue
a	ambOVS	no action no wiggle
b	ambOVS	no action one wiggle
c	ambOVS	one action no wiggle
d	ambOVS	one action one wiggle

DESIGN. The design of Experiment 3 was similar to the design of Experiments 1 and 2. However, instead of the two factors word order and prosody, we now had action and wiggle which left us with four lists. Again, we added counterbalancing (left/right) as a factor resulting in eight lists. Active and passive voice comprehension question were again added across the eight lists such that each item occurred once with a left balancing and an active question and once with a right balancing and a passive question within the same condition. We again pseudo-randomised these eight lists for each participant (24 lists). We used all the fillers described in Experiment 2 for the adults. The fillers each occurred in one of the four experimental conditions. However, in 48 fillers, the action depicted mismatched the action described in the sentence or a character other than the agent wiggled. We decided to use sentence-scene mismatches in the fillers to avoid the that participants can identify the goal of the experiment. For the children, we used 8 fillers instead of 72 to shorten the testing sessions but the experimental items were identical.

PREDICTIONS. Previous studies have shown that adults and children are able to exploit depicted actions to correctly assign thematic roles in OVS sentences (Knoeferle et al., 2005; Münster, 2016; Zhang & Knoeferle, 2012). Additionally, visual saliency (in our case a wiggling animal character) can help to resolve syntactic ambiguities (Coco, 2015). If adults and children are also able to rapidly exploit short-lived visual cues for thematic role assignment in ambiguous OVS sentences, we should replicate the rapid visual context effects reported in previous studies (likely upon hearing the verb: The point in time during which the visual cue was present). If they rapidly exploit depicted actions that are temporally limited in their presentation, then we expected to find more looks to the agent (vs. patient) when the actions were depicted compared to when they were not depicted. If participants are also able to rapidly exploit the wiggle for incremental thematic role assignment, then we should see more looks to the agent (vs. patient) in the wiggle (vs. no wiggle) condition. If one cue is stronger than the other, we should observe more looks to the agent (vs. patient) in one of the two one-cue conditions (depicted action vs. wiggle). If the combination of cues (depicted action and wiggle) has beneficial effects, participants should inspect the agent (vs. patient) more in the two-cue (vs. one-cue) condition.

For post-sentence comprehension questions, we expected a difference in the number of correct responses between children and adults. Case-marking on the second noun-phrase disambiguated the sentence. Prior research and the results from our first experiment suggests that adults are able to use case-marking for thematic role assignment (Kamide, Scheepers, & Altmann, 2003; Kröger et al., 2017; Matzke et al., 2002; Zhang & Knoeferle, 2012). If so, cue presence should not influence the offline data with an overall high number of correct responses in adults. Children's ability to use case-marking for thematic role assignment seems, however, limited to visual contexts in which world knowledge supports their interpretation (Özge et al., 2016 but Dittmar et al., 2008a; Schipke et al., 2011; results from Experiment 1). In the absence of helpful visual cues (no-cue baseline), we expected children to be at chance in responding to post-sentence comprehension questions since our scenes did not provide world knowledge.

If children are able to use the temporally limited visual cues, we should a) replicate the effects of depicted actions found by Zhang and Knoeferle (2012) and Münster (2016): Improved responses when the actions are depicted (vs. not depicted), and b) observe similar effects for the wiggle as for the action: More correct responses when the wiggle is present (vs. absent). If children, however, struggle with pragmatic inferences, as suggested by Trueswell et al. (1999), wiggle presence may not have beneficial effects on accuracies. If one of the two visual cues is stronger than the other, correct responses should be higher for one of the two cues, action or wiggle. If the combination of the two cues (condition d) one action one wiggle) facilitates comprehension more, we should see more correct responses in the two-cue condition as compared to the one-cue conditions.

Existing findings suggest that visual context effects can further be influenced by children's cognitive capacities (Zhang & Knoeferle, 2012 but Münster, 2016). If cognitive abilities influence eye-movements, we may observe more looks to the agent (vs. patient) in children with higher (vs. lower) cognitive abilities. Post-sentence comprehension questions could also reveal differences between the two groups (high vs. low): More correct responses for children with higher (vs. lower) cognitive abilities.

PROCEDURE. In this experiment, we followed a similar procedure as in Experiments 1 and 2. The visual cues were time-locked to verb onset (Figure 9.1). The presentation of the visual cue/ visual cues was temporally limited to the verb region. We decided to use this time frame because it was the point in time when the depicted action was mediated by the verb (the action was a visualisation of the verb; the action is being performed by the character holding the object-the agent). In order to conduct a direct comparison between the depicted action and the wiggle-only/wiggle plus action, we kept the time window in which the visual cues appeared constant for all conditions, even though the wiggle was not mediated by the verb.

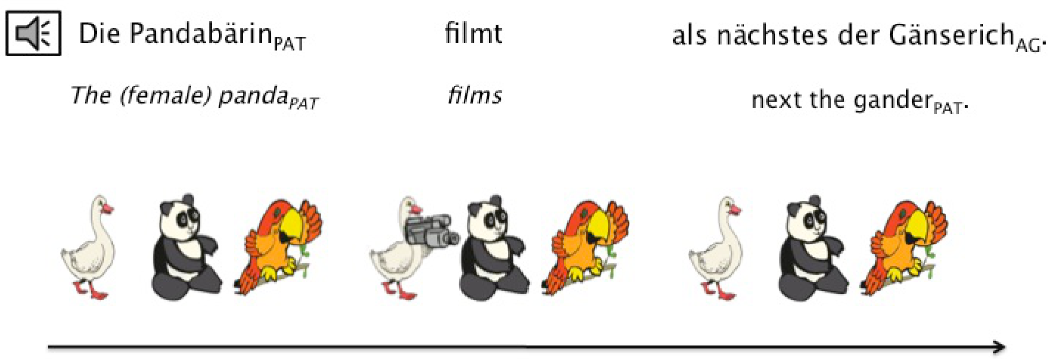


Fig. 9.1: *Visualisation of the sequence during auditory input: Example for action cue.*

ANALYSIS. The analyses were very similar to the analyses for Experiments 1 and 2 (for word on- and offsets see Appendix A.3). We ran two (action) by two (wiggle) repeated-measures ANOVAs by participants and by items. We predefined two word regions of interest: Verb and adverb region. We expected to find effects of the visual cues in these two word regions. We were interested in the verb region because this was the time region in which the visual cue(s) was/were presented. We further analysed the adverb region to capture post-verbal effects. For accuracies, we also ran Mixed Effects Models but with action and wiggle as the two fixed factors (instead of word order and prosody). The mean onsets and the longest word durations were identical to Experiment 2b (see Table 9.2 / Table 8.3).

Table 9.2: Mean Onsets and Longest Word Durations Experiment 3 (in ms)

	NP1	Verb	Adverb	NP2
Mean onsets	0	1458	2696	3982
Longest duration	1673	1469	1625	1515 (1015+500)

9.2 RESULTS AND DISCUSSION

9.2.1 *Adults*

TIME COURSE GRAPHS. During the NP1 region, no preference of looking towards the agent or the patient was observed. From the beginning of the verb region, the results revealed more looks towards the agent (vs. patient) in conditions b) no action one wiggle, c) one action one wiggle, and d) one action one wiggle. Participants anticipated the agent (vs. patient) even more when both cues were available - condition d) - compared to when only one cue was - conditions b), c). In the no-cue baseline (condition a) no action no wiggle), participants inspected both, agent and patient, on equal terms.

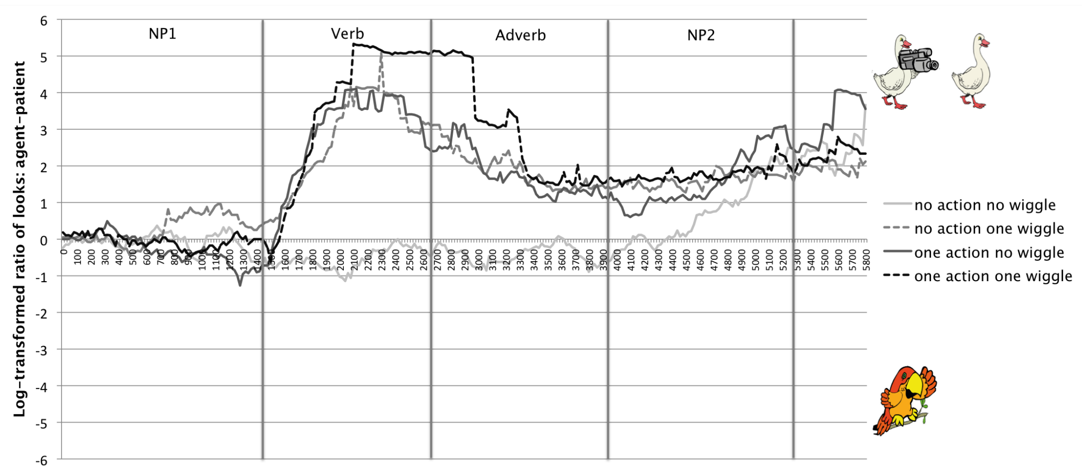


Fig. 9.2: *Experiment 3 adults: Time course of eye-movements including the visual bias (y-axis) towards agent (positive values) relative to patient (negative values) in all conditions (ln(agent/patient)) over time in ms (x-axis). The vertical lines represent mean onsets for each word region.*

Similar gaze-patterns as in the verb region occurred during the adverb region. However, the preference of looking towards the agent (vs. patient) was slightly lower than in the verb region in conditions b), c), and d). During the NP2 region, participants continued to anticipate the agent (vs. patient) in conditions b), c), d), and started to anticipate (and perhaps disambiguate towards) the agent in condition a). Agent preference is almost equally high in all four conditions towards the end of the NP2 region, likely because the agent was mentioned.

WORD REGIONS. In the verb region, the effects of action, wiggle, and the interaction of both were all significant (Table 9.3). We observed more looks to the patient (vs. agent) in condition a) no action no wiggle. In all the other conditions (visual cue available), participants showed a substantial preference to look at the agent (vs. patient). However, in condition b) no action one wiggle, this preference was slightly lower than in condition c) one action no wiggle and condition d) one action one wiggle (Figure 9.3a). A paired sample t-test after Bonferroni (.05/6) revealed a significant difference between condition a) and conditions b), c), and d) ($ps < .01$). No reliable difference was observed between the other conditions.

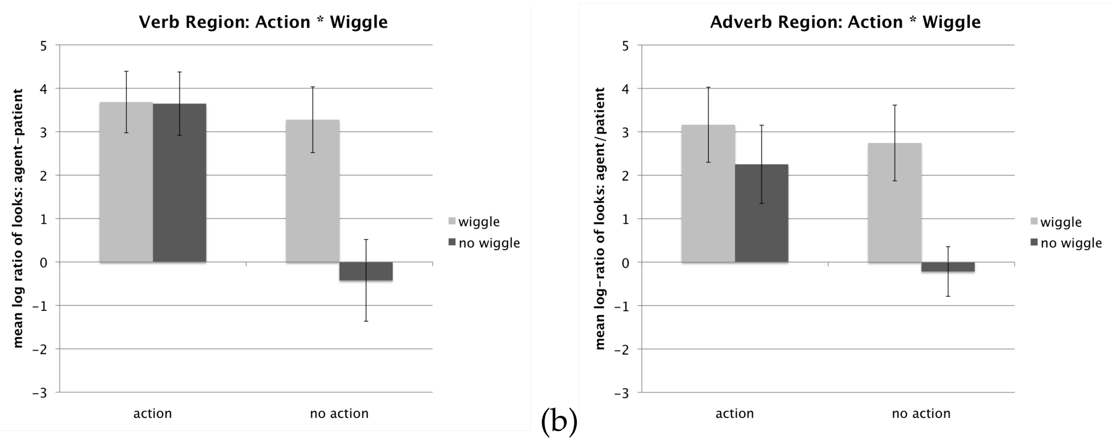
Table 9.3: Experiment 3 adults: ANOVA results by participants and by item during the verb region

participants	**F(1,23)**	**MSE**	**p**	η^2
action	61.595	1.951	< .001	1
wiggle	24.258	3.455	< .001	.997
action*wiggle	19.061	4.228	< .001	.987
item				
action	24.612	4.289	< .001	.997
wiggle	24.572	2.810	< .001	.997
action*wiggle	25.260	3.807	< .001	.998

In the adverb region, we also observed reliable effects of action and wiggle. The interaction of action and wiggle was significant in the analysis by participants but only marginally significant in the analysis by items (Table 9.4). Participants continued to inspect the patient more than the agent when no visual cue was available (condition a). For conditions b), c), and d) we observed a clear preference to inspect the agent (vs. patient). However, this preference was minimally lower in condition c) one action no wiggle, than in conditions b) no action one wiggle and d) one action one wiggle (Figure 9.3b). Comparisons, via paired sample t-tests after Bonferroni, revealed a significant difference ($ps < .01$) between condition a) and conditions b), c), and d). Further comparisons did not reveal significant differences. Descriptively, mean log-ratios were higher during the verb region (compared to the adverb region). Adults' attentional responses to the visual cues were likely very fast (Figure 9.3).

Table 9.4: Experiment 3 adults: ANOVA results by participants and by item during the adverb region

participants	**F(1,23)**	**MSE**	**p**	η^2
action	18.223	2.747	$< .001$	.983
wiggle	29.974	2.998	$< .001$	.997
action*wiggle	8.179	3.077	$< .01$	.782
item				
action	14.763	3.340	$< .01$	.957
wiggle	28.112	2.998	$< .001$	.999
action*wiggle	3.218	2.870	.086	.405



(a) (b)

Fig. 9.3: *Experiment 3 adults: Mean log-ratios of looks towards the agent (positive numbers) relative to the patient (negative numbers) during the verb (a) and adverb (b) region in all four conditions. The graphs represent the interaction of action and wiggle. Error bars reflect 95% confidence intervals.*

In the NP2 region, the effects of action and wiggle or the interaction disappeared in the by participant analyses (action: $F_1(1,23) = 1.236$, $p = .261$, wiggle: $F_1(1,23) = 2.111$, $p = .160$, action and wiggle interaction: $F_1(1,23) = 1.749$, $p = .199$) as did the effect of action and the interaction of action and wiggle in the by items analyses (action: $F_2(1,23) = 2.788$, $p = .109$, action and wiggle interaction: $F_2(1,23) = 1.304$, $p = .265$) but the effect of wiggle was significant in the by items analyses $F_2(1,23) = 6.165$, $MSE = 2.273$, $p < .05$, $\eta^2 = .211$. Participants inspected the agent (vs. patient) more in all four conditions.

ACCURACY. Surprisingly, the overall response accuracy was only 68.9%. In condition a) no action no wiggle 66.7% of the responses were correct, similar to condition b) no action one wiggle in which the percentage of correct responses was 66%. In condition c) one action no wiggle, the data revealed 68.1% correct responses and in condition d) 75% (Figure 9.4). Generalised Linear Mixed Effects Model results revealed a marginal interaction of action and wiggle (Table 9.5): Wiggle presence had beneficial effects on accuracies when actions were also depicted. Following from the descriptive data, correct responses were highest when both cues were present (vs. one or none). Response accuracy was similarly high for active (69.4%) as for passive voice (68.8%) comprehension questions (Figure 9.5).

Table 9.5: Generalised Linear Mixed Effects Model Results: Accuracy Experiment 3 Adults. Maximal Converging Model: accuracy ~ actionC * wiggleC + (1 | subject) + (1 | item).

	Coefficient	Standard Error	z-value	p
Intercept	1.16654	0.63912	1.825	0.0680
action	0.16186	0.37999	0.426	0.6701
wiggle	-0.07818	0.37323	-0.210	0.8341
action*wiggle	1.00072	0.55956	1.788	0.0737

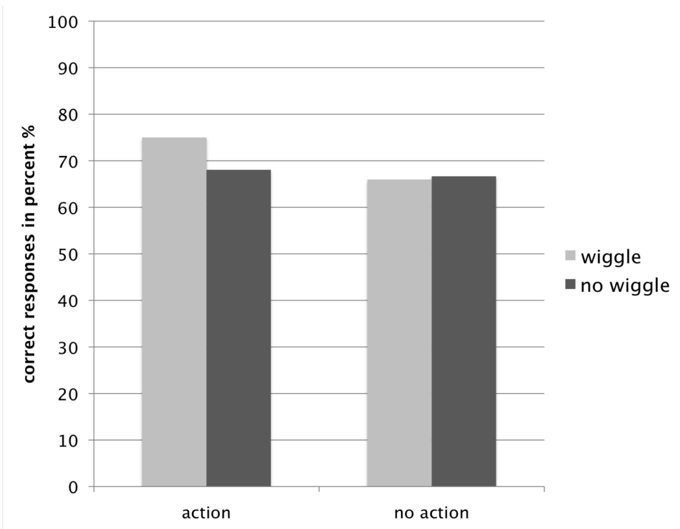


Fig. 9.4: *Experiment 3 adults: Correct responses to post-sentence comprehension questions by condition (x-axis) in percent (y-axis).*

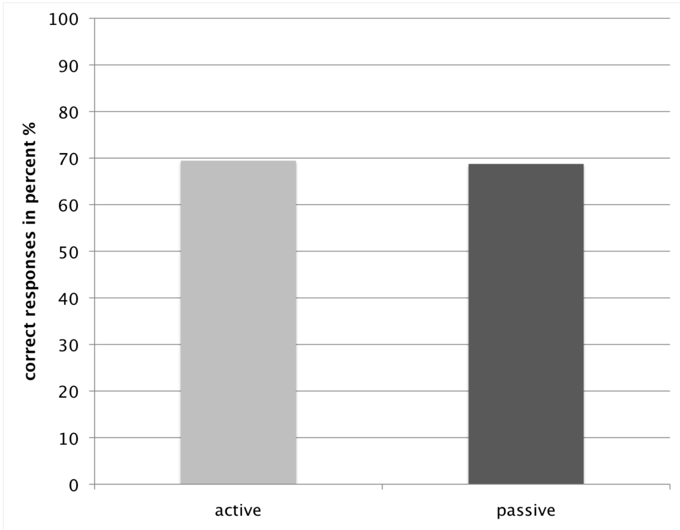


Fig. 9.5: *Experiment 3 adults: Correct responses to post-sentence comprehension questions by active and passive voice (x-axis) in percent (y-axis).*

9.2.2 *Children*

TIME COURSE GRAPHS. During the NP1 region, children did not have a clear preference to look towards the agent or the patient (likely because they inspected the ambiguous character). During the verb region, children inspected the patient slightly more than the agent in condition a) no action no wiggle. In conditions b) no action one wiggle and c) one action no wiggle, the results show a clear preference to inspect the agent (vs. patient). The preference to inspect the agent (vs. patient) was slightly higher for condition d) than conditions b) and c). In other words, children inspected the agent (vs. patient) more when two cues were available (action plus wiggle) compared to when only one cue (action or wiggle) was present.

During the adverb region, children continued to inspect the patient more than the agent when no visual cue was available - condition a). We further observed more looks towards the agent than the patient in conditions b), c), and d). At the beginning of the adverb region, children continued to inspect the agent (vs. patient) more in condition d) compared to conditions b) and c). Towards the end of the adverb region, the number of looks towards the agent (vs. patient) was similar in conditions b), c), and d), however, slightly lower than at the beginning of the adverb region. Towards the middle of the NP2 region, children started to look more at the agent (vs. patient) in condition a) and even more than in conditions b), c) and d) after NP2 offset, indicating late sentence disambiguation. In conditions b), c) and d), children continued to inspect the agent more than the patient. The overall agent preference was, however, lower during the NP2 region than during the verb and adverb region.

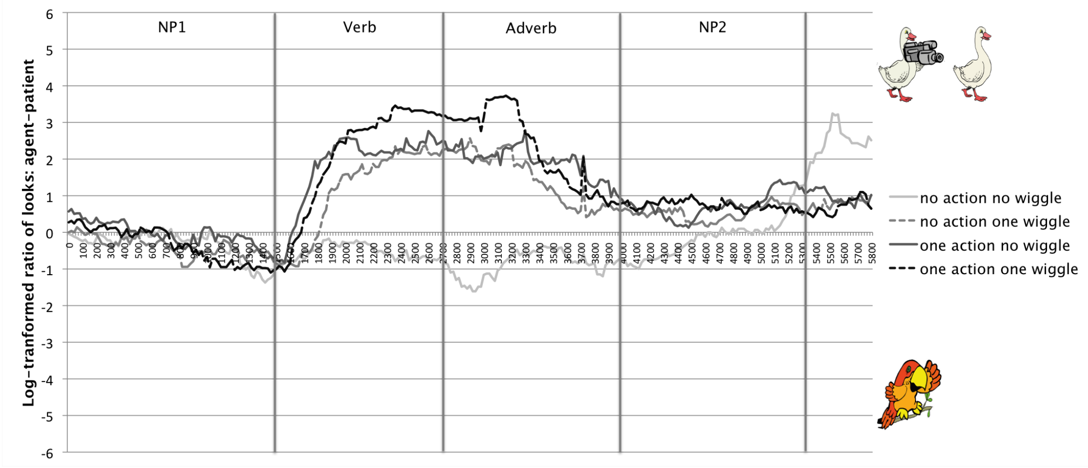


Fig. 9.6: *Experiment 3 children: Time course of eye-movements including the visual bias (y-axis) towards agent (positive values) relative to patient (negative values) in all conditions (ln(agent/patient)) over time in ms (x-axis). The vertical lines represent mean onsets for each word region.*

WORD REGIONS. During the verb region effects of action, wiggle, and the interaction of both were significant (Table 9.6). The results suggest that children inspected the patient (vs. agent) more in condition a) no action no wiggle. In the presence of visual cues, we observed a clear preference of agent (vs. patient) inspections but this preference was slightly lower in condition b) no action one wiggle than in conditions c) and d) (Figure 9.7a). A paired sample t-test after Bonferroni (.05/6) revealed that condition a) was significantly different from conditions b), c), and d) ($ps < .01$), corroborating that cue presence influenced looks to the agent and cue absence looks to the patient.

In the adverb region, the effects of action, wiggle, and the interaction action and wiggle were also significant (Table 9.7). Participants inspected the patient more than the agent in condition a) no action no wiggle, and the agent more than the patient in the presence of visual cues (conditions b), c), and d). However, this agent preference was higher in condition c) compared to condition b) and even higher in condition d) (Figure 9.7b). A paired-samples t-test after Bonferroni (.05/6) repeatedly revealed significant results for the comparisons of condition a) with conditions b), c), and d) ($ps < .01$): Significantly more agent (vs. patient) inspections in conditions b), c), and d) compared to condition a). Furthermore, the difference between conditions b) no action one wiggle and condition d) one action one wiggle ($p < .01$)

Table 9.6: Experiment 3 children: ANOVA results by participants and by item during the verb region

participants	**F(1,23)**	**MSE**	**p**	η^2
action	33.613	1.934	$< .001$	1
wiggle	17.377	1.580	$< .001$	.979
action*wiggle	7.819	3.191	$< .05$	.764
item				
action	18.835	4.764	$< .001$	.986
wiggle	15.075	2.429	$< .01$	.960
action*wiggle	12.625	2.837	$< .01$	.925

was significant. Children's preference to inspect the agent (vs. patient) was significantly higher when both visual cues were present (action plus wiggle) compared to when only the wiggle was - condition b). Descriptively, mean log-ratios were higher during the adverb region compared to the verb region (Figure 9.7).

Table 9.7: Experiment 3 children: ANOVA results by participants and by item during the adverb region.

participants	**F(1,23)**	**MSE**	**p**	η^2
action	33.613	1.934	$< .001$	1
wiggle	17.377	1.580	$< .001$	.979
action*wiggle	7.819	3.191	$< .05$	.764
item				
action	18.835	4.764	$< .001$	.986
wiggle	15.075	2.429	$< .01$	.960
action*wiggle	12.625	2.837	$< .01$	.925

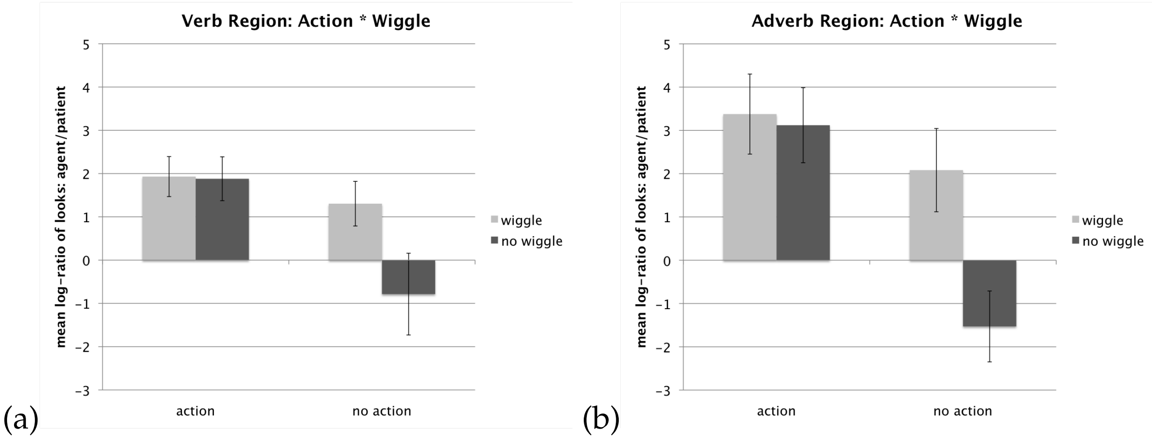


Fig. 9.7: *Experiment 3 children: Mean log-ratios of looks towards the agent (positive values) relative to the patient (negative values) during the verb (a) and adverb (b) region in all four conditions. The graphs represent the interaction of action and wiggle. Error bars reflect 95% confidence intervals.*

Additional effects emerged in the NP2 region (Figure 9.8). We observed an unexpected marginal effect of action in the by participants analysis and a significant effect of action in the by items analysis, $F_1(1, 23) = 4.201$, $MSE = 2.149$, $p = .052$, $\eta^2 = .154$. $F_2(1, 23) = 8.234$, $MSE = 1.576$, $p < .01$, $\eta^2 = .264$. Participants inspected the agent (vs. patient) more when the action had been present (vs. absent). Effects during the NP2 region were unexpected because the name of the target character is uttered in this word region. Therefore, we expected participants to inspect the target regardless of condition. Furthermore, the interaction of action and wiggle was marginally significant in the by participants analysis and significant in the by items analysis, $F_1(1, 23) = 3.530$, $MSE = 2.605$, $p = .073$, $\eta^2 = .133$. $F_2(1, 23) = 8.837$, $MSE = 1.772$, $p < .01$, $\eta^2 = .278$. The effect of wiggle was non-significant in both by participants and by items analyses, $F_1(1, 23) = 2.064$, $p = .164$, $F_2(1, 23) = 2.025$, $p = .168$. Wiggle presence did not result in more looks to the agent (vs. patient) during the NP2 word region.

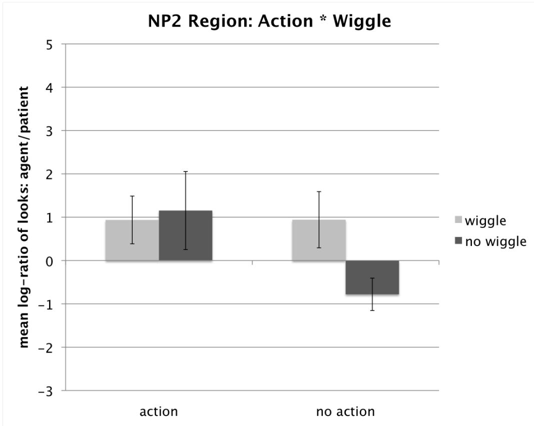


Fig. 9.8: *Experiment 3 children: Mean log-ratios of looks towards the agent (positive values) relative to the patient (negative values) during the NP2 region region in all four conditions. The graphs represent the interaction of action and wiggle. Error bars reflect 95% confidence intervals.*

ACCURACY. For the children, the overall response accuracy was 42.9%. In condition a) no action no wiggle, the percentage of correct responses was 41.7%, in condition b) no action one wiggle 42.4%, in condition c) one action no wiggle 39.6%, and in condition d) one action one wiggle 47.9% (Figure 9.9). Although children answered more questions correctly in condition d) one action one wiggle than in all the other conditions, the results from the Generalised Linear Mixed Effects Model revealed no significant effects of action, wiggle or an interaction of action and wiggle on accuracies.

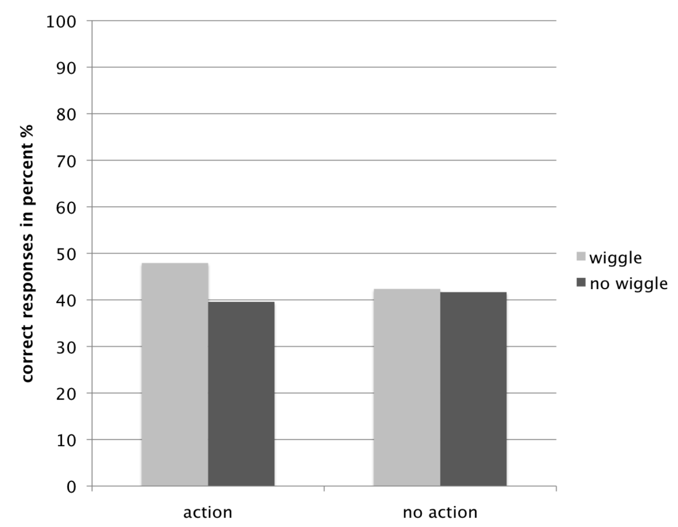


Fig. 9.9: *Experiment 3 children: Correct responses to post-sentence comprehension questions by condition (x-axis) in percent (y-axis).*

However, similar to Münster (2016), we observed a main effect of voice of the comprehension question (Table 9.8). Children's responses were more often correct when comprehension questions were in active voice (60.4%) compared to when they were in passive voice (25.3% more; Figure 9.10). Comprehension questions in active voice directly related to the target character such that the agent was the question target (the character that wiggled, performed an action, or wiggled and performed an action). For example, for the sentence *Die Pandabärin filmt als nächstes der Gänserich* ('The panda$_{\text{patient}}$ films next the gander$_{\text{agent}}$') and the scene gander$_{\text{visual cue}}$->panda->parrot, the answer to the question *Wer filmt hier* ('Who films here?') was *der Gänserich* ('the gander'). For questions in passive voice *Wer wird hier gemalt?* ('Who is being painted here?'), however, the question target was the ambiguous character and the patient - *die Pandabärin* ('the panda') in OVS sentences.

Table 9.8: Generalised Linear Mixed Effects Model Results: Accuracy Experiment 3 Children: plus voice. Maximal Converging Model: accuracy ~ actionC + wiggleC + actpasC + (1 | subject) + (1 | item).

	Coefficient	**Standard Error**	**z-value**	**p**
Intercept	0.3248	0.3274	0.992	0.321
action	0.2032	0.2098	0.969	0.333
wiggle	0.2757	0.2037	1.353	0.176
voice	-1.9519	0.2199	-8.875	0.000

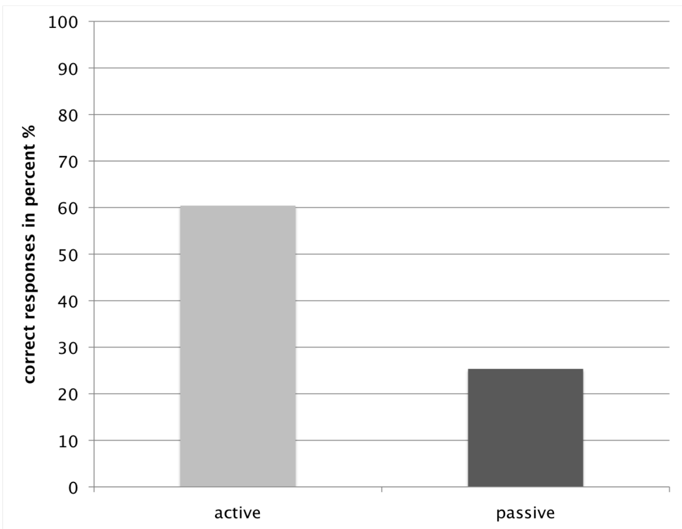


Fig. 9.10: *Experiment 3 children : Correct responses to post-sentence comprehension questions by active and passive voice (x-axis) in percent (y-axis). a - active voice comprehension question.*

Following from the reliable effect of voice, we ran Generalised Linear Mixed Effects Models including the active voice data subset only. The models were identical to the ones used for the whole dataset. The results revealed a marginal effect of wiggle (Table 9.9). Wiggle presence had a positive effect on response accuracy such that responses were more often correct when the wiggle was present (vs. absent; Figure 9.11).

Table 9.9: Generalised Linear Mixed Effects Model Results: Accuracy Experiment 3 Children: active voice only. Maximal Converging Model: accuracy ~ actionC + wiggleC (1 | subject) + (1 | item).

	Coefficient	Standard Error	z-value	p
Intercept	0.4416	0.4128	1.070	0.285
action	0.1846	0.3652	-0.506	0.613
wiggle	0.4872	0.2865	1.700	0.089

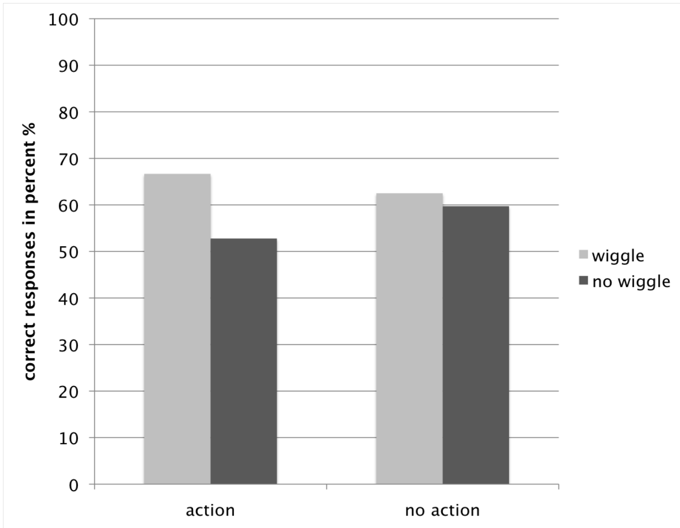


Fig. 9.11: *Experiment 3 Children: Correct responses to active voice post-sentence comprehension questions by condition (x-axis) in percent (y-axis).*

We additionally analysed the passive voice data subset separately using Generalised Linear Mixed Effects Models (Maximal Converging Model: accuracy ~ actionC + wiggleC + (1 | subject)). The results revealed no reliable effects of action or wiggle and no interaction. Although the overall number of correct responses was very low, descriptively, we observed slightly more correct responses when the action was present compared to when it was absent (Figure 9.12).

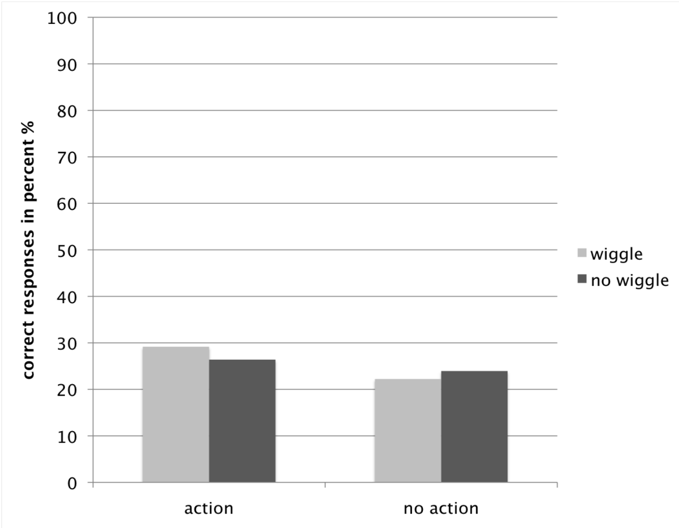


Fig. 9.12: *Experiment 3 children: Correct responses to passive voice post-sentence comprehension questions by condition (x-axis) in percent (y-axis).*

ACCURACY AND EYE-MOVEMENTS. For the verb and adverb region, we further ran two (action) by two (wiggle) repeated-measures ANOVAs including accuracy (high vs. low) as a between-participants factor. The results revealed no significant interaction effects of accuracy with action, wiggle, or action and wiggle (verb region - interaction action and accuracy: $F_1(1, 23) = .099$, $p = .756$; interaction wiggle and accuracy: $F_1(1, 23) = 1.544$, $p = .227$; interaction action, wiggle, and accuracy: $F_1(1, 23) = .001$, $p = .975$; adverb region - interaction action and accuracy: $F_1(1, 23) = .539$, $p = .471$; interaction wiggle and accuracy: $F_1(1, 23) = .169$, $p = .685$; interaction action, wiggle, and accuracy: $F_1(1, 23) = 1.339$, $p = .260$. Therefore, differences in response accuracy did not influence gaze-patterns.

K-ABC. The number of correct responses was highest for the Word Order subtest (91.7%) and the Spatial Memory subtest (88.3%). In the Number Recall subtest, children's correct responses were slightly lower (71.5%; Figure 9.13; see Appendix G, Table G.2 for single subjects results).

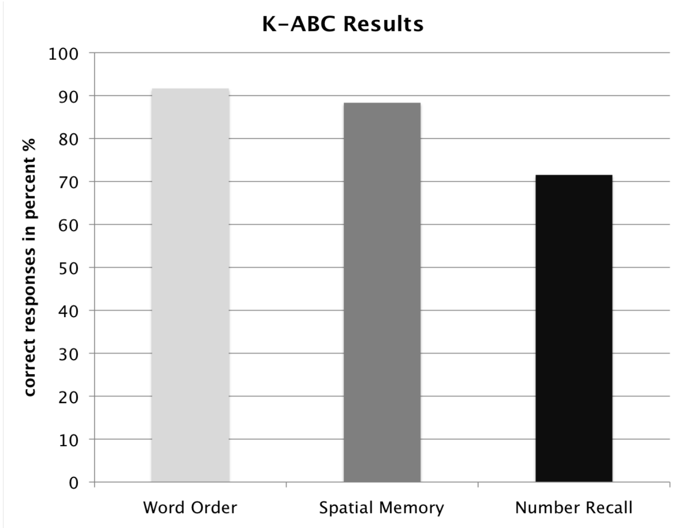


Fig. 9.13: *Experiment 3 children: Results of K-ABC cognitive abilities tests. Overall number of correct responses in percent divided by subtests.*

We conducted a Pearson's correlation coefficient (the data was normally distributed) between the K-ABC results and the accuracy data to examine whether children's cognitive abilities correlated with the accuracy data. The results revealed no correlation between the K-ABC data and the accuracy data (Table 9.10). Thus, children's cognitive abilities did not influence their performance in post-trial comprehension questions.

Table 9.10: Experiment 3 children: Pearson correlation: K-ABC - accuracy data

		Accuracy	K-ABC
Accuracy	Pearson Correlation	1	−.044
	Sig. (2-tailed)		.838
	N	24	24
K-ABC	Pearson Correlation	−.044	1
	Sig. (2-tailed)	.838	
	N	24	24

Descriptively, however, it seems that children with higher (vs. lower) cognitive abilities benefited most from the one-cue conditions (action or wiggle): correct responses were higher in the one-cue conditions compared to the no-cue or two-cue conditions. Children's responses with lower (vs. higher) cognitive abilities were more often correct in the two- and no-cue conditions compared to the one-cue conditions (Figure 9.14). Children with lower cognitive abilities may have struggled with the temporal limitations of cue depiction more than children with higher cognitive abilities resulting slightly lower response accuracy.

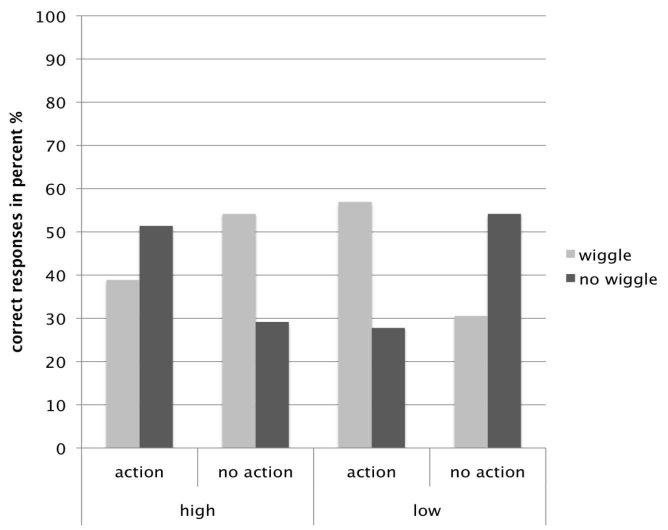


Fig. 9.14: *Experiment 3 children: Correct responses to post-sentence comprehension questions by condition and high/low cognitive abilities (x-axis) in percent (y-axis)*

K-ABC AND EYE-MOVEMENTS. For verb and adverb region, we ran two (action) by two (wiggle) repeated-measures ANOVAs including the K-ABC data (high vs. low) as a between-participants factor. The results of the verb region did not reveal effects of cognitive abilities (interaction action and K-ABC: $F_1(1, 23) = .036$, $p = .850$; interaction wiggle and K-ABC: $F_1(1, 23) = .053$, $p = .819$; interaction action, wiggle, and K-ABC: $F_1(1, 23) = .161$, $p = .692$).

The results of the adverb region, however, revealed a marginal interaction of action and K-ABC, $F_1(1, 23) = 4.108$, $MSE = 8.777$, $p = .055$, $\eta^2 = .157$. Children with higher cognitive abilities inspected the agent (vs. patient) more when the action was depicted (vs. not depicted), compared to children with lower cognitive abilities. We further observed a marginal interaction of action, wiggle, and K-ABC, $F_1(1, 23) = 3.861$, $MSE = 3.165$, $p = .062$, $\eta^2 = .149$. This effect likely resulted from the marginal interaction of action and cognitive abilities.

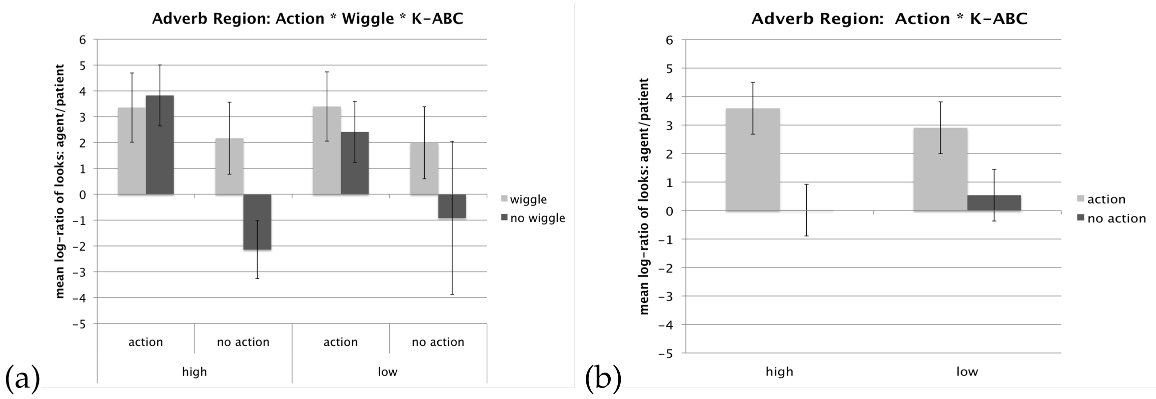


Fig. 9.15: *Experiment 3 children: Mean log-ratios of looks towards the agent (positive values) relative to the patient (negative values) during the adverb region. Graph (a) represents the interaction of action, wiggle, and K-ABC. Graph (b) represents the interaction of action and K-ABC. Error bars reflect 95% confidence intervals.*

9.2.3 *Adults and Children*

We expected participants to inspect the ambiguous role filler upon hearing the first noun-phrase. Figure 9.16 shows that this expected gaze-pattern emerged. The amount of fixations on the ambiguous role filler was very high during NP1 compared to the amount of fixations on the agent and patient role filler. The overall pattern was similar in adults and children, however, there was a difference in the amount of looks to the ambiguous role filler: Overall, adults inspected the ambiguous role filler more than children. Future research could zoom in on why children inspect the ambiguous character less than adults. Does more attention mean better comprehension, more attention to detail, or more critical scrutiny? In the following, we describe the results for analyses that treated participant age as a between-participants factor.

WORD REGIONS. In the verb region (Figure 9.17a), the results revealed a marginal interaction of age and wiggle in the by participants analysis but not in the by items analysis, $F_1(1,47) = 3.045$, $MSE = 2.517$, $p = .088.$, $\eta^2 = .062$, $F_2(1,47) = .974$, $p = .329$. Adults looked more at the agent (vs. patient) than children when the wiggle was present (vs. absent). We did not observe interactions of age and action or age, action, and wiggle (interaction age and action: $F_1(1,47) = 2.166$, $p = .148$, $F_2(1,47) = .071$, $p = .791$; interaction age, action, and wiggle: $F_1(1,47) = 2.138$, $p = .150$, $F_2(1,47) = 2.199$, $p = .145$).

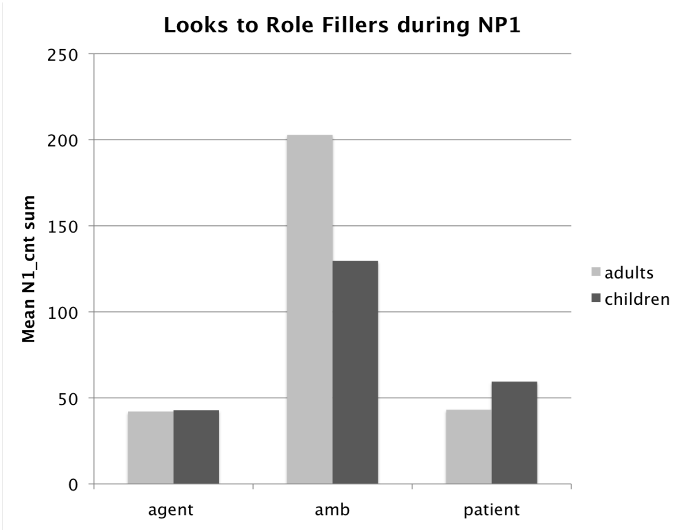


Fig. 9.16: *Experiment 3 adults and children: Mean fixation sum on the three role fillers (agent, ambiguous, patient) during the NP1 region.*

In the adverb region (Figure 9.17), the interaction of age and action was significant in the by participants analysis, $F_1(1, 23) = 10.834$, $MSE = 2.586$, $p < .05$ but not in the by items analysis, $\eta^2 = .191$, $F_2(1, 47) = 2.062$, $p = .158$. The interaction of age, wiggle, and action was significant in the by items analysis, $F_2(1, 23) = 4.049$, $MSE = 3.063$, $p = .05$, $\eta^2 = .081$ but not in the by participants analysis, $F_2(1, 47) = 1.544$, $p = .220$. Children, compared to adults, looked more at the agent (vs. patient) when the action was depicted (vs. not depicted; Figure 9.17).

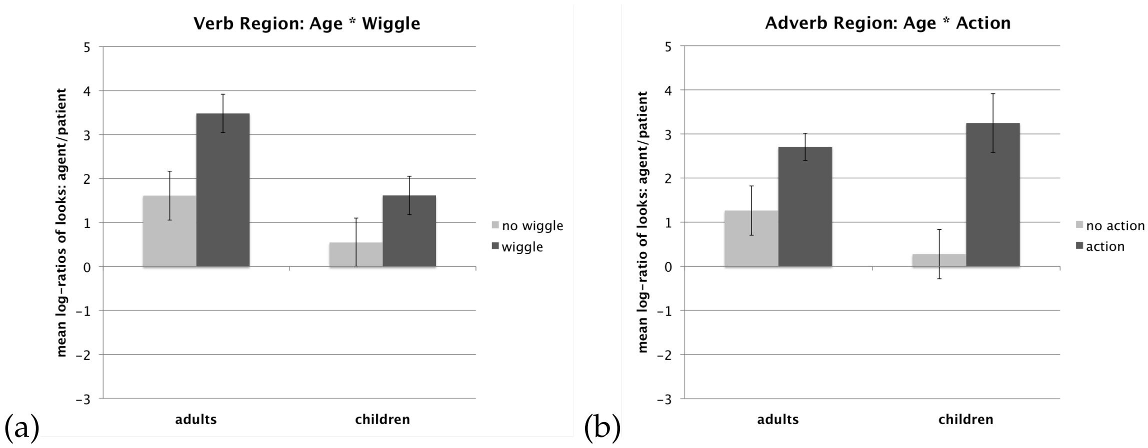


Fig. 9.17: *Experiment 3 adults and children: Mean log-ratios of looks towards the agent (positive values) relative to the patient (negative values) during verb and adverb region. Graph (a) represent the interaction of age and wiggle. Graph (b) represents the interaction of age and action. Error bars reflect 95% confidence intervals.*

ACCURACY. The results revealed no interaction of action or wiggle and age. We did, however, observe a reliable effect of age such that overall children's accuracies were significantly lower than adults' (Table 9.11), corroborating the observations across age groups.

Table 9.11: Generalised Linear Mixed Effects Model Results: Accuracy Experiment 3 Adults vs. Children. Maximal Converging Model: accuracy ~ actionC * wiggleC * ageC + (1 | subject) + (1 | item).

	Coefficient	**Standard Error**	**z-value**	**p**
Intercept	0.82462	0.27381	3.012	0.0026
action	0.07454	0.27245	0.274	0.7844
wiggle	-0.03676	0.27048	-0.136	0.8919
age	-1.22673	0.26739	-4.588	0.000
action*wiggle	0.43717	0.39187	1.116	0.2646
action*age	-0.17721	0.37745	-0.469	0.6387
wiggle*age	0.07070	0.37518	0.188	0.8505
action*wiggle*age	-0.06668	0.53715	-0.124	0.9012

9.3 DISCUSSION

In Experiment 3, we investigated whether adults and children rapidly exploit temporally limited visual cues (action-only, wiggle-only, and action plus wiggle) to assign thematic roles. Additionally, we examined whether one cue is used as a stronger cue for thematic role assignment and whether the combination of cues (action plus wiggle) further influences thematic role assignment. Disambiguating case-markers were not available until the end of the sentence (case-marked NP2). The visual scene did not disambiguate thematic role relations before the visual cue occurred in the verb region. Participants' eye-movements towards the agent or the patient in the scene were recorded while they listened to ambiguously case-marked German OVS sentences.

Eye-Movements. The results corroborated the rapid visual context effects reported by Zhang and Knoeferle (2012) and Münster (2016). Their results revealed that both children and young adults directed more looks to the target character when actions were depicted (vs. not depicted). However, in both studies the effects of depicted actions were delayed by one word region in children (adverb region) compared to adults (verb region). The results from Experiment 3 did not corroborate these findings. For both adults and children, the effects of action, wiggle, and action plus wiggle emerged immediately (during the verb region). We did, however, observe some delay in the preference to inspect the target character (the agent). Descriptively, for adults, cue presence resulted in a higher preference to look at the agent (vs. patient) during the verb region than the adverb region. Children's preference to inspect the agent (vs. patient) was higher in the adverb region (vs. verb region). These differences in the time course of the gaze-pattern likely resulted from the presentation differences. In our study, presentation of the visual cues was limited to the lifetime of the verb of the sentence. The cues had a sudden onset and this dynamic likely boosted attentional responses. By contrast, Zhang and Knoeferle (2012) and Münster (2016) presented the depicted actions throughout the sentence.

In addition to extending effects of depicted actions, the data revealed a significant effect of the wiggle and an interaction of action and wiggle, all during the verb and adverb region. Children's action-based anticipatory looks to the agent (vs. patient) were not modulated by wiggle presence. In the absence of depicted actions, children inspected the agent (vs. patient) more when the wiggle was present (vs. absent). Effects of action (but not of wiggle) extended to the NP2 region, likely because the action disappeared but the target character (the wiggling agent) did not, increasing the memory load for depicted actions (see also Richardson & Spivey, 2000 and Richardson & Kirkham, 2004 for a discussion). Time Course Graphs further indicated late sentence disambiguation: More looks to the agent (vs. patient) in the no action no wiggle condition compared to the other conditions. Perhaps the mention of an unexpected character (the agent and not the previously anticipated patient) influenced the increase in looks to the target in the no-cue baseline condition.

Accuracy. For adults, we observed a marginal interaction of action and wiggle on accuracies such that when actions were depicted, wiggle presence resulted in higher accuracies. By contrast, wiggle presence did not elicit more correct responses to post-sentence comprehension questions when actions were not depicted. Overall, adults' accuracy data was surprisingly low (68.9%). Case-marking on the second noun-phrase did, in principle, disambiguate the OVS sentences. In Experiment 2b, we did not find such a low percentage of correct responses for ambiguous OVS sentences (92.4%; Figure 9.18). The difference in the number of correct responses between the two experiments likely resulted from the difference in materials. In Experiment 3, we used the same fillers for adults as in Experiment 2b, however, accuracies might have been influenced by the language-scene mismatches in Experiment 3. Although we used these mismatches to avoid recognition of the experimental goal, they may have influenced participants' interpretation. Case-marking on the second noun-phrase of an OVS sentence might have been considered as a mismatch, resulting in an SVO interpretation (i.e., incorrect role assignment and responses to post-sentence comprehension questions).

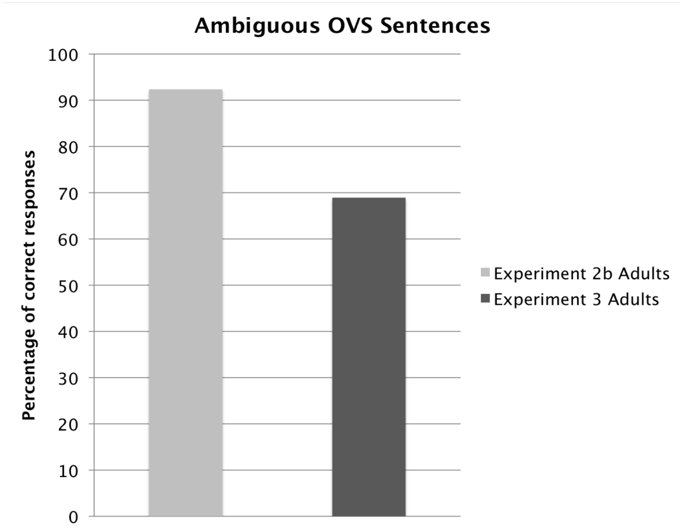


Fig. 9.18: *Experiment 2b adults and Experiment 3 adults: Comparison of accuracy data for ambiguous OVS sentences.*

For children, we did not replicate the improved accuracy when actions were present (vs. absent) as reported by Zhang and Knoeferle (2012) and Münster (2016) although we used similar comprehension questions about 'who does what to whom'. Descriptively, however, our child accuracy data revealed a similar pattern as the adult accuracy data but neither the action nor the wiggle exerted clear effects on accuracies (which were overall also lower than adults'). We did, however, find a clear effect of voice such that children correctly responded significantly more often to active voice compared to passive voice comprehension questions.

Children might have used the relationship between the visual cue representation and the comprehension question: The correct answer to active voice questions was always the agent. Since the agent was either wiggling or performing an action, children may have used this information to answer active voice questions. Following from the reliable effect of voice, we analysed the active data separately from the passive data. In the active voice data, we observed a marginal effect of wiggle. Wiggle presence had a positive effect on accuracies when post-sentence comprehension questions were in active voice.

Since the depiction of the visual cues was temporally limited, children may have kept the more salient cue (the wiggle) in working memory, facilitating its later access during comprehension questions.

However, the eye-gaze data did not corroborate the idea of the wiggle being the more salient cue boosting visual attention. During the adverb region, children's preference to inspect the agent (vs. patient) was higher when the action (vs. the wiggle) was present. Wiggle presence, perhaps, increased the agents' saliency or focus more than action presence, with some delay only, in turn eliciting increased accuracies (when the agent was the question target) but not more inspections of the target character in real-time.

One question, however, remains: Why did we not replicate the boost of actions on accuracies described in Zhang and Knoeferle (2012) and Münster (2016)? One reason for why we failed to observe clear effects of depicted actions might be the temporal limitations of the visual cues in our study. Another possible explanation might be the ambiguity in our sentences. Since children, in general, have a strong SVO word order bias (Dittmar et al., 2008a), understanding role relations in such difficult-to-process sentences may require a full command of their attention. Upon hearing a locally structurally ambiguous OVS sentence, children likely need to restructure the spoken utterance from their initial (and preferred) SVO sentence interpretation. But, if children at the age of five cannot yet use case-marking for thematic role assignment, ambiguity should not have influenced our results, since they need to restructure from their initial SVO interpretation regardless of whether the OVS sentences were ambiguous (ambiguously case-marked) or unambiguous (clearly case-marked). Arguably, it is more likely that the temporal limitation of the depicted actions and not the ambiguity of the sentence caused the differences between our and the results by Zhang and Knoeferle (2012) and Münster (2016).

Descriptively, the passive voice comprehension data did, however, corroborate the findings by Münster (2016) who found an effect of depicted actions in the passive voice comprehension data. The argument was that passive voice comprehension questions are more difficult for children than active voice comprehension questions (i.e., it is more difficult to assign thematic roles). Thus, children used depicted actions to facilitate language comprehension. Since the depicted actions in our study were temporally limited (unlike in Münster, 2016 where the actions were present throughout the sentence), children may have processed the action in less depth, eliminating clear effects on passive voice accuracies.

9.4 SUMMARY AND CONCLUSION

In summary, the results of Experiment 3 revealed significant effects of both visual cues (action and wiggle) and a reliable interaction (all ps < .001; Tables 9.3, 9.4, 9.6, 9.7). The sudden onset of the visual cues likely boosted attentional responses. However, we also observed some differences between children and adults: a) agent preference was higher during the verb for adults and during the adverb for children (Figures 9.7, 9.3), b) during the verb, wiggle presence influenced adults' agent preference more than children's (Figure 9.17a), c) children inspected the agent (vs. patient) more than adults when actions were depicted (vs. not depicted; Figure 9.17), and d) the combination of cues facilitated language comprehension in adults more than one cue (Figure 9.4). For children wiggle presence had beneficial effects on accuracies (when questions were in active voice) and action presence slightly increased accuracies when questions were in passive voice (Figures 9.11, 9.12).

Part IV

DISCUSSION

10

GENERAL DISCUSSION

In six eye-tracking studies, we investigated the influence of prosody, case-marking, and visual cues on thematic role assignment during real-time language comprehension (in five-year-old children in comparison to young adults in Experiments 1 and 3 and young adults only in Experiment 2). We monitored participants' eye-movements whilst they listened to unambiguous SVO and OVS sentences either with biasing or neutral prosodic contours (Experiment 1), unambiguous SVO and OVS sentences (Experiment 2a), unambiguous and ambiguous OVS sentences (Experiment 2b) either with SVO- or OVS-biasing prosodic contours (Experiment 2), and ambiguous OVS sentences with an OVS-biasing prosodic contour (Experiment 3). Visual scenes depicted action ambiguously (Experiment 1 and 2) or visual cues (action, wiggle, or both) were present during the verb region of the sentence.

In Chapter 2 of this thesis, we provided evidence for the interplay of language comprehension and the visual world and highlighted how this bi-directional relationship became acknowledged among psycholinguistic researchers (e.g., Cooper, 1974; Knoeferle et al., 2005; Tanenhaus et al., 1995). Additionally, we discussed evidence for the incrementality of language processing (Allopenna et al., 1998; Sedivy et al., 1999; Tanenhaus et al., 1995). Adult comprehenders are able to rapidly integrate visual and linguistic information to interpret spoken utterances (Chambers et al., 2004; Knoeferle et al., 2005; Sedivy et al., 1999; Tanenhaus et al., 1995). Visual attention is often anticipatory such that comprehenders inspect upcoming referents before its mention on the basis of a verb's selectional restrictions (Altmann & Kamide, 1999). It seems, however, that some pragmatic processes are delayed (Huang & Snedeker, 2009b).

The next chapter (Chapter 3) then highlighted how different cues such as case-marking, prosody, and visual cues can facilitate incremental thematic role assignment and discussed the influence of more than one cue on language comprehension. The evidence provided by these studies, however, came from adult participants. Existing findings on child in comparison to adult language comprehension have shown some similarities (e.g., Arnold et al., 2007; Johnson & Huettig, 2011; Mani & Huettig, 2012) and also some differences (e.g., Dittmar et al., 2008a; Schipke et al., 2011). These differences seem to be especially pronounced in the pragmatic domain during online language comprehension (Huang & Snedeker, 2009b; Trueswell et al., 1999). We then pointed towards the issue that many studies on child language comprehension have relied on offline measures (e.g., Dittmar et al., 2008a; Meroni & Crain, 2003). Thus, the current debate confronting us was the puzzle of how children actually compare to adults during real-time language comprehension. Eye-tracking and the visual-world paradigm measure online language comprehension and were thus suitable to directly compare the time course of language processing in children and adults. Post-sentence comprehension questions further revealed the outcome of the comprehension process.

The first question was whether one main difference between child and adult language comprehension is that children cannot draw pragmatic inferences yet (as suggested by Trueswell and colleagues, 1999) and whether these are delayed in adults. In Experiment 1, we used prosody to investigate this question since we know from previous research that children used prosody offline (e.g., for thematic role assignment when case-marking was unambiguous; Grünloh et al., 2011). However, children's, unlike adults', ability to use case-marking for thematic role assignment so far revealed contradictory results. Some studies reported that children at the age of five cannot yet use case-marking for correct thematic role assignment (Dittmar et al., 2008a; Schipke et al., 2011) whereas others found that four- to -five-year-old and six-year-old children are very well able to use case-marking in online supportive visual contexts (i.e., contexts which include world knowledge; Özge et al., 2016) and offline (during a video-pointing task; Brandt et al., 2016). What we did not yet know, was whether four- to -five-year-old children can use case-marking during online language comprehension of unambiguously case-marked SVO and OVS sentences when the visual context is not supportive but ambiguous (Experiment 1).

The literature review further pointed out that adults can rapidly exploit case-marking in unambiguous sentences and prosody in ambiguous sentences during online language comprehension. Thus, the next question we tried to answer was whether adults rapidly exploit case-marking and/or prosody when the contrast between the two prosodic structures is clear (i.e., either SVO- or OVS-biasing), or whether case-marking is a stronger cue for thematic role assignment/ affects visual attention/language comprehension more than prosody. Additionally, we tried to replicate Weber et al.'s (2006) findings with ambiguous sentences but in comparison to clearly case-marked sentences (Experiment 2).

Children were not able to use referential context for syntactic disambiguation (Trueswell et al., 1999) but used other visual information (i.e., depicted actions which do not require pragmatic inferences) to correctly assign thematic roles even in non-canonical OVS sentences (Münster, 2016; Zhang & Knoeferle, 2012). Sentences in these studies were, however, clearly case-marked. What remained unclear was whether children can use depicted actions in ambiguously case-marked OVS sentences. In addition to the depicted actions, we investigated whether a wiggling target character (a focusing / pragmatic cue) influences thematic role assignment. Comprehenders again had to generate pragmatic inferences (i.e., establish a contrastive interpretation). Evidence provided in chapter 3 suggests that different cues can have similar, different, or combined effects on language comprehension. We thus combined the two cues, action and wiggle, to investigate the extent of distinct visual cues on visual attention and language comprehension.

Our studies make new contributions to the existing literature, especially because a) we used an online measure (eye-tracking), b) we directly compared children with adults, and c) post-sentence comprehension questions verified sentence interpretation. In the following sections, we describe and discuss the effects of different types of cues (case-marking, prosody, and visual cues) on thematic role assignment during real-time language comprehension. We recapitulate the main results from prior research and then point out our findings for adults and children. Finally we describe the main similarities and differences between children's and adults' visual attention and language comprehension and discuss how the differences can be implemented in the Coordinated Interplay Account (Knoeferle & Crocker, 2006, 2007; Knoeferle et al., 2014) which highlights the bi-directional relationship between visual attention and utterance comprehension. Additionally, we propose that scene mediated attention should be an aspect included in the CIA. In the last section, we summarise the main implications of our findings and suggest ideas for future research.

10.1 CASE-MARKING

In German, thematic roles can be distinguished via case-marking. Prior research has shown that for adults, case-marking is a strong cue for online thematic role assignment (Kamide, Scheepers, & Altmann, 2003; Matzke et al., 2002). Results for children, however, were contradictory (Dittmar et al., 2008a; Schipke et al., 2011 but Brandt et al., 2016; Özge et al., 2019, 2016). Only when the visual context supported 'who does what to whom', were four-to five-year-old children able to use case-marking for correct thematic role assignment. One of our questions was whether case-marking effects extend to non-supportive visual contexts. We expected adults to rapidly recruit case-marking and reach ceiling in accuracies. Children might need visual support and thus not use case-marking. Our scenes provided role relations ambiguously such that two role fillers performed identical actions. Furthermore, we avoided stereo-typicality in our scenes. The visual scenes provided a context for both SVO and OVS sentences but did not disambiguate role relations early in the sentence.

10.1.1 *Adults*

In line with previous findings, adults rapidly recruited case-marking for correct thematic role assignment. The results of Experiments 1 and 2a revealed a significant effect of word order: From the verb region onwards, participants inspected the patient (vs. agent) more in SVO sentences and the agent (vs. patient) more in OVS sentences. We did not find a difference in the onset of looks towards the target character between SVO and OVS sentences. In both sentences structures, participants started to inspect the target character at the beginning of the verb region, indicating no difference in processing the two sentence structures. However, in Experiment 1 and 2a, adults inspected the agent (vs. patient) more in OVS sentences than the patient (vs. agent) in SVO sentences. We argued that the salience of the agent (who performed an action) in comparison to the patient (who did not perform an action) resulted in this gaze-pattern. Another possibility is that participants inspected the agent (vs. patient) more to corroborate their interpretation given that OVS sentence structure is less preferred and more difficult to process.

Crucially, unlike previous findings by Kamide, Scheepers, and Altmann (2003) and Zhang and Knoeferle (2012), participants started to inspect the target character almost immediately after the first noun-phrase (from the beginning of the verb region). In Kamide, Scheepers, and Altmann (2003), effects of case-marking were observed during the adverb region although the scenes provided world knowledge about 'who does what to whom'. The authors argued that morpho-syntactic information was rapidly integrated with the verb's semantic constraints. Perhaps this is why anticipatory looks only followed after the verb because the verb's constraints influenced case-marking and thematic role assignment (i.e., world knowledge could only permit thematic role assignment after the verb: Hares eat cabbages and foxes eat hares).

Alternatively, early anticipatory looks to the target character in Experiments 1 and 2a resulted from the depiction of actions. Kamide, Scheepers, and Altmann (2003) did not depict actions. Similar to Kamide, Scheepers, and Altmann (2003), Zhang and Knoeferle (2012) reported effects of case-marking during the adverb region when no actions were depicted. Perhaps the mere fact that the agent was depicted as performing an action and the patient was not permitted participants to rapidly anticipate the target character: Visual information was rapidly integrated with case-marking and knowledge about thematic roles (i.e., agents perform actions and patients are being acted upon).

It seems that case-marking is a very strong cue for thematic role assignment even if the visual context is ambiguous. Bear in mind that our scenes depicted actions ambiguously - two animal characters performed identical actions. Our results thus add nicely to previous findings: Not only can adults use case-marking for thematic role assignment during reading and in supportive visual contexts during comprehension, they can also use case-marking when the visual context is not supportive but ambiguous. Knowledge about thematic roles combined with visual information and unambiguous case-marking seem to influence early target anticipation. For case-marking, ours and previous findings point towards the same idea: Case-marking is a strong cue for thematic role assignment in adult comprehenders (also corroborated by the accuracy data in our experiments).

As we expected, the overall accuracies were very high for adults in Experiment 1 and also in Experiment 2a. Case-marking on the first and on the second noun-phrase pointed towards either SVO or OVS sentences structure. The comprehension data in Experiments 1 and 2 corroborated the eye-tracking data such that anticipatory looks to the patient (vs. agent) resulted in correct responses for SVO sentences and anticipatory looks to the agent (vs. patient) in correct responses for OVS sentences. However, we found slightly fewer correct responses in unambiguously case-marked OVS sentences in Experiment 2a and 2b in comparison to Experiment 1. One of the reasons for this might be the number of SVO sentences in the fillers in Experiment 2. Experiment 1 contained eight fillers, one of which was structured SVO. In Experiment 2, 48 filler sentences were SVO sentences (or coordinated SVO) and only 24 fillers had a different sentence structure. Thus, participants were likely influenced by the number of SVO sentences throughout the experiment. Perhaps they thought SVO sentences were more likely than OVS sentences which then resulted in slightly fewer correct responses.

10.1.2 *Children*

Children in our first experiment did not rapidly recruit case-marking, although sentences were clearly case-marked for either SVO or OVS. In both sentence structures, the results revealed more looks to the patient (vs. agent), indicating an SVO interpretation of the sentences. The accuracy data corroborated the eye-tracking data. Children correctly responded more often to post-sentence comprehension questions for SVO than for OVS sentences. For OVS sentences, response accuracy was below chance level. More often than not, children interpreted OVS sentences as SVO sentences.

Our accuracy results are in line with the offline data reported by Dittmar and colleagues (2008): Response accuracy for OVS sentences was at chance level. In Münster (2016), Zhang and Knoeferle (2012), and Özge et al. (2016), German children correctly interpreted non-canonical (OVS) sentences but all these studies included information (in addition to case-marking) about 'who does what to whom' in the visual scenes (via depicted actions or via world knowledge). Our scenes, however, avoided stereo-typicality and actions were depicted ambiguously. Children likely need additional information in order to correctly assign thematic roles in non-canonical OVS sentences. Since SVO is the preferred sentence structure over OVS (Dittmar et al., 2008a; Hemforth, 1993; Hemforth & Konieczny, 2013), children's response accuracies in SVO sentences were very high. It seems that they used word order (SVO), instead of case-marking, as a cue for thematic role assignment, corroborating the findings by Gertner et al. (2006) and Dittmar et al. (2008b). English and German children used word order as cue for thematic role assignment in transitive sentences. Perhaps children's knowledge about thematic roles is underspecified (influenced by SVO word order) at the age of of five.

Unlike in Münster (2016), we did not find a difference between active and passive voice comprehension questions. This is interesting because, unlike in Münster (2016), we depicted actions (although ambiguously) during the comprehension questions. In Münster (2016), the depicted actions were not present during the questions. Passive constructions seem to be difficult for young children in general since they do not only need to understand who the agent (the actor) but also who the patient (the acted upon) is (Sinclair, Sinclair, & De Marcelus, 1971). Maybe children need some sort of visual support (even if ambiguous) to answer passive voice comprehension questions.

Interestingly, in OVS sentences active voice comprehension questions resulted in slightly higher response accuracy than passive voice comprehension questions. Since non-canonical OVS sentences are difficult-to-process for children at the age of five, questions in passive voice likely added difficulty to the task, resulting in lower response accuracies. Similar arguments were made by Münster (2016). She argued that action depiction influenced response accuracies in passive voice comprehension questions because task difficulty facilitated the use of depicted actions for comprehension.

But why did children respond correctly for some OVS sentences? It seems that children might have used the depicted actions as a support even though they were ambiguous. When a child heard the sentence *The bear pushes...*, looks to the patient indicate they interpreted the sentence as SVO, expecting the worm/patient to be uttered next. Then they did, however, hear *bull* instead of the anticipated *worm* at the end of the sentence. The way in which the actions are depicted (agent+action->ambiguous+action->patient) likely helped children to determine 'who does what to whom'. Another possible explanation is the number of double case-marked OVS sentences (via case-marking on the determiner and suffixation on the first noun). We observed slightly more correct responses for those sentences than for the ones which were only case-marked on the determiner. It seems that this additional information somewhat helped children to determine 'who does what to whom'.

It is also possible that children have difficulties in restructuring the linguistic input and thus correctly responded only to some comprehension questions for OVS. Since SVO is the preferred sentence structure (Dittmar et al., 2008a; Hemforth, 1993; Hemforth & Konieczny, 2013) and the eye-movement results corroborate the idea of an initial SVO interpretation (more looks to the patient in both SVO and OVS sentences), it may be difficult for children to restructure towards OVS, especially when the visual context is ambiguous. Upon hearing an unexpected agent (instead of the anticipated patient), children might need more time to restructure the spoken utterances. Meroni and Crain (2003) made similar arguments for PP-attachment ambiguities. Although offline, they argued that children failed to use referential context because they did not have enough time to revise their initial parsing preferences. This might also hold true for our experiment. Children listened to the sentence and 1500 ms after the sentence, the comprehension question followed. Maybe children's accuracies were low for OVS sentences because they did not have enough time to revise their initial SVO sentence preference.

Another reason why children struggled to correctly interpret OVS sentences might be that OVS sentence structures rarely occur (only in 21%) in child directed speech (Dittmar et al., 2008a). Prior research suggests the importance of statistical learning for language comprehension (e.g., Mitchell & Cuetos, 1991; Saffran, 2003; Saffran et al., 1996; Wells et al., 2009). Trueswell and colleagues (1999) also argued for the the influence of statistical learning (the verb *put* occurs more often in a destination than in a location context). Thus, children in their study might have used lexical information rather than visual information for sentence interpretation. Snedeker and Trueswell (2004) supported this idea by using different verbs in similar sentences. Their results showed that lexical bias influences child language comprehension. If statistical learning plays a role for thematic role assignment, then the fact that OVS sentences are very rare in child directed speech likely influenced children's strong SVO word order bias (agent-verb-patient ordering).

According to early abstraction accounts (Section 4.2.1), children should have been able to link case-marking to thematic roles (early grammatical abstraction). However, unlike in Özge et al. (2016), children did not use case-marking in ambiguous action scenes. Perhaps the absence of world knowledge in the visual scenes (or the presence of it in Özge et al., 2016) has an influence on early abstraction of grammar. If this was the case, early abstraction accounts would have to revised accordingly. Yet, more research is needed to investigate this issue further.

It remains a puzzle, however, why children sometimes do not use case-marking during language comprehension and production (Dittmar et al., 2008a; Kany & Schöler, 2007; Maiworm, 2008; Motsch & Rietz, 2016; Popella, 2005; Schipke et al., 2011) and other times are able to comprehend and produce case-marking (nominative/accusative) in and adult-like fashion (Brandt et al., 2016; Eisenbeiss et al., 2006; Özge et al., 2019, 2016; Szagun, 2004). It is possible that children did not use case-marking in our study because we did not put any stress or emphasis on the case-markers which is typically not the case for German: In prenominal position determiners are unstressed (Jones & Jones, 2019; Szagun, 2004). Szagun (2004) and Brandt et al. (2016) argued that children may not pay much attention to the unstressed determiner, hence, the case-marking. Szagun (2004) further suggested that children are likely to make errors for the accusative and dative case-markers *den* and *dem* because they sound very similar. Now, considering that *der* can be clearly distinguished from *den* and *dem* but the latter two cannot, it might be easier for children to use nominative case-marking because accusative case-marking is less clear or because of the nominative/agent-first bias in German (Dittmar et al., 2008a; Hemforth, 1993; Hemforth & Konieczny, 2013). Kamide, Scheepers, and Altmann (2003) proposed a similar argument for why their data did not reveal a significant effect of case-marking in SVO (vs. OVS) sentences. The authors argued that accusative case may have been misheard as nominative case because case-marking was not acoustically stressed.

Taken together, case-marking is a strong cue for thematic role assignment in adults which rapidly mediated visual attention towards target role fillers as early as the verb. Information about thematic roles in general seems to be rapidly integrated with case-marking. For five-year-old children, case-marking could not mediate early target anticipation or sentence interpretation, word order (SVO) did.

10.2 prosody

In child and adult language comprehension, prosody can be used as a cue to distinguish 'who does what to whom'. For adults, online findings revealed that prosody can overwrite participants' strong SVO bias (Weber et al., 2006). Listeners used the nuclear stress on the first noun-phrase to disambiguate towards OVS. For children, prosody was only helpful in combination with unambiguous case-marking and offline. When case-marking was ambiguous, they fell back on their strong SVO word order bias (Grünloh et al., 2011). We thus investigated the following questions: Can children (and adults) use prosody for correct thematic role assignment in unambiguous sentences during online sentence comprehension? Or do they fail to compute pragmatic inferences? And do adults use prosody when case-marking is ambiguous?

10.2.1 *Adults*

The results reported in this thesis do not corroborate the influential role of prosody on thematic role assignment. We expected prosody to influence the amount and timing of looks to the target role filler but the results of Experiment 1 and Experiment 2 did not reveal significant effects of prosody. Participants anticipated the target character as a result of case-marking but prosody did not contribute to reliably more or earlier looks to the target role filler.

However, in the adverb region of Experiment 1, the preference to look at the target character was higher in the biasing prosody conditions compared to the neutral prosody conditions. Although these effects were not reliable, prosody did somewhat influence looks to the target character. Accentuation (vs. no accentuation) may have influenced listeners' visual attention: When the sentence carried some prosodic marking, participants inspected the target character slightly more, perhaps because accentuation facilitated longer/more fixations. Interestingly, in Experiment 2a the preference to look at the target character was slightly higher in the *non-biasing* prosody conditions (SVO sentences carrying an OVS-biasing prosody and OVS sentences carrying an SVO-biasing prosody) than in the biasing prosody conditions. The fact that the non-biasing prosody elicited slightly more target inspections may have resulted from our design: We depicted actions (unlike Weber et al., 2006), resulting in a different influence of the prosodic structures compared with the findings in Weber et al. (2006).

One of the reasons why we did not find reliable effects of prosody in unambiguously case-marked sentences (Experiment 1, 2a, 2b) might be the fact that case-marking affects visual attention/language comprehension more than prosody. Similar arguments were made by Sedivy et al. (1999) who argued that adjectival modifiers result in contrastive interpretation in general and thus prosody was not necessary for reference resolution. Similarly, if case-marking results in agent/patient anticipation and in SVO/OVS interpretations, then prosody does not have additional beneficial effects.

However, if it was true that case-marking has more pronounced effects on visual attention/language comprehension than prosody, we should have observed reliable effects of prosody in ambiguous sentences (Experiment 2b) but we did not. The SVO-biasing prosody did, however, result in a slight preference of looks towards target character (the agent) during the adverb region. Although we expected the OVS-biasing and not the SVO-biasing prosody to influence looks to the agent (vs. patient), it seems that adults are somewhat sensitive to prosody when case-marking is ambiguous and cannot be used as a cue to thematic role assignment. One of the reasons for this slight agent preference in the SVO-biasing prosody condition might be the ambiguity in the scenes. Recap that there were two animal characters performing the same action. The middle character was role ambiguous and was always mentioned at the beginning of the sentence. The additional stress on the verb might have influenced participants to look at the other character that performs an action (the agent).

Another possible explanation for why we, unlike Weber et al. (2006), did not find reliable effects of prosody in Experiment 2b (ambiguous OVS sentences) is the difference in materials. In Weber and colleagues' (2006) study, the visual scenes included world knowledge. Their scenes depicted, for example, a cat, a dog, and a bird. Cats are likely to chase birds as much as cats are likely to be chased by dogs. Our scenes did not include world knowledge or stereo-typicality. Bears are not very likely to push worms, and similarly, bears are not likely to be pushed by bulls. Maybe comprehenders need some sort of world knowledge support in the visual scene to use prosody as an effective cue for thematic role assignment. Additionally, Weber et al. (2006) did not depict actions whereas we depicted actions ambiguously. The additional stress on the verb may have had a different function in ours compared to Weber et al.'s (2006) study. When actions are not displayed the additional stress on the verb can refer to new information whereas when actions are depicted ambiguously, it might put focus on the other character which also performs an action (the agent instead of the patient) because the other character performing an action was already mentioned at the beginning of the sentence. Furthermore, we compared the biasing prosody conditions to non-biasing prosody conditions. In Weber et al. (2006) no such comparison was conducted. Participants either heard SVO sentences with an SVO-biasing prosody, or OVS sentences with an OVS-biasing prosody. The addition of a clear contrast between biasing and non-biasing prosody might actually give a clearer idea of the effects of prosody on thematic role assignment since it shows what happens when a sentence structure biasing prosodic structure is assigned (SVO-SVO biasing, OVS-OVS biasing) in a direct comparison to when prosody biases towards a different sentence structure (SVO-OVS biasing, OVS-SVO biasing).

At the beginning of the adverb region in Experiment 1 and the end of the adverb region in Experiment 2a, we observed a decrease in the preference to look at the patient (vs. agent) in SVO sentences such that participants looked almost equally as much at the agent and the patient. This equal preference likely resulted from the prosodic structure. Remember that SVO sentences were assigned an L*+H accent on the first noun-phrase and an H* accent on the verb. The main accent on the verb might have influenced people to look at the other character performing an action (the agent). All in all, it seems that adults are somewhat sensitive to prosody.

10.2.2 *Children*

For children, we also did not find reliable effects of prosody. We previously described that children did not use case-marking, resulting in more looks to the patient (vs. agent) in all four conditions. However, similarly to adults, children looked more at the patient (vs. agent) in the biasing compared to the neutral prosodic conditions (during the adverb region). It seems that children, similarly to adults, are also somewhat sensitive to prosody but instead of using prosody to correctly assign thematic roles in OVS sentences, both biasing prosodic structures increased children's preference to look at the patient (vs. agent).

Unlike Grünloh et al. (2011), we did not find effects of prosody on thematic role assignment in unambiguously case-marked sentences. The difference between our and Grünloh and colleagues (2011) findings likely resulted from the difference in materials. Recap that Grünloh et al. (2011) used an offline video-pointing task. Children watched two video clips in which two animal characters performed novel actions. These videos were then followed by two static images of the scenes in which agent and patient roles were reversed. Thus, the stimuli provided a direct contrast between the two possible sentence interpretations. Our stimuli provided information ambiguously (i.e., both possible interpretations depicted within one image). The ambiguity in our visual scenes and the fact that we recorded online measures likely influenced our results.

Arnold (2008) found that prosody is an effective cue for reference resolution but their prosodic manipulations were not restricted to accent vs. no accent. The target word was followed by an intonational phrase break and in addition other words in the utterance were longer in duration. The authors argued that listeners may have used all or only one of these cues. It thus remains unclear whether it was actually the accentuation alone that influenced reference resolution or whether the other prosodic means contributed to an effect of prosody on reference resolution. Maybe accentuation alone is not a strong cue after all. Additionally, reference resolution involves different processes than thematic role assignment which could have further contributed to the null effects of prosody in the experiments reported in this thesis.

Similarly to adults in Experiment 1 and Experiment 2a, children's preference to look at the patient decreased during the adverb region. For adults, the same tendency was found at the beginning at the adverb region, for children throughout the adverb region. It seems that prosody did slightly influence visual attention (via the main accent on the verb and the depicted actions).

It is also worth mentioning that for both age groups in Experiment 1 and for adults in Experiment 2, the time course graphs indicate a slight preference to look at the agent (vs. patient) during the first noun-phrase. In Experiment 3, looks to the agent (vs. patient) are rather at chance. It is possible that this preference derived from the design of our materials: The role filler of the agent always performed an action whereas the role filler of the patient did not. Thus, the agent might have been slightly more salient than the patient which possibly attracted attention during the first noun-phrase. In contrast, no actions were depicted during the first noun-phrase in Experiment 3. Note, however, that participants mostly looked at the ambiguous role filler during NP1 (Figure 9.16). Hence, the overall number of looks towards the agent and the patient is very low and thus not very representative overall.

Prosody may not be as straightforward for thematic role assignment as, for example, case-marking which may make prosody altogether a more subtle cue. For instance, the realisation of different prosodic structures (e.g., contrastive stress) varies a lot among language producers such that different speakers use different prosodic structures to express contrast (Grice et al., 2017). It might thus be difficult for comprehenders to rapidly integrate one type of prosodic structure for one type of sentence structure. Weber et al. (2006) also argued that the prosodic structure they used in their experiment might not be the only structure that biases towards an SVO or OVS interpretation. Since we have seen that participants are somewhat sensitive to prosody, it was either the prosodic structure we used or the fact that prosody includes too much variation which led to null effects of prosody during online thematic role assignment.

In section 3.2, we briefly described the variability in the use of prosody. Perhaps the fact that prosody can have many different functions for language comprehension and production has led to very subtle effects in our experiments, likely because prosody is altogether a more subtle cue. The person recording the audio files was phonetically trained which resulted in a clear contrast between the prosodic conditions. However, to confirm that the different prosodic contours were realised accurately, a phonological evaluation by experts would likely have been helpful. We can thus not exclude the possibility that the realisation of the prosodic contours was not as clear-cut as we thought (since we did not obtain independent ratings that corroborated a perceptible contrast between the prosodic conditions). If independent ratings had corroborated a perceptible contrast, we could be more certain that the weak prosodic effects for real-time language processing were not a result of the absence of a perceptible prosodic contrast but a result of the role prosody plays on language processing and visual attention in general.

It is also possible that regional aspects of prosody additionally influenced our results. In German, intonation patterns differ among different dialects (Dafydd, 1998). Since the adult participants were all university students, their origins and likely their dialects and typical intonation patterns differed. Descriptively speaking, we observed individual differences in looks to the agent or the patient such that some participants were influenced differently by the prosodic structure than others (Figure 8.4). Maybe this is another reason why we did not find reliable effects of prosody. Individual differences possibly cancelled out any clear effects of prosody. Although individual differences play a role in most eye-tracking studies, in the domain of prosody, these differences may be especially pronounced.

All in all, it seems that prosody is a subtle cue for incremental thematic role assignment when scenes do not support role disambiguation. Across both age groups, we observed only subtle and non-reliable effects of prosody on target anticipation and no clear effects on utterance comprehension. Relating one prosodic contour to one sentence structure may be problematic because prosody contains a lot of variation among speakers and listeners. Whether children (and adults) failed to compute pragmatic inferences, remains an open question at this point.

10.3 VISUAL CUES

Adult comprehenders can rapidly recruit visual information (referential context, contrast between objects, depicted actions and events) for the interpretation, syntactic structuring, or thematic role assignment of spoken utterances (Chambers et al., 2004; Knoeferle et al., 2005; Tanenhaus et al., 1995). Five-year-old children, on the other hand, did not use the referential context to disambiguate PP-attachment ambiguities, suggesting they cannot yet compute pragmatic inferences online (Trueswell et al., 1999). However, it has been suggested that other visual information helps children to correctly assign thematic roles. Children used depicted actions to correctly interpret unambiguous German SVO and OVS sentences (Münster, 2016; Zhang & Knoeferle, 2012) and case-marking in combination with world knowldge for thematic role assignment (Özge et al., 2016). We thus investigated whether visual information can help children to overcome their SVO word order bias in ambiguous OVS sentences when a) no pragmatic inference is (depicted actions) or b) pragmatic inferences (a wiggling target character) are required. Is one cue perhaps stronger than the other? Since visual cues seem to have similar, different, or combined effects on language comprehension (Knoeferle et al., 2007; Kreysa et al., 2014; Münster et al., 2015; Staudte et al., 2014), we further examined whether two cues (action plus wiggle) are better than one (action or wiggle).

10.3.1 *Adults*

In Experiment 3, we observed reliable effects of action, wiggle, and a reliable interaction of action and wiggle during both regions of interest (verb and adverb). Participants inspected the agent (vs. patient) significantly more when visual cues were present compared to when they were absent (no cue condition). When the action was present (vs. absent), wiggle presence influenced looks to the agent. The visual cues boosted attentional responses which is not surprising since the visual cues had a sudden onset which likely attracted listeners attention towards the target role filler (the agent).

For accuracies, we observed a marginal interaction of action and wiggle. When the action was depicted, an added wiggle (vs. no wiggle) elicited more correct responses to post-sentence comprehension questions. By contrast, wiggle presence did not influence accuracies when no action was depicted. Overall, the accuracies were highest for the combination of cues. Descriptively, a similar pattern can be observed in the eye-tracking data (Figure 9.3). In the adverb region, the agent preference is highest in condition d) one action one wiggle. Although post-hoc tests did not reveal reliable effects, the time course data supports the idea that participants inspected the agent (vs. patient) more in condition d) one action one wiggle compared to the other visual cue conditions (conditions b) and c); Figure 9.2) resulting in more correct responses to post-sentence comprehension questions.

The seemingly combined effect of action and wiggle on eye-movements differs from the findings by Kreysa et al. (2014) who reported that speaker gaze and depicted actions have similar but not combined effects on language processing. Perhaps the nature of the two cues (wiggle vs. speaker gaze) differs substantially, resulting in different effects on language processing. Although both cues are not mediated by language (e.g., unlike the depicted action which is mediated by the verb), speaker gaze is triggered by another entity (a speaker) which guides listeners (visual) attention towards a potential referent. The wiggle, on the other hand, occurs right at the location of the target character (the agent) which is then relevant for developing expectations (i.e., when a role filler has not yet been mentioned, the wiggle already attracts the listeners' attention to the relevant role filler - the agent). Descriptively, it seems that the combined action and wiggle guided listeners' attention more towards the target role filler than one action or one wiggle only. Speaker gaze and action depiction did not result in more looks to the target character than speaker gaze or depicted actions alone. Overall, it seems that the wiggle in combination with depicted actions facilitates language comprehension differently than speaker gaze in combination with depicted actions.

It is, however, also possible that the two cues (speaker gaze vs. wiggle) have different or similar effects on language processing in combination with depicted actions because Kreysa et al. (2014) used unambiguous SVO sentences and we used ambiguous OVS sentences. Not only are OVS sentences non-canonical, the ambiguity likely requires restructuring from an initial SVO interpretation. Thus, the processes involved differ and might have led to different effects of the different visual cues (speaker gaze vs. wiggle). Additionally, the two cues differed in their timing. The wiggle was only present throughout the lifetime of the verb, whereas the speaker gaze shift towards the target character started at verb onset and the gaze remained on the target character until sentence offset. Again, it is possible that the temporal limitations of the visual cue depiction in our study resulted in different effects compared to other visual cues which are not temporally limited.

However, in Staudte et al. (2014), speaker gaze was found to have similar effects on language comprehension as an arrow pointing towards the target object. The authors argued for a *Visual Account* and not a *Intentional Account* since they found no reliable differences when participants remapped gaze cues or arrow cues. When the two cues had similar timings, effects on language comprehension of gaze and arrow cues were similar. In our study, the timing of the cues was also identical such that the wiggle cue, the action cue, and the combination of the two cues occurred at the same time (the lifetime of the verb). For the action and the wiggle alone, we also found no reliable differences. Our results thus support the suggested Visual Account such that both the action and the wiggle function as purely visual cues which increased the saliency of the target object.

The overall accuracies for adults were, however, surprisingly low (68.9%) although case-marking on the second noun-phrase disambiguated the sentences. From our first experiment and from previous research (Kamide, Scheepers, & Altmann, 2003; Matzke et al., 2002) we know that adults can use case-marking for thematic role assignment. Since case-marking on the first noun-phrase was ambiguous, adults might have initially expected an SVO sentence because SVO is overall preferred over OVS (Hemforth & Konieczny, 2013). Independent of the visual cue, we expected adults' accuracies to be high because of the case-marked NP2. It is possible that the design of the fillers elicited lower accuracies as expected. Bear in mind that our fillers contained many SVO sentences (48) which might have led people to believe that they are encountering an SVO sentence in addition to their strong SVO word order bias. This would, however, suggest that adults ignored case-marking on the second noun-phrase which seems unlikely on the basis of existing findings on case-marking.

The low accuracies may have, however, resulted from the frequent language-scene mismatches in the fillers. The fillers depicted actions which mismatched the spoken utterances (e.g., for the verb 'bake' the scene depicted a sawing action). Additionally, a character which was depicted in the scene but not part of the spoken utterance wiggled or a character other than the agent of the sentence wiggled. These mismatches may have had the effect that participants interpreted case-marking on the second noun-phrase as a mismatch, resulting in incorrect thematic role assignment and responses to comprehension questions.

What is interesting, is that the mismatching wiggles in the fillers did not influence target anticipation or accuracies in a negative way: Target preference was not lower in the wiggle (vs. action) condition (Figure 9.2), response accuracies were similarly low in both conditions (9.4). Unlike mismatching actions which can safely be identified as a verb-action mismatch, the mismatching wiggles may not be safely identified as a mismatch because a) the wiggle was not mediated by the verb but occurred during the verb and b) the wiggle did not exclusively relate to the agent because characters other than the agent wiggled in the fillers. The fact that mismatching wiggles in the filler items did not hinder processing suggests that participants identified the wiggling character in the experimental items as the agent, perhaps as a result of the ambiguity in the sentence.

Although the mismatches in the fillers might have influenced our results, case-marking on the second noun-phrase in our study seems to support disambiguation processes less than the action plus wiggle cue. Knoeferle et al. (2007) reported ERPs during comprehension and observed similar effects (even though temporally different) of visual and linguistic cues on sentence disambiguation. Although the measure they used differs from eye-tracking, it seems that the results of Experiment 3 do not corroborate the idea that visual and linguistic cues influence sentence disambiguation in a similar way (case-marking on the second noun-phrase vs. action and wiggle cue; Figure 9.4).

10.3.2 *Children*

Children also looked at the agent (vs. patient) more when visual cues were present (vs. absent). We observed reliable effects of action, wiggle, and an interaction of action and wiggle during the verb and adverb region. The results corroborate the rapid visual context effects reported by Zhang and Knoeferle (2012) and Münster (2016) who found more looks to the target character when actions were depicted in comparison to when they were not depicted. Unlike Zhang and Knoeferle (2012) and Münster (2016), looks to the agent (vs. patient) were not delayed in Experiment 3. Descriptively, we did, however, observe a similar pattern: For adults, the preference to look at the agent (vs. patient) was higher during the verb than the adverb region; Children's preference of agent inspections was higher during the adverb than the verb region. One of the reasons why attentional responses were immediate in our study might be explained by the sudden onset of the visual cues. By contrast, Zhang and Knoeferle (2012) and Münster (2016) presented the depicted actions throughout the sentence, resulting in delayed effects of depicted actions in children versus adults.

The effects of the visual cues were all reliable during the verb and adverb region. However, during the NP2 region, we observed a marginal effect of action and an interaction of action and wiggle. These effects were unexpected such that we expected participants to inspect the agent upon its mention, regardless of condition. Perhaps the effect of action persisted and the effect of wiggle disappeared because the action disappeared from the scene whereas the agent (the wiggling character) did not, facilitating looks to where the action had been (with the agent). Existing findings by Richardson and Spivey (2000) on adult language processing suggest that participants fixate empty locations in which an event was presented previously when they respond to questions about the event (see Richardson & Spivey, 2004 for a discussion). Crucially, such a behaviour was also observed when the locations move around, even in young children (Richardson & Kirkham, 2004). In that sense the location of the depicted action is spatially indexed in the brain and fixated as a result of the memory activating the depicted action when the character (that performed the action) was mentioned at the end of the sentence.

We did not replicate reliable effects of depicted actions on accuracies, as suggested by Zhang and Knoeferle (2012) and Münster (2016). Unlike in the adult data of Experiment 3 and the child data of Experiment 1 but similar to the findings by Münster (2016), we observed an unpredicted significant effect of voice. Children's responses were more often correct for active in comparison to passive voice comprehension questions. It is possible that children simply used the relationship between the visual cue and the agent since responses to active voice comprehension questions were the agent. Children may have used this direct relationship between the visual cue and the comprehension question. Another possibility is that the main effect of voice resulted from the lack of visual support during the questions. Recap that in Experiment 1 children did not struggle with passive voice comprehension questions (unlike in Münster, 2016) which was possibly influenced by the fact that the actions (although ambiguous) were depicted during the comprehension question. The main effect of voice in Experiment 3 supports the idea that children need visual support for correct responses to the more difficult passive voice comprehension questions. The visual cues were, however, not present during the comprehension questions (similarly to Münster, 2016), resulting in lower response accuracies.

Following from the difference in correct responses between active and passive voice questions, we examined the subsets of active and passive voice data separately. The results of the active data revealed a marginal effect of wiggle such that responses were more often correct when the wiggle was present (vs. absent). Perhaps children used the more salient cue (the wiggle) to keep the target character (the agent) in working memory, which then facilitated access in response to comprehension questions. The eye-gaze data, however, did not corroborate the idea that the wiggle, in particular, boosted attentional responses. In the adverb region, children looked significantly more at the agent (vs. patient) when one action (vs. one wiggle) was depicted. More research is needed to investigate this interpretation further.

A wiggling target character could, however, also hinder thematic role assignment. From an early age, children acquire verb-argument structure and associated abstract knowldge (e.g., Bencini & Valian, 2008; Messenger, Branigan, McLean, & Sorace, 2012; Peter, Chang, Pine, Blything, & Rowland, 2015. Therefore, children may know that a transitive verb requires two arguments (e.g., agent and patient). The depicted action on the one hand is compatible with the argument structure of the verb because it represents a two-argument event (e.g., the bear$_{\text{PATIENT}}$ pushes the bull$_{\text{AGENT}}$). The wiggle, on the other hand, could be incompatible with the argument structure of the verb because it may introduce a one-argument event (e.g., the bull wiggles). Wiggle effects on question-accuracies, however, suggest that participants integrated the wiggle into the argument structure of the verb by at least the end of the sentence. Perhaps the fact that the wiggle occurred throughout the lifetime of the verb facilitated a wiggle-argument structure integration.

It remains a puzzle, however, why depicted actions did not boost children's responses to comprehension questions as it was the case in Zhang and Knoeferle (2012) and Münster (2016). One of the reasons why we failed to observe beneficial effects of depicted actions on accuracies might be the temporal limitations of the cue depiction. Another possible explanation is that children simply cannot use depicted actions in locally structurally ambiguous sentences (previous findings relied on unambiguous sentences). Ambiguous OVS sentences are also very rare in child directed speech (under 1%; Dittmar et al., 2008a) and are thus likely difficult-to-process for five-year-old children. Children in general have a strong SVO word order bias (Dittmar et al., 2008a). Understanding 'who does what to whom' in such difficult-to-process sentences possibly requires all of the child's attention: When encountering locally structurally ambiguous sentences, children likely need to restructure from their initial SVO interpretation which in turn requires all of their attention.

However, if children cannot use case-marking for correct thematic role assignment, as suggested by Dittmar et al. (2008a) and the results of Experiment 1, we should have observed effects of depicted actions, regardless of whether sentences are unambiguous (clearly case-marked) or ambiguous (ambiguously case-marked) since children need to restructure from their initial SVO interpretation, regardless. Arguably, it is more likely that the short presentation of the depicted action and not the ambiguity of the sentences caused the difference between our and previous findings.

Analyses on the subset of passive voice comprehension questions revealed no reliable effects of action or wiggle. Descriptively, however, children's accuracies were higher when the action was depicted (vs. not depicted). The depicted action is mediated by the verb whereas the wiggle directly cues the agent. Perhaps children used the depicted action slightly more when the question required the answer to be the patient: The wiggle cues the agent and not the patient and the depicted action included the actor and the acted upon. Passive voice questions in relation to thematic roles might also be more difficult for children because they not only have to understand 'who does what to whom' but also that the question does not require the agent. Similarly, Münster (2016) observed an effect of depicted action during passive voice comprehension questions. She argued that questions in passive voice enhanced the difficulty of the task, resulting in an influence of depicted actions on thematic role assignment. Even though the accuracy results of Experiment 3 are not reliable, they corroborate the idea that when comprehension questions are difficult (in addition to difficult-to-process, non-canonical OVS sentences) depicted actions facilitate language comprehension.

Another argument found in the literature is that children cannot resolve syntactic ambiguities because they prefer early-arriving cues over late-arriving cues (Choi & Trueswell, 2010). The authors argued that children's cognitive control abilities are not yet fully developed. An early-arriving cue could be case-marking but since children at the age of five seem to struggle with case-marking in non-supportive contexts (Dittmar et al., 2008a, results Experiment 1), they are likely to use word order (SVO) as an early-arriving cue. Gertner et al. (2006) and Dittmar et al. (2008b) suggested that children use word order as a cue for thematic role assignment from a very early age. However, our results do not fully corroborate the idea that early-arriving cues are preferred over late-arriving cues. If word order (for children SVO) in our study was the early-arriving cue, children would have used word order only and ignored the late-arriving visual cues. We did, however, find some influences of depicted actions (in passive voice comprehension questions) and even more of the wiggle (in active voice comprehension questions) such that the late-arriving visual cues facilitated language comprehension to an extent.

Prior research suggests that one of the main differences between child and adult language processing results from children's inability to compute pragmatic inferences online (Huang & Snedeker, 2009b; Trueswell et al., 1999). Our results do not necessarily corroborate these findings. The wiggle in our study was not mediated by the verb (it only occurred during the verb region) and can function as a pragmatic/focusing cue that highlights the target role filler. Children then have to infer that the the wiggling target character was intended to signal a different sentence structure (OVS instead of the preferred SVO). If that was the case, we present evidence against the suggestion that one of the main differences between child and adult language processing is children's inability to make pragmatic inferences (Huang & Snedeker, 2009b; Trueswell et al., 1999) since we observed marginal effects of the wiggle on language comprehension. In Trueswell et al. (1999) children failed to use the two-referent context in which two frogs are potential referents: One frog on a napkin and one frog on the table. Similarly, our scenes depicted two possible agents (agent character and ambiguous character). The wiggle then highlights the agent like the napkin highlights the frog. However, children's thematic role assignment was only influenced by the wiggle when post-sentence comprehension questions were in active voice: When there was a direct relation between the wiggling character and the correct answer (the agent). Thus, the wiggle may have directly cued the target character without any further computations.

Weighall and Altmann (2011) suggested that only higher span children can use action contexts for different sentence structures (center-embedded and right-branching sentences) whereas lower span children can only use action contexts when the comprehension question is directly related to the action. Although our questions did not directly relate to the action, the active questions did directly relate to the role filler performing an action or wiggling slightly. In Experiment 3, analyses including cognitive abilities (K-ABC high vs. low) revealed a marginal interaction of action, wiggle, and K-ABC such that wiggle presence influenced looks to the target character in children with low in comparison to children with high K-ABC results. We observed that children with high K-ABC scores looked more at the agent when the action was depicted but the wiggle was not present (condition c) one action no wiggle) in comparison to the other conditions. However, the preference to look at the agent was highest when both cues were combined (condition d) one action one wiggle) for children with low K-ABC scores.

Cognitive abilities did, however, not only influence eye-movements. Although we found no reliable effects of K-ABC on accuracies, descriptively, we found some differences. Children with high K-ABC scores had higher response accuracies when only one visual cue was depicted (condition b) no action one wiggle and c) one action no wiggle). For children with low K-ABC scores, accuracies were highest in condition d) one action one wiggle and in condition a) no action no wiggle. It is possible that these differences between high and low scoring K-ABC children resulted in no clear effects of the visual cues on accuracies. Again, this leads us to the conclusion that the temporal limitations of the visual cue depictions influenced our results such that children with high K-ABC scores were less influenced by the limitations of the cue whereas response accuracies were higher when only one cue was presented in children with low K-ABC scores (Figure 9.14).

In summary, the visual cues immediately boosted attentional responses in both children and adults. However, for adults the temporally limited action plus wiggle facilitated comprehension. Perhaps the temporal limitations resulted in less in-depth processing of the single cues and only the combination of verb-mediated and focus-mediated attention could facilitate comprehension (marginal interaction of action and wiggle on response-accuracies; Table 9.5, Figure 9.4). For children, on the other hand, the wiggle facilitated comprehension but only for active voice comprehension questions. Further research is needed to disentangle whether the wiggle facilitated comprehension because it directly cued or because it focused the target. If the latter was true, children computed pragmatic inferences online. But regardless of wiggle function, we observed differences between child and adult language comprehension. What we do not yet know is the exact reason for why children did not use the depicted action for correct thematic role assignment. Further research is needed to disentangle whether the ambiguity in sentence structure or the shortness of depiction of the visual cue influenced the results.

10.4 AGE DIFFERENCES

One of the main questions raised at the beginning of this thesis was whether online language comprehension is similar or different in children compared to adults. In our Experiments 1 and 3, we observed that children process language similarly to adults such that they anticipated the target character before it is mentioned (during the verb region). Thus, our results corroborate previous findings on child language processing. Nation et al. (2003) and Mani and Huettig (2012) suggested that ten-year-old children rapidly anticipate an upcoming target but at the age of two this was only the case for skilled (vs. unskilled) language producers. We did, however, observe slight differences between children and adults' preference to look at the agent (vs. patient) in Experiment 3. For children, the mean log-ratios were higher in the adverb than the verb region whereas for adults they were higher in the verb than the adverb region. In Zhang and Knoeferle (2012) and Münster (2016), effects of depicted actions were delayed by word region for children compared to adults. Although our results showed a similar pattern, the sudden onset of the visual cues likely influenced our results. Nevertheless, we did observe a difference in the preferences to look at the agent. It seems that for adults the visual cues immediately had the strongest influence on visual attention during the verb region and for children slightly later during the adverb region.

Although the processing mechanisms seem similar, adults and children rely on different cues during processing. In Experiment 1, adults rapidly exploited case-marking for correct thematic role assignment. They anticipated the patient in SVO sentences and the agent in OVS sentences. Interestingly, target anticipation started from the beginning of the verb region onwards which suggests an additional influence of thematic role knowledge. Children, on the other hand, relied on word order (SVO). Although their eye-movements were anticipatory, they inspected the patient (vs. agent) more in both, SVO and OVS, sentences. We have argued that children may need additional visual information to be able to rapidly exploit case-marking for correct thematic role assignment, as it was suggested in previous studies (Münster, 2016; Özge et al., 2016; Zhang & Knoeferle, 2012). Since our study depicted actions ambiguously, no additional visual information was provided which is likely why we failed to observe an influence of case-marking on children's incremental thematic role assignment. We further argued that a visual bias may have influenced children's eye movements towards the patient. Perhaps the posture of the NP1 role filler (which children looked at upon its mention) directed attention towards the patient (vs. agent) and thus effectively garden-pathed children towards an SVO interpretation of OVS sentences.

Children and adults' visual attention was immediately (at the verb) guided by the visual cues in Experiment 3 but the cues facilitated language comprehension differently in the two age groups. Adults' accuracies were highest in the action plus wiggle condition, indicating beneficial effects of a combined action and wiggle when the cues are short-lived. Children's responses were more often correct in active versus passive voice comprehension questions and more often when the wiggle was present (vs. absent). We argued that one of the main reasons for this difference might have resulted from the temporal limitations of the cue depiction.

For both age groups, we did not find reliable effects of prosody on correct thematic role assignment. We did, however, observe a similar increase of target inspections (although patient inspections for OVS sentences in children) in children and adults, indicating that both age groups are somewhat sensitive to prosody.

In summary, our studies provide new evidence for online language comprehension differences between children and adults. The overall mechanisms of visual attention seem similar in children and adults but they rely on different cues and thus language comprehension is facilitated in different ways. These findings are especially interesting for accounts of visually situated language comprehension because age seems to be a factor which influences the way different linguistic and visual cues are processed. In the following section, we discuss how our results can enrich an account such as the CIA.

10.5 implications for the cia/scia

The Coordinated Interplay Account highlights the tight temporal coordination of sentence interpretation and perception of the visual context during real-time language comprehension (Knoeferle & Crocker, 2007; Knoeferle et al., 2014). Recently, Münster and Knoeferle (2018) suggested the importance of properties of the comprehender (ProCom) and the speaker during language comprehension. The authors argued that properties of the comprehender can either be biological (e.g., age) or experiential (e.g., literacy). Our results can potentially enrich the CIA/sCIA in terms of age. Case-marking as a cue for thematic role assignment was rapidly integrated by adults but not by children. We thus provide further evidence for the age factor to be included in the properties of the comprehender (ProCom) in the CIA, as suggested by Münster and Knoeferle (2018).

Regardless of case-marking, children's expectations about the target role filler were formed on the basis of their strong SVO word order bias when the context was not supportive. Currently the age of the comprehender influences the interpretation (Interpretation of word$_i$ based on ProCom) and the expectations (Expectations based on ProCom). However, the interpretation is additionally based on the linguistic constraints and the expectations are based on linguistic/long-term knowledge. But, the linguistic constraints and the linguistic/long-term knowledge are also influenced by age or rather age influences what the linguistic constraints, linguistic long-term knowledge are. For five-year-old children, for example, case-marking is not yet a linguistic constraint that changes their interpretation. Word order (SVO), however, is a linguistic constraint which influences their sentence interpretation. Thus, the age of the comprehender influences many aspects of *Utterance-mediated attention*.

In both Experiment 1 and 3, we found a difference in visual attention between children and adults (e.g., adults' preference to look at the target role filler during the first noun-phrase was higher than children's; adults preference to look at the agent (vs. patient) was higher during the verb region of the sentence whereas children's preference was higher during the adverb region), such that each aspect within *Utterance-mediated attention* is influenced by the age of the comprehender. The last stage *Integration* is then also influenced by age. We have seen different results of the visual cues on children and adults' accuracies such that children do not only reconcile their interpretation and expectations with the scene different from adults but also they revise their interpretation based on the visual cues different from adults. The wiggle, for example, influenced children's interpretation of the sentences whereas the combined action and wiggle elicited comprehension about 'who does what to whom' in adults.

Experiences of the comprehender are also included in ProCom (Münster & Knoeferle, 2018). However, age (as a biological characteristic) may sometimes compete with experience. It has been repeatedly suggested in previous research that statistical learning plays a role in incremental language processing. Borovsky et al. (2012), for example, suggested that vocabulary knowledge stands in relation to incremental language processing such that higher vocabulary knowledge resulted in more and earlier looks to the target. Trueswell et al. (1999) argued that children did not use the referential context because the verb put occurs more often in a destination context than in a location context.

The idea that children rely on lexical bias during language processing was further supported by Snedeker and Trueswell (2004) who found that children rely on verb bias rather than referential context. Wells et al. (2009) showed improved reading speed for difficult-to-process object relative clauses after reading experience. Dittmar et al. (2008a) and we argued that children at the age of five rather use word order (SVO) than case-marking because OVS sentence structure rarely occurs in child-directed speech. Overall, these results all point towards a role of statistical knowledge or learning during language processing. Therefore, *Experience* should be weight against *age* within ProCom. Depending on the context, one of the two (age or experience) may weigh more than the other: In Borovsky et al. (2012), for example, older children and children with higher vocabulary knowledge anticipated the target object earlier than younger children or children with low vocabulary knowledge. Probabilities (0,1) of age and experience within ProCom could signal which aspect weights more.

Another aspect which influences language comprehension is prosody. Münster and Knoeferle (2018) suggested to include acoustic properties in the *Speaker Characteristics* as a factor that influences language comprehension in real-time. The way in which language is uttered can influence a comprehenders expectations. Snedeker and Trueswell (2003), for example, found that speakers use prosody when it necessary and listeners exploit prosody when it is available. Prosody is another factor which should be included in the *Speaker Characteristics*. Existing findings by Grice et al. (2017) suggest that different speakers realise, for example, contrastive stress differently. However, not only individual differences but also regional differences play a role in the use and of prosody. Although we did not find reliable effects of prosody for thematic role assignment, our results indicate that children and adults are somewhat sensitive to prosody (more looks to the agent/patient in the biasing prosody conditions compared to the neutral prosody conditions). Evidence from previous research suggests an influence of prosody on ambiguity and reference resolution (Dahan et al., 2002; Weber et al., 2006). In the CIA prosody could potentially influence the comprehenders' expectations. More research is needed to be able to precisely include the role of prosody on incremental language processing.

In Experiment 3, we observed that the wiggle influenced participants' visual attention (children and adults looked at the wiggling target character) and also their interpretation. We argue that the wiggle is not utterance-mediated but purely scene-mediated. Perhaps even the depicted action in this specific design is scene-mediated because of the temporal limitations of action depiction. If the wiggle was utterance-mediated, effects on eye-movements (looks to the agent) would have occurred later, not immediately at verb onset Therefore, target anticipation can be attributed to the visual manipulation alone (not language). However, future research is needed to elaborate on this (e.g., leaving the verb out of the sentence could show whether target anticipation is similar to when the sentence has a verb and thus possibly support the idea of scene-mediated attention). Similarly, existing findings of an arrow as a cue have shown that participants looked towards a target referent solely on the basis of the arrow pointing towards the referent (Staudte et al., 2014). The current version of the CIA does not include a component of scene-mediated attention.

We suggest that it may be helpful to relabel the second stage from *Utterance-mediated attention* to *Mediated attention* because visual attention can not only be utterance- but also scene-mediated. We further suggest to add a *saliency, visually motivated search* aspect within the *Mediated-attention* stage. The existing mechanisms (Referential search, Anticipatory search, Merger, and Decay) would then be indexed by *Um* (Utterance-mediated) or *Sm* (Scene-mediated). The *Referential search* and the *Anticipatory search* would be indexed for *Um* whereas the *Merger*, *Decay* would be indexed for both *Um* and *Sm*, and the *saliency, visually motivated search* would be indexed by *Sm*.

10.5.1 *Example: Scene mediated attention*

We have argued earlier that mediated attention can either be utterance-mediated or scene-mediated. The results of Experiment 3 support the idea that a visually mediated cue (the wiggle) influences visual attention (more looks to the agent than the patient) but elicits different results on accuracies in children and adults. In this section, we provide an example of how scene-mediated attention can be included in the CIA, using the ambiguously case-marked sentence *Das Käferchen schubst gerade der Stier* ('The (neuter) bug$_{\text{patient}}$ pushes immediately the bull$_{\text{agent}}$') from Experiment 3. The visual scene contains three animal characters. The target character (the bull - the agent) wiggles slightly up and down during the verb region of the sentence (see example image Experiment 3: Figure 9.1).

Upon hearing the verb *schubst* ('pushes') an interpretation of the word is formed based on what has been mentioned before (int$_{i''}$), the linguistic constraints, and the age which yields int$_{i''+1}$. Expectations are based on previous expectations (ants$_{i''}$), linguistic/long-term knowledge, and age. In this example both, children and adults, would expect an upcoming patient (SVO word order preference). The second stage, newly labelled *Mediated attention* then depends on the type of visual information (e.g., depicted actions are verb-mediated and thus utterance-mediated whereas a wiggling target character is scene-mediated). The wiggling target character leaves the referential and anticipatory search out (both indexed for *Um*) but elicits a saliency/visually motivated search (indexed for *Sm*) which again is influenced by age. Our results have shown that children's preference to look at the target character is lower than adults' during the verb region. Newly attended scene information (the wiggle) is then merged with the previously seen scene$_{i''}$ to scene$_{i''+1}$. In the next stage, the interpretation of the verb *schubst* ('pushes') and the expectations are reconciled with the scene$_{i''+1}$.

Then the next word, the adverb *gerade* ('immediately') follows and initiates a new processing cycle. Again, the interpretation of the adverb and the expectations are formed on the basis of the previous processing cycle (int$_{i''+1}$; ant$_{i''+1}$) which yield int$_{i''+2}$ and ant$_{i''+2}$. In the second stage, *Mediated attention*, the newly attended scene is merged with the previously seen scene$_{i''+1}$. The

$scene_{i''+1}$ is meanwhile held in working memory and thus still influences looks to the target character. Decay of objects, events, or saliency which are no longer present is indexed for *Sm* because the visual cues no longer mediate visual attention (the information held in Working Memory does). Again interpretation and expectations are reconciled with the $scene_{i''+2}$.

Upon hearing the second noun-phrase *der Stier* ('the $bull_{ag}$), an interpretation of the word is formed based on what has been previously heard, $int_{i''+2}$, linguistic constraints, and age which yields $int_{i''+3}$. Linguistic constraints influence children and adults' interpretation in different ways: Adults use case-marking whereas children use word order for incremental thematic role assignment. In the second stage, *Mediated attention*, adults and children perform a referential search based on new referring expressions (they hear the word *bull* and look at the *bull*). The referential search is indexed *Um* because it is utterance mediated. The working memory component still carries $scene_{i''+1}$. In the last step *Integration* the interpretation and expectations are reconciled with the scene and the scene in working memory. Finally, the processing cycles end and the ambiguous OVS sentence is interpreted.

Taken together, we have pointed out how the current version of the CIA can be enriched by modulating ProCom and adding scene-mediated attention to the processing cycle. The current version of the CIA does not yet include scene-mediated attention. Thus, adding this aspect to the CIA provides a more detailed description of the interplay between language comprehension and visual attention. Crucially, it helps us understand language processing in more detail.

11

CONCLUSION

In this thesis, we have shown that children at the age of five, unlike adults, are not yet able to rapidly exploit case-marking as a cue for incremental thematic role assignment when the visual scene does not support 'who does what to whom' via world knowledge or unambiguous depicted actions. This result has implications for language acquisition and language education. Five-year-olds have not yet fully acquired German case-marking (i.e., accusative case on the first noun-phrase indicates patient/agent ordering - object-verb-subject sentence structure) which might not, at last, be influenced by the rarity of this structure in child-directed speech. In an educational and clinical settings, our findings could help to improve children's understanding of case-marking by a) tutors using OVS sentences more often (frequency) and b) visual support.

However, the time course and amount/type of visual support (visual cues) play a role in language comprehension. Depicted actions, for example, can facilitate language comprehension in children when the depiction is not temporally limited (Münster, 2016; Zhang & Knoeferle, 2012). Temporal limitations in the presentation (e.g., during the verb of the sentence) elicit less in-depth processing of the action, resulting in no benefit on comprehension of 'who does what to whom'. A wiggling target character, which is not mediated by language, helps children but only when the comprehension question directly connects to the target. For adults, the combination of action and wiggle elicits more correct responses although adults should be able to use case-marking on the second noun-phrase for correct thematic role assignment.

In the discussion, we proposed that children likely need visual support but possibly without temporal limitations in order to correctly assign thematic roles and use case-marking. Future research could disentangle this issue by manipulating case-marking (ambiguous vs. unambiguous) and depicted actions (present vs. absent) or world knowledge (present vs. absent). The results could then shed light on whether a) children only struggle with case-marking when the visual scene is not supportive, b) sentence ambiguity influences the use of visual cues.

Another possibility to disentangle whether children need visual support to be able to use case-marking for thematic role assignment or whether they cannot use case-marking yet, is to prosodically mark (vs. not mark) the case-marker itself (instead of the first noun-phrase or the verb) and depict actions (vs. no actions) in unambiguous German SVO and OVS sentences. Prosodic marking would include a pitch accent on and a pause after the case-marker which would clearly highlight case-marking. Depicting actions vs. no actions, could then tease apart whether it is case-marking or the visual support or a combination of both which facilitates comprehension.

Furthermore, our results indicate that there is no one-to-one mapping between prosody (the prosodic structures used in our experiments) and thematic role assignment. Individual and regional speaker and comprehender differences likely influence the use of prosody. Prosody, unlike case-marking, is not as static and thus likely to influence each person in a different way, especially if pitch accents are used in isolation (which are in themselves produced differently by different speakers; Grice et al., 2017) and not in combination with other prosodic means (e.g., pauses, lengthening, etc.).

Looking back at the big question raised at the beginning of whether children and adults rely on the same processing mechanisms during language comprehension: The results suggest that children's and adults' eye-movements are similarly anticipatory but different cues (case-marking, visual cues) influence language comprehension in different ways. One of the main differences between child and adult online language comprehension was suggested to be children's inability to make pragmatic inferences. Our results do not necessarily corroborate this idea: Although children used the wiggle (pragmatic/focusing cue) to facilitate language comprehension, it may have functioned as a direct cue to the target character, including no further computations (i.e., only when the comprehension questions directly related to the wiggling character - the agent).

Overall, the findings of this thesis have important implications for accounts of situated language processing such as the Coordinated Interplay Account (Knoeferle & Crocker, 2006, 2007; Knoeferle et al., 2014) which highlights the interplay between language comprehension and the visual world. We have shown and discussed how the age of the comprehender influences which linguistic constraints/linguistic long-term knowledge come into play when forming an interpretation and developing expectations of the spoken utterance. Children at the age of five are rather influenced by word order whereas adults can use case-marking to assign thematic roles. Further research is needed to explore, in more detail, which linguistic constraints influence child language processing and what exactly their linguistic/long-term knowledge entails. Crucially, scene-mediated attention can also influence language comprehension (Exp. 3).

Language in itself is a complex construct. Although children acquire language rapidly throughout their first years of life, a full command of their native language takes time to develop. Thus, understanding 'who does what to whom' can be challenging at the beginning. Since children learn a lot from their environment (e.g., when they play), it is not surprising that visual rather than morpho-syntactic information can be helpful to correctly interpret spoken utterances.

12

GERMAN SUMMARY

Diese Doktorarbeit befasst sich mit der Sprachverarbeitung von Kindern im direkten Vergleich zu Erwachsenen. Wir haben mit Hilfe von Messungen der Blickbewegungen genauer untersucht, ob fünfjährige Kinder und Erwachsene Kasusmarkierungen, Prosodie oder visuelle Informationen schnell nutzen, um thematische Rollen inkrementell zuzuweisen.

Bestehende experimentelle Studien haben gezeigt, dass Erwachsene gesprochene Sprache inkrementell verarbeiten und währenddessen eine Reihe verschiedener sprachlicher oder visueller Informationen einbeziehen können. Diese sprachlichen oder visuellen Informationen tragen dann dazu bei, wie wir die gesprochene Sprache interpretieren. So kann zum Beispiel die Kasusmarkierung im Deutschen unmittelbar unsere thematische Rollenzuweisung (wer macht was mit wem) beinflussen (Kamide, Scheepers, & Altmann, 2003; Matzke et al., 2002). Aber nicht nur morphosyntaktische Informationen spielen eine Rolle, wenn wir uns untereinander verständigen möchten, auch das Gesprochene kann zweideutig sein. Dann kann die Art und Weise, wie wir bestimmte Teile eines Satzes betonen, dazu beitragen, eine solche Zweideutigkeit aufzulösen. Weber et al. (2006) haben zum Beispiel gezeigt, dass junge Erwachsene Prosodie benutzen, um vorzeitig eine strukturelle Zweideutigkeit, die durch eine feminine Kasusmarkierung hervorgerufen wurde, aufzulösen.

Wie bereits erwähnt, können Erwachsene auch nichtsprachliche (visuelle) Informationen nutzen, um frühzeitig eine Zweideutigkeit aufzulösen, thematische Rollen zuzuordnen oder einen Bezug herzustellen. Nichtsprachliche Informationen, die unmittelbar unser Sprachverständnis beinflussen können, sind unter anderem a) ein visueller Kontrast zwischen Gegenständen, b) ein visueller Verweisungszusammenhang (referentieller Kontext) oder c) dargestellte/abgebildete Handlungen (Knoeferle et al., 2005; Spivey-Knowlton & Sedivy, 1995; Tanenhaus et al., 1995). Allerdings scheint es eine zeitliche Verzögerung zwischen semantischer Interpretation und pragmatischen Rückschlüssen zu geben (Huang & Snedeker, 2009a) und auch eine Diskrepanz zwischen verschiedenen visuellen Informationen, das heißt, dass unterschiedliche visuelle Informationen verschieden oder auch ähnlich auf das Sprachverständnis einwirken (z.B., Kreysa et al., 2014; Münster, 2016).

Obwohl der Erstspracherwerb bei Kindern sehr schnell verläuft, sind trotzdem nicht direkt alle Sprachaspekte verfügbar. Aus diesem Grund haben Sprachforscher in den letzten Jahren untersucht, ob Kinder Sprache genau so verarbeiten und verstehen wie Erwachsene. Obwohl Kinder genauso wie Erwachsene unmittelbar ihre Aufmerksamkeit auf dargestellte Objekte richten, über die gesprochenen worden ist, gibt es trotzdem Unterschiede zwischen Kindern und Erwachsenen in der Sprachverarbeitung. Wir haben allerdings festgestellt, dass viele wissenschaftliche Studien Messmethoden benutzt haben, die nicht den zeitlichen Ablauf der Sprachverarbeitung beschreiben. Außerdem wurde häufig kein direkter Vergleich zwischen Kindern und Erwachsenen gezogen. Wobei genau dieser notwendig ist, um genaue Aussagen über Unterschiede oder Gemeinsamkeiten im zeitlichen Ablauf der Sprachverarbeitung zu untersuchen. Das heißt, es ist weiterhin unklar, wie genau sich die Sprachverarbeitung von Kindern und Erwachsenen unterscheidet oder eben nicht.

Eine der ersten Studien, die Kinder im direkten Vergleich zu Erwachsenen untersucht hat, wurde im Jahr 1999 von Truesell und seinen Kollegen durchgeführt. In dieser Studie wurde untersucht, ob Kinder auch unmittelbar visuelle Verweiszusammenhänge in ihr Sprachverständnis integrieren, um zweideutige syntaktische Strukturen zu vereindeutigen. Die Ergebnisse dieser Studie deuten darauf hin, dass Kinder, nicht so wie Erwachsene, visuelle Verweiszusammenhänge nutzen, sondern unabhängig vom visuellen Kontext die bevorzugte Struktur zur Interpretation nutzen. Verschiedene Nachfolgeuntersuchungen haben versucht zu klären, woher dieser Unterschied zwischen Kindern und Erwachsenen rührt. Eine durchaus plausible Erklärung ist, dass Kinder im Alter von fünf Jahren noch keine pragmatischen Rückschlüsse zwischen der gesprochenen Sprache und den visuellen Verweiszusammenhängen ziehen können.

In Anbetracht der Tatsache, dass es eine Spannung zwischen semantischer Interpretation und pragmatischen Rückschlüssen im Sprachverständnis von Erwachsenen zu geben scheint und pragmatische Rückschlüsse generell problematisch für Kinder zu sein scheinen, war es das Ziel dieser Doktorarbeit, diese Aspekte mit Hilfe von Prosodie (die ebenfalls pragmatische Rückschlüsse erfordern kann), Kasusmarkierungen und visuellen Informationen zu untersuchen.

Kapitel 2 dieser Arbeit beschreibt die Entwicklung des Zusammenspiels zwischen Blickbewegungen und Sprachverarbeitung. Hierzu beschreiben wir zunächst die Methode an sich und betrachten wissenschaftliche Studien, die gezeigt haben, wie a) gesprochene Sprache unsere Blickbewegungen auf Objekte in einer visuellen Darstellung beeinflusst und b) wie dann diese betrachteten Objekte unsere Interpretation der gesprochenen Sprache beinflusst. Hierzu haben wir uns angeschaut, welche verschiedenen Informationsquellen eine Rolle während des Verstehens der Sprache spielen.

In Kapitel 3 beschreiben wir dann mehrere experimentelle Beweise, die gezeigt haben, dass Kasusmarkierungen, Prosodie und visuelle Informationen unsere Sprachverarbeitung unmittelbar beeinflussen, wobei hier ein besonderer Fokus auf den Einfluss der Informationsquellen auf die thematischen Rollenzuweisungen gelegt wird. Als nächstes argumentieren wir, dass viele der gesichteten Studien sich nur auf eine Informationsquelle bezogen haben und beschreiben zudem Studien, die mehr als eine untersuchten. Anhand dessen stellen wir heraus, dass es sowohl Unterschiede als auch Gemeinsamkeiten in der schnellen Nutzung verschiedener Informationsquellen gibt.

In Kapitel 4 befassen wir uns dann mit der Sprachverarbeitung und dem Sprachverständnis von Kindern. Zuerst schauen wir uns die Gemeinsamkeiten an. Hierbei betonen wir, dass Kinder auch so wie Erwachsene verschiedene Informationen schnell in das Sprachverständnis einbetten können (z.B. grammatische oder synatiktische Informationen). In dem darauffolgenden Abschnitt beschreiben wir dann die Unterschiede. Auch hier liegt ein spezieller Fokus auf dem Einfluss von Kasusmarkierungen, Prosodie und visuellen Infromationen. Des Weiteren stellen wir auch noch Entwicklungsaspekte heraus.

Kapitel 5 beschreibt dann den Coordinated Interplay Account, ein Modell der Sprachverarbeitung in visuellen Kontexten. Hier fassen wir die wesentlichen Schritte dieses Accounts zusammen, beschreiben eine Erweiterung des Accounts (dem social Coordinated Interplay Account) und legen dar, dass unsere Studien den CIA potentiell im Bezug auf das Alter einer Person oder auf pragmatische Aspekte der Sprachverarbeitung erweitern könnte.

In Kapitel 6 fassen wir dann noch einmal zusammen, was die Sichtung der Literatur für Fragen offen gelassen hat, um dann im Folgekapitel (Kapitel 7) unseren eigenen Beitrag zur Forschung darzustellen. Die Experimente sind in drei Blöcke aufgeteilt, die jeweils 2 eye-tracking Studien beinhalten. Im ersten Block haben wir untersucht, ob fünfjährige Kinder und Erwachsene Kasusmakierungen und/oder Prosodie benutzen, um thematische Rollen zuzuweisen. Während des Experiments betrachteten die Teilnehmer Szenen und hörten kurz darauf gleichzeitig einen Subjekt-Verb-Objekt (SVO) oder Objekt-Verb-Subjekt (OVS) Satz, der entweder, so wie bei der Studie von Weber et al. (2006), eine Satzstruktur unterstützende oder neutrale Prosodie zugewiesen bekam (SVO: L*+H auf der NP1 und H* auf dem Verb, OVS: L+H* auf der NP1). Die Szenen zeigten drei Tiercharaktere, von denen zwei die gleiche Aktion ausführten (z.B. für das Verb *filmen* hielten die beiden Tiere eine Kamera in der Hand). Das mittlere Tier wurde immer am Anfang des Satzes benannt und war somit rollenzweideutig, da es entweder die Rolle des Agens oder des Patiens füllen konnte (es führte eine Handlung aus und eine Handlung wurde mit ihm ausgeführt). Der andere Charakter, der ebenfalls die gleiche Handlung ausführte konnte nur die Rolle des Agens füllen und der, der keine Aktion ausführte konnte nur die Rolle des Patiens füllen. Während die Versuchsteilnehmer den Satz hörten und die Szene betrachteten, haben wir ihre Augenbewegungen aufgenommen, bei denen wir uns die Blicke zum Agens und Patiens angeschaut haben, nicht aber zu dem mittleren Tier, welches beide Rollen füllen konnte. Kurz nach Ende des Satzes haben die Teilnehmer dann noch eine Frage im Aktiv oder Passiv zu den thematischen Rollen beantwortet (z.B. *Wer filmt hier?* oder *Wer wird hier gefilmt?*).

Die Ergebnisse haben gezeigt, dass Erwachsene, aber nicht Kinder, die Kasusmarkierungen genutzt haben, um unmittelbar nach der NP1 die thematischen Rollen richtig zuzuweisen. Kinder hingegen waren durch ihre bevorzugte Satzstruktur (SVO) mehr beeinflusst als durch die Kasusmarkierungen. Die Prosodie hat keinen eindeutigen Einfluss auf die unmittelbare Zuweisung der thematischen Rollen gezeigt. Wir haben zwar beobachtet, dass sowohl Kinder und Erwachsene leicht mehr Blicke auf den Agens oder Patiens gerichtet haben, wenn eine Satzstruktur unterstützende Prosodie vorlag, aber die Datenanalyse ergab keinen eindeutigen Einfluss der Prosodie. Jetzt stellte sich die Frage, warum Prosodie keinen Einfluss auf die thematische Rollenzuweisung hatte. Es könnte sein, dass die Kasusmarkierung generell einen stärkeren Einfluss auf die Rollenzuweisung hat als Prosodie. Allerdings haben die Kinder in unserer Studie gar keine Kasusmarkierungen genutzt, was eventuell auf der Mehrdeutigkeit des Bildes basiert.

Folglich haben wir untersucht, ob Kasusmarkierungen eine stärkere Informationsquelle liefern als Prosodie und was passiert, wenn die Kasusmarkierungen eine Mehrdeutigkeit hervorrufen und somit erst mal nur die Prosodie zur thematischen Rollenzuweisung genutzt werden kann. Da Kinder keine Kasusmarkierungen genutzt haben, waren die Versuchspersonen in diesem

Block alles junge Erwachsene. Um dieses Fragestellungen zu untersuchen, haben wir den Kontrast zwischen den beiden prosodischen Bedingungen erhöht, in dem wir gegensätzliche Satzstruktur unterstützende Prosodien verwendet haben. Das heißt alle SVO- und OVS-Sätze erhielten entweder eine SVO oder OVS unterstützende Prosodie. In einem zweiten Experiment haben wir dann ein- und mehrdeutige OVS-Sätze untersucht, die die gleiche Prosodie erhielten wie im ersten Experiment in diesem Block. Die visuellen Materialien waren denen aus dem ersten Experiment sehr ähnlich.

Ähnlich wie bei unserem ersten Experiment haben Erwachsene die Kasusmarkierungen, wenn sie eindeutig waren, genutzt, um die thematischen Rollen unmittelbar nach der NP1 zuzuordnen. Effekte von Prosodie auf die Rollenzuweisung blieben weitestgehend aus. Allerdings konnten wir beobachten, dass, wenn die Kasusmarkierungen zweideutig waren und die Sätze eine SVO unterstützende Prosodie trugen, die Teilnehmer etwas mehr den Blick auf den Zielcharakter (den Agens) richteten, was auf eine frühzeitige Agens-Rollenzuweisung hindeutet. Allerdings blieben auch diese Ergebnisse nur Tendenzen.

In einem letzten Block haben wir dann untersucht, ob Kinder und Erwachsene unterschiedliche visuelle Information nutzen können, um thematische Rollen in mehrdeutigen Sätzen schnell zuzuweisen (vor dem Satzende, die NP2 trug eine eindeutige Kasusmarkierung). Hierzu hörten die Teilnehmer zeitweise mehrdeutige OVS-Sätze und betrachteten visuelle Szenen. Die Szenen zeigten wieder drei Tiercharactere, aber diesmal unterschieden sich die Szenen in ihrer Darstellung. Während des Verbs passierte entweder gar nichts, der Agens führte eine Akion aus, wackelte hoch und runter oder beides.

Die Ergebnisse weisen darauf hin, dass die kurzzeitigen visuellen Informationen die Augenbewegungen direkt beeinflussen. Allerdings haben sie unterschiedliche Auswirkungen auf das Sprachverständnis. Bei Erwachsenen konnten wir beobachten, dass das Zusammenfügen zweier verschiedener visueller Informationen einen marginell positiven Einfluss auf die Beantwortung der Verständnisfragen hatte. Bei Kindern konnten wir nur einen marginell positiven Einfluss des wackelnden Agens feststellen, als wir uns die Aktivfragen gesondert von den Passivfragen angeschaut haben.

Alles in allem konnten wir durch unsere Untersuchungen herausstellen, dass es Unterschiede in der Sprachverarbeitung zwischen Kindern und Erwachsenen gibt. Während Erwachsene die Kasusmarkierungen nutzen, um ihre visuelle Aufmerksamkeit schon kurz nach der NP1 auf den richtigen Charakter, der die richtige Rolle füllt, zu richten, sind Kinder doch mehr beeinflusst durch ihre bevorzugte Satzstruktur (SVO). Auch wenn beide Altersgruppen weitestgehend unberührt von Prosodie geblieben sind, scheint doch eine Sensibiliät für die Betonung zu bestehen. Kurzzeitige visuelle Informationen haben die visuelle Aufmerksamkeit zum Zielcharakter angetrieben, allerdings mit unterschiedlichen Auswirkungen auf das Sprachverständnis.

In Kapitel 8 greifen wir die Ergebnisse unserer Studien noch einmal auf und diskutieren diese im Zusammenhang mit den aus den ersten Kapiteln hervorgehenden Beweisen. Hierbei legen wir einen speziellen Fokus auf die herausgestellten Forschungslücken und versuchen diese anhand unserer Forschung aufzuklären oder abzugrenzen. Zusammenfassend konzentrieren wir uns dann auf die genauen Unterschiede zwischen Kindern und Erwachsenen im situierten Sprachverstehen. In einem letzten Schritt versuchen wir dann unsere Erkenntnisse in die Darstellung von situiertem Sprachverstehen (den CIA) zu intergieren und schlagen Erweiterungen dieser Darstellung durch den Faktor Alter sowie visuell geleitete Aufmerksamkeit und Interpretationen vor.

In einem letzten Schritt fassen wir zusammen, was unsere Ergebnisse in Bezug auf inkrementelle thematische Rollenzuweisung gezeigt haben und implizieren die Wichtigkeit der Ergebnisse für den Spracherwerb und die Spracherziehung. Des Weiteren argumentieren wir, welchen Einfluss unsere Ergebnisse auf unser Verständnis des situierten Sprachverstehens haben und diskutieren weitere Möglichkeiten, die unsere Ergebnisse und Vorschläge weiter untersuchen können.

Part V

APPENDIX

A

WORD ON- AND OFFSETS

The following tables provide the on- and offsets for each word in the experimental sentences by condition and divided by Experiments.

A.1 EXPERIMENT 1

Table A.1: Word On- and Offsets Experiment 1

item	word order	prosody	NP1 on-set	NP1 off-set	verb on-set	verb off-set	adv. on-set	adv. off-set	NP2 on-set	NP2 off-set
1	SVO	biasing	0	778	1241	1850	2381	3062	3565	4293
1	OVS	neutral	0	981	1241	1938	2381	3127	3565	4324
1	SVO	neutral	0	742	1241	1941	2381	3083	3565	4207
1	OVS	biasing	0	981	1241	1938	2381	3127	3565	4136
2	SVO	biasing	0	933	1403	2350	2803	3571	4067	4884
2	OVS	neutral	0	955	1403	2363	2803	3609	4067	4841
2	SVO	neutral	0	947	1403	2306	2803	3595	4067	4881
2	OVS	biasing	0	993	1403	2298	2803	3525	4067	4860
3	SVO	biasing	0	770	1357	2337	2949	3654	4186	4908
3	OVS	neutral	0	993	1357	2472	2949	3701	4186	4812
3	SVO	neutral	0	879	1357	2459	2949	3747	4186	4903
3	OVS	biasing	0	872	1357	2364	2949	3691	4186	4762
4	SVO	biasing	0	1081	1463	2237	2808	3531	4039	4778
4	OVS	neutral	0	1003	1463	2228	2808	3506	4039	4835
4	SVO	neutral	0	971	1463	2253	2808	3552	4039	4810
4	OVS	biasing	0	978	1463	2275	2808	3538	4039	4872
5	SVO	biasing	0	916	1416	2179	2627	3209	3642	4485
5	OVS	neutral	0	965	1416	2087	2627	3209	3642	4304
5	SVO	neutral	0	886	1416	2100	2627	3107	3642	4466
5	OVS	biasing	0	893	1416	2092	2627	3100	3642	4314

Table A.1: EXP 1

item	word order	prosody	NP1 on-set	NP1 off-set	verb on-set	verb off-set	adv. on-set	adv. off-set	NP2 on-set	NP2 off-set
6	SVO	biasing	0	852	1373	1995	2538	3037	3618	4435
6	OVS	neutral	0	871	1373	1907	2538	3065	3618	4265
6	SVO	neutral	0	859	1373	1995	2538	3103	3618	4372
6	OVS	biasing	0	1026	1373	2046	2538	3157	3618	4416
7	SVO	biasing	0	1183	1664	2612	3119	3692	4134	4931
7	OVS	neutral	0	1273	1664	2685	3119	3659	4134	4920
7	SVO	neutral	0	1169	1664	2666	3119	3661	4134	4870
7	OVS	biasing	0	1194	1664	2710	3119	3619	4134	4940
8	SVO	biasing	0	709	1298	2013	2558	3109	3527	4166
8	OVS	neutral	0	961	1298	1972	2558	3057	3527	4236
8	SVO	neutral	0	823	1298	2075	2558	3082	3527	4138
8	OVS	biasing	0	1039	1298	2030	2558	3107	3527	4166
9	SVO	biasing	0	855	1420	2293	2758	3228	3772	4651
9	OVS	neutral	0	1068	1420	2288	2758	3307	3772	4643
9	SVO	neutral	0	955	1420	2288	2758	3288	3772	4682
9	OVS	biasing	0	969	1420	2257	2758	3330	3772	4665
10	SVO	biasing	0	739	1303	2330	2796	3346	3824	4623
10	OVS	neutral	0	845	1303	2353	2796	3299	3824	4589
10	SVO	neutral	0	806	1303	2285	2796	3338	3824	4711
10	OVS	biasing	0	763	1303	2240	2796	3388	3824	4594
11	SVO	biasing	0	930	1456	2245	2702	3219	3707	4449
11	OVS	neutral	0	1094	1456	2189	2702	3345	3707	4457
11	SVO	neutral	0	976	1456	2196	2702	3234	3707	4452
11	OVS	biasing	0	976	1456	2183	2702	3280	3707	4462
12	SVO	biasing	0	849	1394	2274	2663	3221	3701	4379
12	OVS	neutral	0	1033	1394	2183	2663	3231	3701	4424
12	SVO	neutral	0	917	1394	2166	2663	3213	3701	4387
12	OVS	biasing	0	917	1394	2090	2663	3171	3701	4455

Table A.1: EXP 1

item	word order	prosody	NP1 onset	NP1 offset	verb onset	verb offset	adv. onset	adv. offset	NP2 onset	NP2 offset
13	SVO	biasing	0	791	1352	2169	2696	3762	4248	5329
13	OVS	neutral	0	817	1352	2169	2696	3814	4248	4997
13	SVO	neutral	0	813	1352	2172	2696	3751	4248	5188
13	OVS	biasing	0	843	1352	2187	2696	3786	4248	4945
14	SVO	biasing	0	856	1348	2153	2738	3839	4294	4874
14	OVS	neutral	0	887	1348	2099	2738	3830	4294	5124
14	SVO	neutral	0	863	1348	2244	2738	3809	4294	4934
14	OVS	biasing	0	784	1348	2093	2738	3818	4294	5112
15	SVO	biasing	0	1024	1570	2276	2672	3858	4209	5079
15	OVS	neutral	0	1189	1570	2216	2672	3747	4209	5131
15	SVO	neutral	0	1084	1570	2192	2672	3726	4209	5055
15	OVS	biasing	0	1219	1570	2246	2672	3795	4209	5119
16	SVO	biasing	0	969	1510	2052	2513	3556	4006	4932
16	OVS	neutral	0	1298	1510	2072	2513	3579	4006	4708
16	SVO	neutral	0	1023	1510	2026	2513	3519	4006	4769
16	OVS	biasing	0	1012	1510	2040	2513	3651	4006	4737
17	SVO	biasing	0	847	1411	2077	2557	3612	4243	4931
17	OVS	neutral	0	1012	1411	2265	2557	3873	4243	5024
17	SVO	neutral	0	926	1411	2071	2557	3751	4243	4926
17	OVS	biasing	0	905	1411	2129	2557	3757	4243	5021
18	SVO	biasing	0	791	1344	1889	2394	3526	3991	4864
18	OVS	neutral	0	922	1344	1980	2394	3595	3991	5011
18	SVO	neutral	0	858	1344	1901	2394	3506	3991	4919
18	OVS	biasing	0	791	1344	1853	2394	3418	3991	4930
19	SVO	biasing	0	859	1391	2357	2868	3776	4308	5286
19	OVS	neutral	0	868	1391	2465	2868	3705	4308	5191
19	SVO	neutral	0	905	1391	2382	2868	3822	4308	5256
19	OVS	biasing	0	815	1391	2237	2868	3659	4308	5213

Table A.1: EXP 1

item	word order	prosody	NP1 on-set	NP1 off-set	verb on-set	verb off-set	adv. on-set	adv. off-set	NP2 on-set	NP2 off-set
20	SVO	biasing	0	1192	1870	2805	3286	4284	4842	5795
20	OVS	neutral	0	1349	1870	2821	3286	4311	4842	5658
20	SVO	neutral	0	1353	1870	2823	3286	4354	4842	5950
20	OVS	biasing	0	1337	1870	2704	3286	4272	4842	5686
21	SVO	biasing	0	1025	1600	2294	2808	3669	4122	4988
21	OVS	neutral	0	1208	1600	2275	2808	3655	4122	4741
21	SVO	neutral	0	1114	1600	2322	2808	3633	4122	4891
21	OVS	biasing	0	1150	1600	2194	2808	3622	4122	4755
22	SVO	biasing	0	863	1455	2282	2777	3643	4163	5170
22	OVS	neutral	0	1064	1455	2323	2777	3916	4163	5263
22	SVO	neutral	0	968	1455	2290	2777	3676	4163	5206
22	OVS	biasing	0	985	1455	2386	2777	3722	4163	5243
23	SVO	biasing	0	775	1355	1998	2567	3465	3910	4703
23	OVS	neutral	0	866	1355	1938	2567	3431	3910	4779
23	SVO	neutral	0	868	1355	2080	2567	3422	3910	4659
23	OVS	biasing	0	782	1355	1974	2567	3356	3910	4860
24	SVO	biasing	0	1021	1500	2386	2845	3709	4171	5037
24	OVS	neutral	0	1203	1500	2312	2845	3678	4171	4929
24	SVO	neutral	0	1013	1500	2335	2845	3684	4171	5057
24	OVS	biasing	0	1024	1500	2239	2845	3689	4171	4921

A.2 EXPERIMENT 2A

Table A.2: Word On- and Offsets Experiment 2a

item	word order	prosody	NP1 on-set	NP1 off-set	verb on-set	verb off-set	adv. on-set	adv. off-set	NP2 on-set	NP2 off-set
1	SVO	SVO biasing	0	1034	1338	2097	2398	3257	3600	4346
1	OVS	SVO biasing	0	968	1338	1995	2398	3131	3600	4345
1	SVO	OVS biasing	0	838	1338	1898	2398	3100	3600	4309
1	OVS	OVS biasing	0	938	1338	2000	2398	3144	3600	4438
2	SVO	SVO biasing	0	992	1388	2346	2715	3554	4032	4873
2	OVS	SVO biasing	0	974	1388	2309	2715	3512	4032	4655
2	SVO	OVS biasing	0	888	1388	2217	2715	3532	4032	4881
2	OVS	OVS biasing	0	978	1388	2331	2715	3557	4032	4835
3	SVO	SVO biasing	0	875	1313	2384	2782	3519	4084	4858
3	OVS	SVO biasing	0	913	1313	2272	2782	3548	4084	4716
3	SVO	OVS biasing	0	864	1313	2358	2782	3521	4084	4848
3	OVS	OVS biasing	0	905	1313	2328	2782	3522	4084	4681
4	SVO	SVO biasing	0	964	1376	2141	2613	3493	3853	4698
4	OVS	SVO biasing	0	984	1376	2078	2613	3397	3853	4735
4	SVO	OVS biasing	0	876	1376	2113	2613	3353	3853	4704
4	OVS	OVS biasing	0	1024	1376	2114	2613	3352	3853	4701
5	SVO	SVO biasing	0	1017	1465	2161	2646	3194	3695	4629
5	OVS	SVO biasing	0	1021	1465	2119	2646	3261	3695	4466
5	SVO	OVS biasing	0	966	1465	2145	2646	3196	3695	4654
5	OVS	OVS biasing	0	988	1465	2108	2646	3180	3695	4428
6	SVO	SVO biasing	0	795	1234	1830	2273	2836	3256	4121
6	OVS	SVO biasing	0	880	1234	1853	2273	2848	3256	3997
6	SVO	OVS biasing	0	734	1234	1773	2273	2756	3256	4087
6	OVS	OVS biasing	0	821	1234	1781	2273	2813	3256	3976
7	SVO	SVO biasing	0	1163	1630	2621	3053	3656	4120	4956
7	OVS	SVO biasing	0	1151	1630	2593	3053	3591	4120	4983
7	SVO	OVS biasing	0	1130	1630	2553	3053	3620	4120	4949

Table A.2: EXP 2a

item	word order	prosody	NP1 on-set	NP1 off-set	verb on-set	verb off-set	adv. on-set	adv. off-set	NP2 on-set	NP2 off-set
7	OVS	OVS biasing	0	1166	1630	2495	3053	3584	4120	4913
8	SVO	SVO biasing	0	943	1478	2252	2712	3281	3783	4509
8	OVS	SVO biasing	0	989	1478	2221	2712	3282	3783	4461
8	SVO	OVS biasing	0	978	1478	2212	2712	3283	3783	4493
8	OVS	OVS biasing	0	993	1478	2236	2712	3268	3783	4545
9	SVO	SVO biasing	0	948	1464	2334	2800	3374	3585	4728
9	OVS	SVO biasing	0	964	1464	2229	2800	3390	3858	4755
9	SVO	OVS biasing	0	964	1464	2300	2800	3358	3858	4824
9	OVS	OVS biasing	0	976	1464	2265	2800	3317	3858	4700
10	SVO	SVO biasing	0	891	1362	2331	2755	3368	3794	4673
10	OVS	SVO biasing	0	910	1362	2313	2755	3324	3794	4602
10	SVO	OVS biasing	0	862	1362	2255	2755	3294	3794	4624
10	OVS	OVS biasing	0	911	1362	2284	2755	3291	3794	4545
11	SVO	SVO biasing	0	989	1377	2103	2583	3176	3662	4521
11	OVS	SVO biasing	0	1037	1377	2116	2583	3163	3662	4461
11	SVO	OVS biasing	0	878	1377	2083	2583	3163	3662	4479
11	OVS	OVS biasing	0	955	1377	2030	2583	3156	3662	4445
12	SVO	SVO biasing	0	989	1515	2281	2802	3379	3882	4679
12	OVS	SVO biasing	0	985	1515	2253	2802	3390	3882	4709
12	SVO	OVS biasing	0	1016	1515	2303	2802	3382	3882	4662
12	OVS	OVS biasing	0	974	1515	2331	2802	3366	3882	4703
13	SVO	SVO biasing	0	855	1286	2150	2556	3677	4074	5007
13	OVS	SVO biasing	0	861	1286	2103	2556	3641	4074	4693
13	SVO	OVS biasing	0	786	1286	2056	2556	3574	4074	5047
13	OVS	OVS biasing	0	829	1286	2063	2556	3607	4074	4779
14	SVO	SVO biasing	0	831	1361	2140	2605	3728	4204	4925
14	OVS	SVO biasing	0	897	1361	2120	2605	3701	4204	4970
14	SVO	OVS biasing	0	861	1361	2105	2605	3704	4204	4912

Table A.2: EXP 2a

item	word order	prosody	NP1 on-set	NP1 off-set	verb on-set	verb off-set	adv. on-set	adv. off-set	NP2 on-set	NP2 off-set
14	OVS	OVS biasing	0	875	1361	2062	2605	3648	4204	5033
15	SVO	SVO biasing	0	999	1458	2214	2603	3723	4188	5131
15	OVS	SVO biasing	0	980	1458	2094	2603	3714	4188	5115
15	SVO	OVS biasing	0	958	1458	2103	2603	3687	4188	5126
15	OVS	OVS biasing	0	976	1458	2174	2603	3669	4188	5058
16	SVO	SVO biasing	0	1080	1491	2056	2500	3605	4013	4813
16	OVS	SVO biasing	0	1025	1491	2062	2500	3607	4013	4748
16	SVO	OVS biasing	0	991	1491	2000	2500	3514	4013	4860
16	OVS	OVS biasing	0	1017	1491	1997	2500	3534	4013	4756
17	SVO	SVO biasing	0	1011	1538	2229	2710	3813	4273	4963
17	OVS	SVO biasing	0	1074	1538	2160	2710	3812	4273	4998
17	SVO	OVS biasing	0	1038	1538	2210	2710	3773	4273	4965
17	OVS	OVS biasing	0	1031	1583	2083	2710	3786	4273	5048
18	SVO	SVO biasing	0	867	1253	1850	2303	3406	3928	4852
18	OVS	SVO biasing	0	915	1253	1861	2303	3393	3928	4842
18	SVO	OVS biasing	0	830	1253	1806	2303	3493	3928	4906
18	OVS	OVS biasing	0	864	1253	1759	2303	3329	3928	4800
19	SVO	SVO biasing	0	860	1320	2275	2773	3559	4106	5144
19	OVS	SVO biasing	0	888	1320	2278	2773	3638	4016	5086
19	SVO	OVS biasing	0	820	1320	2273	2773	3606	4106	5074
19	OVS	OVS biasing	0	854	1320	2201	2773	3647	4106	5015
20	SVO	SVO biasing	0	1153	1623	2467	2929	3771	4201	4991
20	OVS	SVO biasing	0	1153	1623	2429	2929	3804	4201	4962
20	SVO	OVS biasing	0	1124	1623	2428	2929	3701	4201	4986
20	OVS	OVS biasing	0	1195	1623	2366	2929	3758	4201	4844
21	SVO	SVO biasing	0	930	1357	2065	2485	3350	3859	4571
21	OVS	SVO biasing	0	963	1357	2045	2485	3338	3859	4490
21	SVO	OVS biasing	0	858	1357	1986	2485	3359	3859	4633

Table A.2: EXP 2a

item	word order	prosody	NP1 on-set	NP1 off-set	verb on-set	verb off-set	adv. on-set	adv. off-set	NP2 on-set	NP2 off-set
21	OVS	OVS biasing	0	954	1357	1984	2485	3342	3859	4508
22	SVO	SVO biasing	0	989	1553	2384	2802	3700	4165	5036
22	OVS	SVO biasing	0	972	1553	2378	2802	3589	4165	5041
22	SVO	OVS biasing	0	1052	1553	2303	2802	3665	4165	5137
22	OVS	OVS biasing	0	1126	1553	2315	2802	3627	4165	5018
23	SVO	SVO biasing	0	786	1300	1989	2440	3349	3812	4644
23	OVS	SVO biasing	0	800	1300	1923	2440	3303	3812	4710
23	SVO	OVS biasing	0	800	1300	1940	2440	3312	3812	4657
23	OVS	OVS biasing	0	807	1300	1906	2440	3349	3812	4743
24	SVO	SVO biasing	0	1043	1605	2474	2959	3767	4297	5164
24	OVS	SVO biasing	0	1053	1605	2406	2959	3762	4297	5042
24	SVO	OVS biasing	0	1105	1605	2459	2959	3797	4297	5211
24	OVS	OVS biasing	0	1061	1605	2426	2959	3780	4297	5082

A.3 EXPERIMENT 2B AND EXPERIMENT 3

The following table shows all on- and offsets for Experiment 2b and also includes the on- and offsets for Experiment 3 (the ambiguous OVS (aOVS) sentences).

Table A.3: Word On- and Offsets Experiment 2b / 3

item	word order	prosody	NP1 on-set	NP1 off-set	verb on-set	verb off-set	adv. on-set	adv. off-set	NP2 on-set	NP2 off-set
1	aOVS	SVO biasing	0	1104	1438	2097	2498	3257	3700	4346
1	OVS	SVO biasing	0	968	1438	2095	2498	3231	3700	4445
1	aOVS	OVS biasing	0	1149	1438	2147	2498	3289	3700	4403
1	OVS	OVS biasing	0	938	1438	2100	2498	3244	3700	4538
2	aOVS	SVO biasing	0	1138	1488	2346	2815	3689	4132	5027
2	OVS	SVO biasing	0	974	1488	2409	2815	3612	4132	4755

Table A.3: EXPs 2b / 3

item	word order	prosody	NP1 on-set	NP1 off-set	verb on-set	verb off-set	adv. on-set	adv. off-set	NP2 on-set	NP2 off-set
2	aOVS	OVS biasing	0	1105	1488	2365	2815	3615	4132	5071
2	OVS	OVS biasing	0	978	1488	2431	2815	3657	4132	4935
3	aOVS	SVO biasing	0	1073	1464	2346	2800	3416	3858	4811
3	OVS	SVO biasing	0	964	1464	2229	2800	3390	3858	4755
3	aOVS	OVS biasing	0	1158	1464	2369	2800	3399	3858	4818
3	OVS	OVS biasing	0	976	1464	2265	2800	3317	3858	4700
4	aOVS	SVO biasing	0	1178	1478	2257	2712	3381	3783	4499
4	OVS	SVO biasing	0	989	1478	2221	2712	3282	3783	4461
4	aOVS	OVS biasing	0	1126	1478	2263	2712	3273	3783	4466
4	OVS	OVS biasing	0	993	1478	2236	2712	3268	3783	4545
5	aOVS	SVO biasing	0	1099	1465	2171	2646	3265	3695	4536
5	OVS	SVO biasing	0	1021	1465	2119	2646	3261	3695	4466
5	aOVS	OVS biasing	0	1057	1465	2189	2646	3227	3695	4501
5	OVS	OVS biasing	0	988	1465	2108	2646	3180	3695	4428
6	aOVS	SVO biasing	0	1028	1334	1902	2373	2998	3356	4170
6	OVS	SVO biasing	0	880	1334	1953	2373	2948	3356	4097
6	aOVS	OVS biasing	0	954	1334	1960	2373	2966	3356	4092
6	OVS	OVS biasing	0	821	1334	1881	2373	2913	3356	4092
7	aOVS	SVO biasing	0	986	1286	2162	2556	3741	4074	4934
7	OVS	SVO biasing	0	861	1286	2103	2556	3641	4074	4693
7	aOVS	OVS biasing	0	976	1286	2152	2556	3771	4074	4984
7	OVS	OVS biasing	0	829	1286	2063	2556	3607	4074	4779
8	aOVS	SVO biasing	0	1118	1515	2298	2802	3360	3882	4667
8	OVS	SVO biasing	0	985	1515	2253	2802	3390	3882	4709
8	aOVS	OVS biasing	0	1153	1515	2314	2802	3369	3882	4697
8	OVS	OVS biasing	0	974	1515	2331	2802	3366	3882	4703
9	aOVS	SVO biasing	0	1011	1313	2422	2782	3635	4084	4777
9	OVS	SVO biasing	0	913	1313	2272	2782	3548	4084	4716

Table A.3: EXPs 2b / 3

item	word order	prosody	NP1 on-set	NP1 off-set	verb on-set	verb off-set	adv. on-set	adv. off-set	NP2 on-set	NP2 off-set
9	aOVS	OVS biasing	0	904	1313	2344	2782	3636	4084	4782
9	OVS	OVS biasing	0	905	1313	2328	2782	3522	4084	4681
10	aOVS	SVO biasing	0	1003	1376	2115	2613	3429	3853	4758
10	OVS	SVO biasing	0	984	1376	2078	2613	3397	3853	4735
10	aOVS	OVS biasing	0	877	1376	2198	2613	3462	3853	4725
10	OVS	OVS biasing	0	1024	1376	2114	2613	3352	3853	4701
11	aOVS	SVO biasing	0	954	1377	2174	2583	3210	3662	4495
11	OVS	SVO biasing	0	1037	1377	2116	2583	3163	3662	4461
11	aOVS	OVS biasing	0	901	1377	2135	2583	3252	3662	4552
11	OVS	OVS biasing	0	955	1377	2030	2583	3156	3662	4445
12	aOVS	SVO biasing	0	909	1362	2280	2755	3355	3794	4673
12	OVS	SVO biasing	0	910	1362	2313	2755	3324	3794	4602
12	aOVS	OVS biasing	0	991	1362	2423	2755	3365	3794	4709
12	OVS	OVS biasing	0	911	1362	2284	2755	3291	3794	4545
13	aOVS	SVO biasing	0	1320	1630	2672	3053	3682	4120	5051
13	OVS	SVO biasing	0	1151	1630	2593	3053	3591	4120	4983
13	aOVS	OVS biasing	0	1310	1630	2602	3053	3632	4120	5030
13	OVS	OVS biasing	0	1166	1630	2495	3053	3584	4120	4913
14	aOVS	SVO biasing	0	1160	1461	2263	2705	3850	4304	5161
14	OVS	SVO biasing	0	897	1461	2220	2705	3801	4304	5070
14	aOVS	OVS biasing	0	1099	1461	2209	2705	3908	4304	5197
14	OVS	OVS biasing	0	875	1461	2162	2705	3748	4304	5133
15	aOVS	SVO biasing	0	1100	1420	2370	2873	3869	4206	5221
15	OVS	SVO biasing	0	888	1420	2378	2873	3738	4206	5186
15	aOVS	OVS biasing	0	1118	1420	2372	2873	3839	4206	5221
15	OVS	OVS biasing	0	854	1420	2301	2873	3746	4206	5115
16	aOVS	SVO biasing	0	1359	1673	2555	2979	3977	4251	4926
16	OVS	SVO biasing	0	1153	1673	2479	2979	3854	4251	5012

Table A.3: EXPs 2b / 3

item	word order	prosody	NP1 onset	NP1 offset	verb onset	verb offset	adv. onset	adv. offset	NP2 onset	NP2 offset
16	aOVS	OVS biasing	0	1350	1673	2565	2979	3969	4251	4935
16	OVS	OVS biasing	0	1195	1673	2416	2979	3808	4251	4894
17	aOVS	SVO biasing	0	1064	1538	2316	2710	3858	4273	5118
17	OVS	SVO biasing	0	1074	1538	2160	2710	3812	4273	4998
17	aOVS	OVS biasing	0	1148	1538	2316	2710	3865	4273	5096
17	OVS	OVS biasing	0	1031	1538	2083	2710	3786	4273	5048
18	aOVS	SVO biasing	0	998	1353	1877	2403	3580	4028	4953
18	OVS	SVO biasing	0	915	1353	1961	2403	3493	4028	4942
18	aOVS	OVS biasing	0	1006	1353	1925	2403	3601	4028	4941
18	OVS	OVS biasing	0	864	1353	1859	2403	3429	4028	4900
19	aOVS	SVO biasing	0	986	1300	1965	2440	3452	3812	4796
19	OVS	SVO biasing	0	800	1300	1923	2440	3303	3812	4710
19	aOVS	OVS biasing	0	968	1300	1921	2440	3443	3812	4795
19	OVS	OVS biasing	0	807	1300	1906	2440	3349	3812	4743
20	aOVS	SVO biasing	0	1123	1553	2430	2802	3765	4165	5054
20	OVS	SVO biasing	0	972	1553	2378	2802	3589	4165	5041
20	aOVS	OVS biasing	0	1172	1553	2351	2802	3792	4165	5100
20	OVS	OVS biasing	0	1126	1553	2315	2802	3627	4165	5018
21	aOVS	SVO biasing	0	1246	1558	2273	2703	3849	4288	5285
21	OVS	SVO biasing	0	980	1558	2194	2703	3814	4288	5279
21	aOVS	OVS biasing	0	1249	1558	2213	2703	3889	4288	5126
21	OVS	OVS biasing	0	976	1558	2274	2703	3769	4288	5158
22	aOVS	SVO biasing	0	1066	1491	2066	2500	3709	4013	4754
22	OVS	SVO biasing	0	1025	1491	2062	2500	3607	4013	4748
22	aOVS	OVS biasing	0	1001	1491	2037	2500	3657	4013	4768
22	OVS	OVS biasing	0	1017	1491	1997	2500	3534	4013	4756
23	aOVS	SVO biasing	0	1048	1605	2455	2959	3907	4297	5102
23	OVS	SVO biasing	0	1053	1605	2406	2959	3762	4297	5042

Table A.3: EXPs 2b / 3

item	word order	prosody	NP1 on-set	NP1 off-set	verb on-set	verb off-set	adv. on-set	adv. off-set	NP2 on-set	NP2 off-set
23	aOVS	OVS biasing	0	1105	1605	2459	2959	3797	4297	5211
23	OVS	OVS biasing	0	1061	1605	2426	2959	3780	4297	5082
24	aOVS	SVO biasing	0	1202	1507	2240	2635	3656	4009	4721
24	OVS	SVO biasing	0	963	1507	2195	2635	3488	4009	4640
24	aOVS	OVS biasing	0	1188	1507	2212	2635	3620	4009	4698
24	OVS	OVS biasing	0	954	1507	2134	2635	3492	4009	4658

B

ITEMS EXPERIMENT 1

Experimental sentences for each visual scene: a) SVO sentence structure and b) OVS sentence structure.

Item 1
1a) Der Bär schubst gerade den Wurm.
1b) Den Bären schubst gerade der Stier.
Item 2
2a) Der Vogel beköstigt gerade den Delfin.
2b) Den Vogel beköstigt gerade der Krebs.
Item3
3a) Der Affe fotografiert gerade den Igel.
3b) Den Affen fotografiert gerade der Hai.
Item4
4a) Der Kater kitzelt gerade den Adler.
4b) Den Kater kitzelt gerade der Hamster.
Item 5
5a) Der Hase krönt bald den Koala.
5b) Den Hasen krönt bald der Dino.
Item 6
6a) Der Wolf boxt bald den Tiger.
6b) Den Wolf boxt bald der Hahn.
Item 7
7a) Der Grashüpfer verarztet bald den Stier.
7b) Den Grashüpfer verarztet bald der Delfin.
Item 8
8a) Der Bär begießt bald den Hai.
8b) Den Bären begießt bald der Wurm.
Item9
9a) Der Vogel besprüht bald den Hamster.
9b) Den Vogel besprüht bald der Koala.
Item 10
10a) Der Affe verzaubert bald den Krebs.
10b) Den Affen verzaubert bald der Adler.

Item11

11a) Der Kater bedient bald den Dino.

11b) Den Kater bedient bald der Igel.

Item 12

12a) Der Hase beschenkt bald den Hahn.

12b) Den Hasen beschenkt bald der Tiger.

Item 13

13a) Der Wolf beliefert als nächstes den Pinguin.

13b) Den Wolf beliefert als nächstes der Schwan.

Item 14

14a) Der Frosch bewirft als nächstes den Hund.

14b) Den Frosch bewirft als nächstes der Gepard.

Item 15

15a) Der Eisbär filmt als nächstes den Papagei.

15b) Den Eisbären filmt als nächstes der Gänserich.

Item 16

16a) Der Elefant kämmt als nächstes den Esel.

16b) Den Elefanten kämmt als nächstes der Rabe.

Item 17

17a) Der Löwe föhnt als nächstes den Fisch.

17b) Den Löwen föhnt als nächstes der Spatz.

Item 18

18a) Der Fuchs küsst als nächstes den Schmetterling.

18b) Den Fuchs küsst als nächstes der Maulwurf.

Item 19

19a) Der Frosch beobachtet sogleich den Gänserich.

19b) Den Frosch beobachtet sogleich der Schmetterling.

Item 20

20a) Der Grashüpfer befragt sogleich den Schwan.

20b) Den Grashüpfer befragt sogleich der Hund.

Item 21

21a) Der Eisbär malt sogleich den Spatz.

21b) Den Eisbären malt sogleich der Fisch.

Item 22

22a) Der Löwe beschirmt sogleich den Maulwurf.

22b) Den Löwen beschirmt sogleich der Papagei.

Item 23

23a) Der Fuchs wäscht sogleich den Raben.

23b) Den Fuchs wäscht sogleich der Pinguin.

Item 24

24a) Der Elefant zeichnet sogleich den Gepard.

24b) Den Elefanten zeichnet sogleich der Esel.

C

ITEMS EXPERIMENT 2 AND 3

Experimental sentences for each visual scene: a) SVO and b) OVS sentence structure.

C.1 EXPERIMENT 2A: UNAMBIGUOUS

Item 1
1a) Der Käfer schubst gerade den Wurm.
1b) Den Käfer schubst gerade der Stier.
Item 2
2a) Der Vogel beköstigt gerade den Delfin.
2b) Den Vogel beköstigt gerade der Krebs.
Item 3
3a) Der Biber fotografiert gerade den Igel.
3b) Den Biber fotografiert gerade der Hai.
Item 4
4a) Der Kater kitzelt gerade den Adler.
4b) Den Kater kitzelt gerade der Hamster.
Item 5
5a) Der Storch krönt bald den Koala.
5b) Den Storch krönt bald der Dino.
Item 6
6a) Der Wolf boxt bald den Tiger.
6b) Den Wolf boxt bald der Hahn.
Item 7
7a) Der Grashüpfer verarztet bald den Stier.
7b) Den Grashüpfer verarztet bald der Delfin.
Item 8
8a) Der Käfer begießt bald den Hai.
8b) Den Käfer begießt bald der Wurm.
Item 9
9a) Der Vogel besprüht bald den Hamster.
9b) Den Vogel besprüht bald der Koala.
Item 10
10a) Der Biber verzaubert bald den Krebs.

10b) Den Biber verzaubert bald der Adler.

Item 11

11a) Der Kater bedient bald den Dino.

11b) Den Kater bedient bald der Igel.

Item 12

12a) Der Storch beschenkt bald den Hahn.

12b) Den Storch beschenkt bald der Tiger.

Item 13

13a) Der Wolf beliefert als nächstes den Pinguin.

13b) Den Wolf beliefert als nächstes der Schwan.

Item 14

14a) Der Frosch bewirft als nächstes den Hund.

14b) Den Frosch bewirft als nächstes der Gepard.

Item 15

15a) Der Panda filmt als nächstes den Papagei.

15b) Den Panda filmt als nächstes der Gänserich.

Item 16

16a) Der Kraken kämmt als nächstes den Esel.

16b) Den Kraken kämmt als nächstes der Rabe.

Item 17

17a) Der Seehund föhnt als nächstes den Fisch.

17b) Den Seehund föhnt als nächstes der Spatz.

Item 18

18a) Der Fuchs küsst als nächstes den Schmetterling.

18b) Den Fuchs küsst als nächstes der Maulwurf.

Item 19

19a) Der Frosch beobachtet sogleich den Gänserich.

19b) Den Frosch beobachtet sogleich der Schmetterling.

Item 20

20a) Der Grashüpfer befragt sogleich den Schwan.

20b) Den Grashüpfer befragt sogleich der Hund.

Item 21

21a) Der Panda malt sogleich den Spatz.

21b) Den Panda malt sogleich der Fisch.

Item 22

22a) Der Seehund beschirmt sogleich den Maulwurf.

22b) Den Seehund beschirmt sogleich der Papagei.

Item 23
23a) Der Fuchs wäscht sogleich den Raben.
23b) Den Fuchs wäscht sogleich der Pinguin.
Item 24
24a) Der Kraken zeichnet sogleich den Gepard.
24b) Den Kraken zeichnet sogleich der Esel.

C.2 Experiment 2b: Unambiguous and Ambiguous, Experiment 3: Ambiguous

For Experiment 2b, we changed the order of the sentences and images: for Experiment 2b, see the unambiguous OVS (b) and ambiguous OVS (b′) sentences, for Experiment 3, see only ambiguous OVS sentences (b′).

Item 1
1b) Den Käfer schubst gerade der Stier.
1b′) Das Käferchen schubst gerade der Stier.
Item 2
2b) Den Vogel beköstigt gerade der Krebs.
2b′) Das Vögelchen beköstigt gerade der Krebs.
Item 3
3b) Den Vogel besprüht bald der Koala.
3b′) Das Vögelchen besprüht bald der Koala.
Item 4
4b) Den Käfer begießt bald der Wurm.
4b′) Das Käferchen begießt bald der Wurm.
Item 5
5b) Den Storch krönt bald der Dino.
5b′) Die Störchin krönt bald der Dino.
Item 6
6b) Den Wolf boxt bald der Hahn.
6b′) Die Wölfin boxt bald der Hahn.
Item 7
7b) Den Wolf beliefert als nächstes der Schwan.
7b′) Die Wölfin beliefert als nächstes der Schwan.
Item 8
8b) Den Storch beschenkt bald der Tiger.
8b′) Die Störchin beschenkt bald der Tiger.
Item 9
9b) Den Biber fotografiert gerade der Hai.

9b′) Die Biberin fotografiert gerade der Hai.

Item 10
10b) Den Kater kitzelt gerade der Hamster.
10b′) Die Katze kitzelt gerade der Hamster.
Item 11
11b) Den Kater bedient bald der Igel.
11b′) Die Katze bedient bald der Igel.
Item 12
12b) Den Biber verzaubert bald der Adler.
12b′) Die Biberin verzaubert bald der Adler.
Item 13
13b) Den Grashüpfer verarztet bald der Delfin.
13b′) Das Grashüpferchen verarztet bald der Delfin.
Item 14
14b) Den Frosch bewirft als nächstes der Gepard.
14b′) Das Fröschchen bewirft als nächstes der Gepard.
Item 15
15b) Den Frosch beobachtet sogleich der Schmetterling.
15b′) Das Fröschchen beobachtet sogleich der Schmetterling.
Item 16
16b) Den Grashüpfer befragt sogleich der Hund.
16b′) Das Grashüpferchen befragt sogleich der Hund.
Item 17
17b) Den Seehund föhnt als nächstes der Spatz.
17b′) Die Seehündin föhnt als nächstes der Spatz.
Item 18
18b) Den Fuchs küsst als nächstes der Maulwurf.
18b′) Die Füchsin küsst als nächstes der Maulwurf.
Item 19
19b) Den Fuchs wäscht sogleich der Pinguin.
19b′) Die Füchsin wäscht sogleich der Pinguin.
Item 20
20b) Den Seehund beschirmt sogleich der Papagei.
20b′) Die Seehündin beschirmt sogleich der Papagei.
Item 21
21b) Den Panda filmt als nächstes der Gänserich.
21b′) Die Pandabärin filmt als nächstes der Gänserich.

Item 22

22b) Den Kraken kämmt als nächstes der Rabe.

22b′) Die Krakin kämmt als nächstes der Rabe.

Item 23

23b) Den Kraken zeichnet sogleich der Esel.

23b′) Die Krakin zeichnet sogleich der Esel.

Item 24

24b) Den Panda malt sogleich der Fisch.

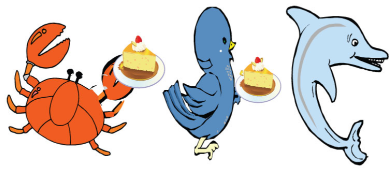
24b′) Die Pandabärin malt sogleich der Fisch.

D

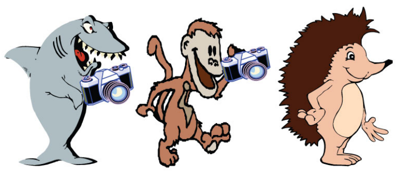
ITEMS EXPERIMENT 1: IMAGES

The experimental items were balanced either left or right. The images below show the right balancing (all animal characters facing right).

Item 1

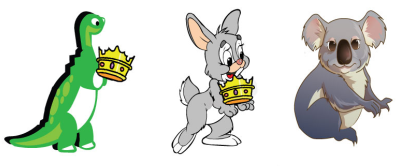


Item 2

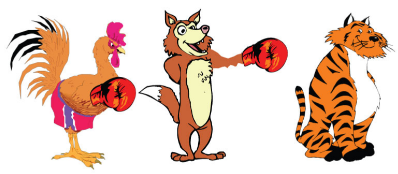


Item 3

Item 4

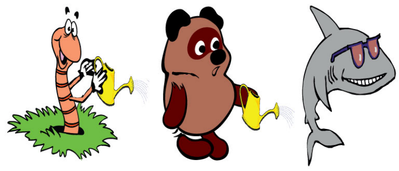


Item 5

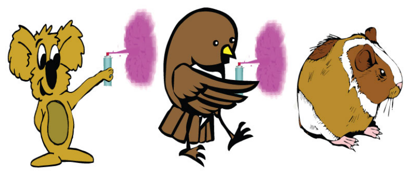


Item 6

Item 7

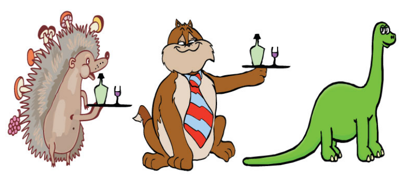


Item 8

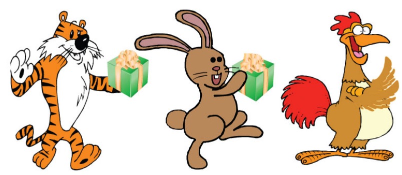


Item 9

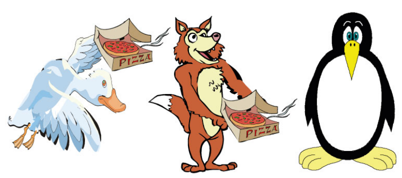
Item 10



Item 11

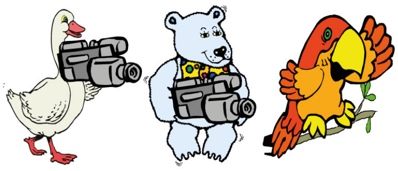


Item 12

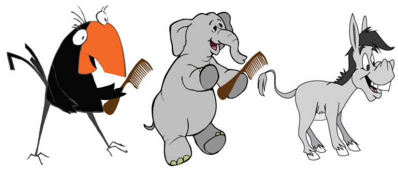


Item 13

Item 14

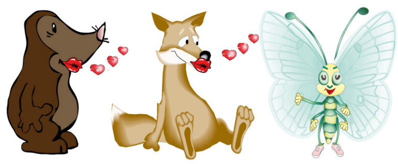


Item 15



Item 16

Item 17



Item 18

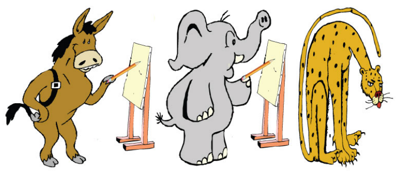
Item 19

Item 20

Item 21

Item 22

Item 23

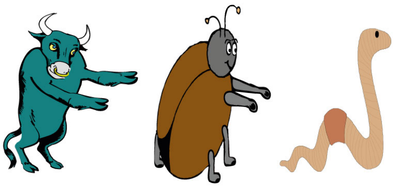


Item 24

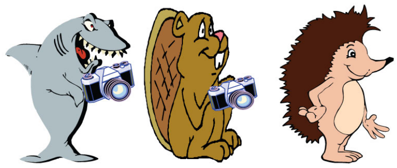
E

ITEMS EXPERIMENT 2: IMAGES

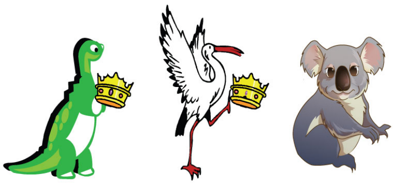
For Experiment 2, we exchanged 12 ambiguous characters (see item 1, 3, 5, 8, 10, 12, 15, 16, 17, 21, 22, and 24 below). The other images did not change (see Appendix D for items 2, 4, 6, 7, 9, 11, 13, 14, 18, 19, 20, and 23). For Experiment 2b, the order of the items changed but not the images themselves.



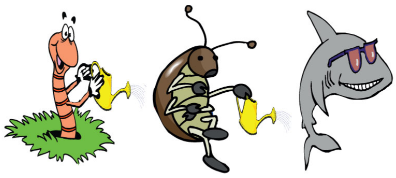
Item 1



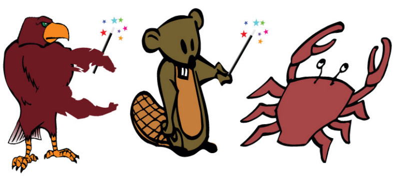
Item 3



Item 5



Item 8

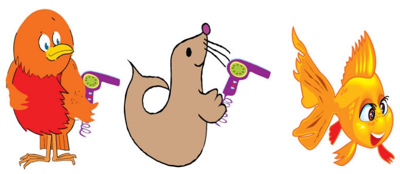


Item 10

Item 12

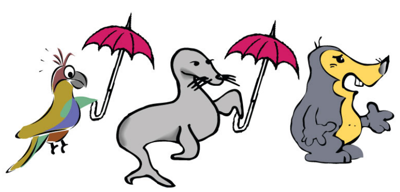
Item 15

Item 16

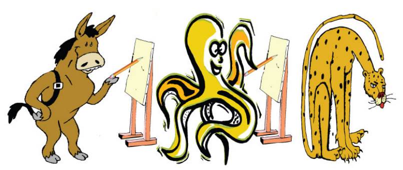


Item 17

Item 21



Item 22

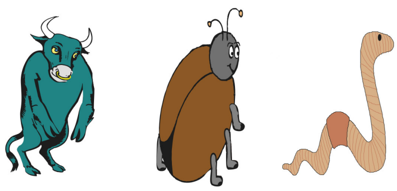


Item 24

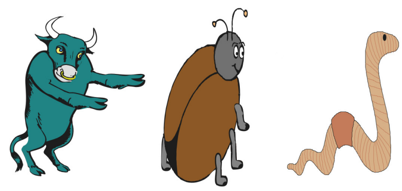
F

ITEMS EXPERIMENT 3: IMAGES

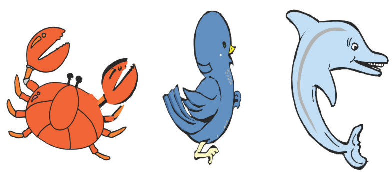
Experimental items for the two conditions a) no action no wiggle and c) one action no wiggle. In conditions b) no action one wiggle and d) one action one wiggle, the target character (the character on the left in the examples) wiggles up and down a couple of inches (see procedure Experiment 3: Figure 9.1).



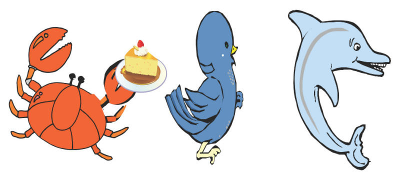
Item 1 no action



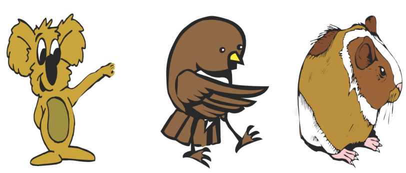
Item 1 one action



Item 2 no action



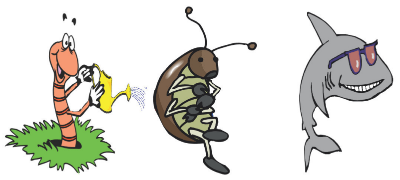
Item 2 one action



Item 3 no action

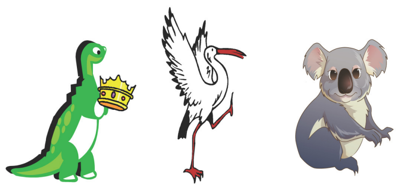
Item 3 one action

Item 4 no action



Item 4 one action

Item 5 no action



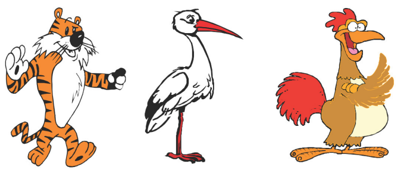
Item 5 one action

Item 6 no action

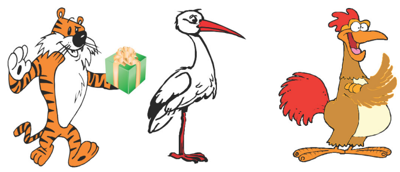
Item 6 one action

Item 7 no action

Item 7 one action



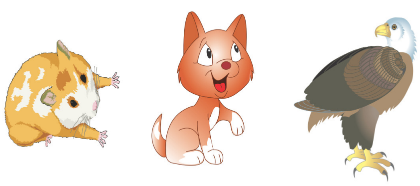
Item 8 no action



Item 8 one action

Item 9 no action

Item 9 one action

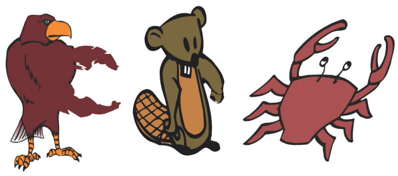


Item 10 no action

Item 10 one action

Item 11 no action

Item 11 one action

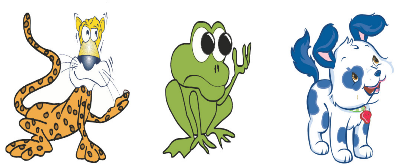


Item 12 no action

Item 12 one action

Item 13 one action

Item 13 one action



Item 14 one action

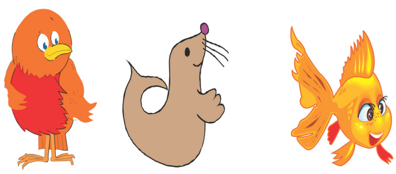
Item 14 one action

Item 15 no action

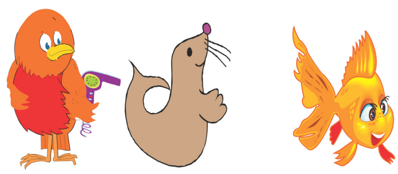
Item 15 one action

Item 16 no action

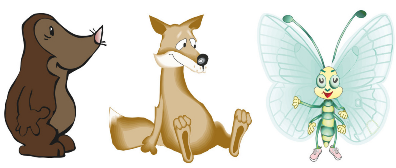
Item 16 one action



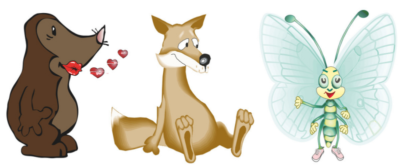
Item 17 one action



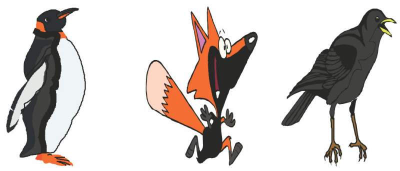
Item 17 one action



Item 18 one action

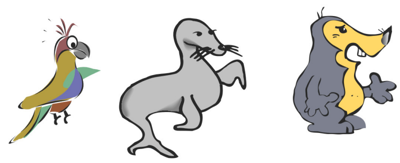


Item 18 one action

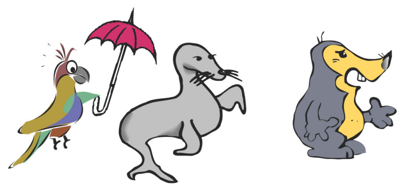


Item 19 no action

Item 19 one action



Item 20 no action



Item 20 one action

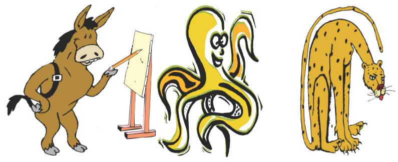
Item 21 one action

Item 21 one action

Item 22 one action

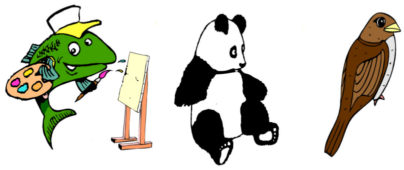
Item 22 one action

Item 23 no action



Item 23 one action

Item 24 no action



Item 24 one action

G

K-ABC RESULTS: SUBTEST

The following tables provide the responses for all three K-ABC subtest by participant. 1 represents a correct response and 0 a false response.

Table G.1: K-ABC Results: Experiment 1

Participant	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Word Order	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	0	1	1	1	1	1	1	1	0	1	1	1	1	0	1	1	1	1	1	1	0	1	1	1
	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1	0	1	1	1
	0	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	0	1	1
Spatial Memory	1	1	1	1	1	1	0	0	1	0	1	0	1	1	1	1	0	0	1	1	1	1	1	1
	1	1	1	1	1	1	0	0	1	1	1	0	1	1	1	0	0	0	1	1	1	1	1	1
	1	0	1	1	1	1	0	1	1	0	1	1	1	1	1	1	0	1	0	1	1	1	1	1
	1	1	1	1	1	1	0	1	0	1	0	0	1	0	1	1	0	0	1	1	1	1	1	1
	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1	1	0	0	1	0	1	1	1	1
Number Recall	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	1	1	1	1	1	0	0	1	1	0	1	0	1	1	0	1	0	1	1	1	0	1	1	1
	0	1	1	1	1	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1
	0	1	1	1	1	0	0	0	1	0	1	0	0	1	0	1	0	1	1	1	0	1	1	1
	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	1	1	0	0	1	0	1	1	1
	0	0	0	1	1	0	0	0	0	0	1	0	0	1	1	1	1	0	1	1	0	1	1	1
	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0	1	0	1	0	1	0	1	0	1

Table G.2: K-ABC Results: Experiment 3

Participant	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Word Order	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1
	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1
	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	1	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	1	1	1	1	0	1
	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1
Spatial Memory	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	1	0	0	1	1
	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	0	1	1	1	1	1	1
	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1
	1	0	1	1	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1
Number Recall	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1
	1	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1
	1	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1
	0	0	1	1	1	1	0	0	0	1	1	1	1	1	1	1	0	0	1	0	1	1	1	1
	1	0	1	1	1	1	1	1	0	0	0	1	0	1	1	1	0	1	0	0	1	1	1	1
	1	0	1	1	1	1	0	0	0	0	0	0	1	1	1	1	0	1	0	0	1	1	1	1
	0	0	1	1	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1	0	0
	0	0	1	0	1	1	0	0	0	0	0	0	1	0	1	0	0	1	0	0	1	1	0	0
	0	0	1	1	1	1	0	0	0	0	0	0	1	0	1	0	0	0	0	0	1	1	1	0

H

MAXIMAL CONVERGING MODELS

The following table shows the maximal converging models divided by experiment. *A* stands for adults, *C* for children, *A/C* for adult-children comparisons, *act* for active voice comprehension questions, and *pass* for passive voice comprehension questions.

Table H.1: Maximal Converging Models in all Experiments

<table>
<tr><th>EXP</th><th>Model</th></tr>
<tr><td>1 A</td><td>accuracy ~ wordorderC + prosodyC + (1 | subject)</td></tr>
<tr><td>1 C</td><td>accuracy ~ wordorderC + prosodyC + (1 | subject) + (1 | item)</td></tr>
<tr><td>1 A/C</td><td>accuracy ~ wordorderC + prosodyC + ageC + (1 | subject) + (1 | item)</td></tr>
<tr><td>2a A</td><td>accuracy ~ wordorderC * prosodyC + (1 | subject) + (1 | item)</td></tr>
<tr><td>2b A</td><td>accuracy ~ wordorderC + prosodyC + (1 | subject) + (1 | item)</td></tr>
<tr><td>3 A</td><td>accuracy ~ actionC * wiggleC + (1 | subject) + (1 | item)</td></tr>
<tr><td>3 C</td><td>accuracy ~ actionC + wiggleC + actpasC + (1 | subject) + (1 | item)</td></tr>
<tr><td>3 C act</td><td>accuracy ~ actionC + wiggleC (1 | subject) + (1 | item)</td></tr>
<tr><td>3 C pass</td><td>accuracy ~ actionC + wiggleC + (1 | subject)</td></tr>
<tr><td>3 A/C</td><td>accuracy ~ actionC * wiggleC * ageC + (1 | subject) + (1 | item)</td></tr>
</table>

BIBLIOGRAPHY

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of memory and language*, *38*(4), 419–439.

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264.

Altmann, G. T., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*(4), 502–518.

Altmann, G. T., & Kamide, Y. (2009). Discourse-mediation of the mapping between language and the visual world: Eye movements and mental representation. *Cognition*, *111*(1), 55–71.

Ambridge, B., & Lieven, E. V. (2011). *Child language acquisition: Contrasting theoretical approaches*. Cambridge, England: Cambridge University Press.

Arai, M., Van Gompel, R. P., & Scheepers, C. (2007). Priming ditransitive structures in comprehension. *Cognitive Psychology*, *54*(3), 218–250.

Arnold, J. E. (2008). The bacon not the bacon: How children and adults understand accented and unaccented noun phrases. *Cognition*, *108*(1), 69–99.

Arnold, J. E., Brown-Schmidt, S., & Trueswell, J. C. (2007). Children's use of gender and order-of-mention during pronoun comprehension. *Language and Cognitive Processes*, *22*(4), 527–565.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The celex lexical database. linguistic data consortium. *University of Pennsylvania, Philadelphia*.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3), 255–278.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *Journal of Statistics Software*, *67*(1), 1-48.

Bates, E., & MacWhinney, B. (1987). Competition, variation, and language learning. *Mechanisms of language acquisition*, 157–193.

Bates, E., & MacWhinney, B. (1989). Functionalism and the competition model.

The crosslinguistic study of sentence processing, *3*, 73–112.

Baumann, S., Becker, J., Grice, M., & Mücke, D. (2007). Tonal and articulatory marking of focus in german. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th international congress of phonetic sciences* (pp. 1029–1032). Saarbrücken.

Baumann, S., Grice, M., & Benzmüller, R. (2000). Gtobi-a phonological system for the transcription of german intonation. In *Prosody* (pp. 21–28).

Bencini, G. M., & Valian, V. V. (2008). Abstract sentence representations in 3-year-olds: Evidence from language production and comprehension. *Journal of Memory and Language*, *59*(1), 97–113.

Bornkessel, I., Schlesewsky, M., & Friederici, A. D. (2002). Grammar overrides frequency: Evidence from the online processing of flexible word order. *Cognition*, *85*(2), B21–B30.

Borovsky, A., Elman, J. L., & Fernald, A. (2012). Knowing a lot for one's age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of experimental child psychology*, *112*(4), 417–436.

Brandt, S., Lieven, E., & Tomasello, M. (2016). German children's use of word order and case marking to interpret simple and complex sentences: testing differences between constructions and lexical items. *Language Learning and Development*, *12*(2), 156–182.

Brennan, S. E., Chen, X., Dickinson, C. A., Neider, M. B., & Zelinsky, G. J. (2008). Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition*, *106*(3), 1465–1477.

Büring, D., & Gutiérrez-Bravo, R. (2001). Focus-related constituent order variation without the NSR: A prosody-based crosslinguistic analysis. *Syntax at Santa Cruz*, *3*, 41–58.

Cameron-Faulkner, T., Lieven, E., & Tomasello, M. (2003). A construction based analysis of child directed speech. *Cognitive Science*, *27*(6), 843–873.

Carlson, K. (2009). How prosody influences sentence comprehension. *Language and Linguistics Compass*, *3*(5), 1188–1200.

Carminati, M. N., & Knoeferle, P. (2013). Effects of speaker emotional facial expression and listener age on incremental sentence processing. *PloS ONE*, *8*(9), e72559.

Chambers, C. G., Tanenhaus, M. K., & Magnuson, J. S. (2004). Actions and affordances in syntactic ambiguity resolution. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(3), 687–696.

Choi, Y., & Trueswell, J. C. (2010). Children's (in)ability to recover from garden paths in a verb-final language: Evidence for developing control in sentence processing. *Journal of experimental child psychology*, *106*(1), 41–61.

Clahsen, H. (1984). Der Erwerb von Kasusmarkierungen in der deutschen Kindersprache. *Linguistische Berichte*(89), 1–31.

Coco, F., Moreno I & Keller. (2015). The interaction of visual and linguistic saliency during syntactic ambiguity resolution. *The Quarterly Journal of Experimental Psychology*, *68*(1), 46–74.

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*(1), 84–107.

Dafydd, G. (1998). Intonation in German. In D. Hirst & A. Di Cristo (Eds.), *Intonation systems: A survey of twenty languages* (pp. 78–95). Cambridge, MA: Cambridge University Press.

Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language*, *47*(2), 292–314.

Dittmar, M., Abbot-Smith, K., Lieven, E., & Tomasello, M. (2008a). German children's comprehension of word order and case marking in causative sentences. *Child Development*, *79*(4), 1152–1167.

Dittmar, M., Abbot-Smith, K., Lieven, E., & Tomasello, M. (2008b). Young german children's early syntactic competence: a preferential looking study. *Developmental Science*, *11*(4), 575–582.

Du Bois, J. W., Kumpf, L. E., & Ashby, W. J. (2003). *Preferred argument structure: Grammar as architecture for function* (Vol. 14). Amsterdam: John Benjamins Publishing.

Dürr, M., & Schlobinski, P. (2006). *Deskriptive Linguistik: Grundlagen und Methoden*. Göttingen, Germany: Vandenhoeck & Ruprecht.

Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of psycholinguistic research*, *24*(6), 409–436.

Eisenbeiss, S., Bartke, S., & Clahsen, H. (2006). Structural and lexical case in child German: Evidence from language-impaired and typically developing children. *Language acquisition*, *13*(1), 3–32.

Fanselow, G. (2000). Optimal exceptions. In B. Stiebels & D. Wunderlich (Eds.), *Lexicon in focus* (pp. 173–209). Berlin: Akademie Verlag.

Fernald, A. (1993). Approval and disapproval: Infant responsiveness to vocal affect in familiar and unfamiliar languages. *Child development*, *64*(3), 657–674.

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental psychology*, *42*(1), 98.

Fernald, A., Pinto, J. P., Swingley, D., Weinbergy, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science*, *9*(3), 228–231.

Fernald, A., Thorpe, K., & Marchman, V. A. (2010). Blue car, red car: Developing efficiency in online interpretation of adjective–noun phrases. *Cognitive psychology*, *60*(3), 190–217.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive psychology*, *47*(2), 164–203.

Fiebach, C. J., Schlesewsky, M., & Friederici, A. D. (2001). Syntactic working memory and the establishment of filler-gap dependencies: Insights from ERPs and fMRI. *Journal of psycholinguistic research*, *30*(3), 321–338.

Field, A. (2009). *Discovering statistics using SPSS*. London: SAGE Publications Ltd.

Fisher, C., & Gleitman, L. R. (2002). Language acquisition. *Stevens' Handbook of Experimental Psychology*.

Gertner, Y., Fisher, C., & Eisengart, J. (2006). Learning words and rules: Abstract knowledge of word order in early sentence comprehension. *Psychological science*, *17*(8), 684–691.

Gibson, J. J. (1977). The theory of affordances. In R. Shaw, J. Bransford, & J. Bransford (Eds.), *Perceiving acting and knowing* (pp. 67–82). London: Routldge.

Goldberg, J. H., & Wichansky, A. M. (2003). Eye tracking in usability evaluation: A practitioner's guide. In *The mind's eye* (pp. 493–516). Amsterdam: Elsevier.

Grice, M., Ritter, S., Niemann, H., & Roettger, T. B. (2017). Integrating the discreteness and continuity of intonational categories. *Journal of Phonetics*, *64*, 90–107.

Grodner, D. J., Klein, N. M., Carbary, K. M., & Tanenhaus, M. K. (2010). "Some," and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition*, *116*(1), 42–55.

Grünloh, T., Lieven, E., & Tomasello, M. (2011). German children use prosody to identify participant roles in transitive sentences. *Cognitive Linguistics*, 22(2), 393-419.

Haider, H. (2010). *The syntax of German*. Cambridge, England: Cambridge University Press.

Hanssen, J., Peters, J., & Gussenhoven, C. (2008). Prodosic Effect of Focus in Dutch Declaratives. In P. A. Barbosa, S. Madureira, & C. Reis (Eds.), *Proceedings of the speech prosody 2008 conference* (pp. 609–612). Campinas, Brazil.

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

Hemforth, B., & Konieczny, L. (2013). *German sentence processing* (Vol. 24). Dordrecht, Netherlands: Springer Science & Business Media.

Huang, Y. T., & Snedeker, J. (2009a). Online interpretation of scalar quantifiers: Insight into the semantics–pragmatics interface. *Cognitive psychology*, *58*(3), 376–415.

Huang, Y. T., & Snedeker, J. (2009b). Semantic meaning and pragmatic interpretation in 5-year-olds: Evidence from real-time spoken language comprehension. *Developmental psychology*, *45*(6), 1723.

Huettig, F., & Altmann, G. T. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition*, *96*(1), B23–B32.

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta psychologica*, *137*(2), 151–171.

Ito, K., Jincho, N., Minai, U., Yamane, N., & Mazuka, R. (2012). Intonation facilitates contrast resolution: Evidence from Japanese adults and 6-year olds. *Journal of Memory and Language*, *66*(1), 265–284.

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

Jaeger, T. (2011). Random slopes in LME. *ling-r-lang-L–A mailing list for language researchers who use the R statistical programming language for modeling and data analysis. Available: https://mailman. ucsd. edu/pipermail/ling-r-lang-l/2011-February/000225. html. Accessed 2018 March 13th.*

Johnson, E. K., & Huettig, F. (2011). Eye movements during language-mediated visual search reveal a strong link between overt visual attention and lexical processing in 36-month-olds. *Psychological Research*, *75*(1), 35–42.

Johnson, E. K., McQueen, J. M., & Huettig, F. (2011). Toddlers' language-mediated visual search: They need not have the words for it. *The Quarterly Journal of Experimental Psychology*, *64*(9), 1672–1682.

Jones, H., & Jones, M. (2019). *The Oxford Guide to Middle High German.* Oxford: Oxford University Press.

Kamide, Y., Altmann, G. T., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*(1), 133–156.

Kamide, Y., Scheepers, C., & Altmann, G. T. (2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from german and english. *Journal of Psycholinguistic Research*, *32*(1), 37–55.

Kany, W., & Schöler, H. (2007). *Fokus: Sprachdiagnostik: Leitfaden zur Sprachstandardbestimmung im Kindergarten*. Berlin: Cornelsen.

Keppel, G., & Zedeck, S. (1982). Correction for multiple comparisons. *Design and analysis: a researcher's handbook*, 144–168.

Kidd, E., & Bavin, E. L. (2007). Lexical and referential influences on on-line spoken language comprehension: A comparison of adults and primary-school-age children. *First Language*, *27*(1), 29–52.

Kjelgaard, M. M., & Speer, S. R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language*, *40*(2), 153–194.

Knoeferle, P. (2007). Comparing the time course of processing initially ambiguous and unambiguous German SVO/OVS sentences in depicted events. In R. P. Van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 517–533). Amsterdam: Elsevier.

Knoeferle, P. (2015). Visually situated language comprehension in children and in adults. In R. K. Mishra, N. Srinivasan, & F. Huettig (Eds.), *Attention and vision in language processing* (pp. 57–75). New Delhi: Springer.

Knoeferle, P., & Crocker, M. W. (2006). The coordinated interplay of scene, utterance, and world knowledge: Evidence from eye tracking. *Cognitive Science*, *30*(3), 481–529.

Knoeferle, P., & Crocker, M. W. (2007). The influence of recent scene events on spoken comprehension: Evidence from eye movements. *Journal of Memory and Language*, *57*(4), 519–543.

Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2003). Actions and roles: Using depicted events for disambiguation and reinterpretation in German and English. In *Proceedings of the 25th Annual Meeting of the Cognitive Science Society* (pp. 681–686).

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*(1), 95–127.

Knoeferle, P., Habets, B., Crocker, M. W., & Münte, T. F. (2007). Visual scenes trigger immediate syntactic reanalysis: evidence from ERPs during situated spoken comprehension. *Cerebral Cortex*, *18*(4), 789–795.

Knoeferle, P., Urbach, T. P., & Kutas, M. (2011). Comprehending how visual context influences incremental sentence processing: Insights from ERPs and picture-sentence verification. *Psychophysiology*, *48*(4), 495–506.

Knoeferle, P., Urbach, T. P., & Kutas, M. (2014). Different mechanisms for role relations versus verb–action congruence effects: Evidence from ERPs in picture–sentence verification. *Acta psychologica*, *152*, 133–148.

Kreysa, H., Knoeferle, P., & Nunneman, E. M. (2014). Effects of speaker gaze versus depicted actions on visual attention during sentence comprehension. In *Proceedings of the 36th Annual Meeting of the Cognitive Science Society* (pp. 2513–2518).

Kröger, J. M., Münster, K., Burigo, M., & Knoeferle, P. (2018). Wiggle, Wiggle, Wiggle: How Visual Cues Influence Thematic Role Assignment in Children and Adults. In *Proceedings of the 40th Annual Meeting of the Cognitive Science Society* (pp. 1969–1974).

Kröger, J. M., Münster, K., & Knoeferle, P. (2017). The Influence of Prosody and Case Marking on Thematic Role Assignment in Ambiguous Action Scenes: Adults versus Children. In *Proceedings of the 39th Annual Meeting of the Cognitive Science Society* (pp. 2463–2468).

Kügler, F. (2008). The role of duration as a phonetic correlate of focus. In P. A. Barbosa, S. Madureira, & C. Reis (Eds.), *Proceedings of the speech prosody 2008 conference* (pp. 591–594). Campinas, Brazil.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205.

Lambrecht, K. (1996). *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents* (Vol. 71). Cambridge: Cambridge University Press.

Lew-Williams, C., & Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, *18*(3), 193–198.

Lieven, E. V., Pine, J. M., & Baldwin, G. (1997). Lexically-based learning and early grammatical development. *Journal of child language*, *24*(1), 187–219.

Luck, S. J. (2014). *An introduction to the event-related potential technique* (2nd ed.). Cambridge, MA: MIT press.

MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of verbal learning and verbal behavior*, *23*(2), 127–150.

MacWhinney, B., Pléh, C., & Bates, E. (1985). The development of sentence interpretation in Hungarian. *Cognitive psychology*, *17*(2), 178–209.

Maiworm, B. (2008). *Grammatische Fähigkeiten sprachnormaler Grundschulanfänger. empirische Überprüfung mit der ESGRAF-Diagnostik* (Unveröffentlichte Diplomarbeit). Universiät zu Köln.

Mani, N., & Huettig, F. (2012). Prediction during language processing is a piece of cake—but only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(4), 843–847.

Matzke, M., Mai, H., Nager, W., Rüsseler, J., & Münte, T. (2002). The costs of freedom: an ERP–study of non-canonical sentences. *Clinical Neurophysiology*, *113*(6), 844–852.

Melchers, P., & Preuß, U. (1994). *Kaufman assessment battery for children: K-ABC; deutschsprachige fassung*. Leiden: Swets und Zeitlinger.

Meroni, L., & Crain, S. (2003). How children avoid kindergarten paths. In *Proceedings of 4th Tokyo Conference on Psycholinguistics* (pp. 159–184).

Messenger, K., Branigan, H. P., McLean, J. F., & Sorace, A. (2012). Is young children's passive syntax semantically constrained? Evidence from syntactic priming. *Journal of Memory and Language*, *66*(4), 568–587.

Milles, A. E. (1985). The acquisition of German. In D. I. Slobin (Ed.), *The cross-linguistic study of language acquisition* (Vol. 1, pp. 141–254). Hillsdale, New Yersey: Lawrence Erlbaum Associates.

Mishra, R. K., Olivers, C. N., & Huettig, F. (2013). Spoken language and the decision to move the eyes: To what extent are language-mediated eye movements automatic? In *Progress in brain research* (Vol. 202, pp. 135–149). Amsterdam: Elsevier.

Mitchell, D. C., & Cuetos, F. (1991). The origins of parsing strategies. *Current issues in natural language processing*, 1–12.

Motsch, H.-J., & Becker, L.-M. (2014). Grammatikerwerb deutschsprachiger Kinder zwischen 4 und 9 Jahren (GED 4–9). *Vierteljahresschrift für Heilpädagogik (VHN)*, *1*, 71–73.

Motsch, H.-J., & Rietz, C. (2016). *ESGRAF 4-8 Grammatiktest für 4-bis 8-jährige Kinder*. München: Ernst Reinhardt.

Münster, K. (2016). *Effects of emotional facial expressions and depicted actions on situated language processing across the lifespan.* (PhD Thesis). Bielefeld

University.

Münster, K., Carminati, M. N., & Knoeferle, P. (2015). The Effect of Facial Emotion and Action Depiction on Situated Language Processing. In *Proceedings of the 37th Annual Meeting of the Cognitive Science Society* (pp. 1673–1678).

Münster, K., & Knoeferle, P. (2018). Extending situated language comprehension (accounts) with speaker and comprehender characteristics: towards socially situated interpretation. *Frontiers in psychology*, *8*, 2267.

Nation, K., Marshall, C. M., & Altmann, G. T. (2003). Investigating individual differences in children's real-time sentence comprehension using language-mediated eye movements. *Journal of Experimental Child Psychology*, *86*(4), 314–329.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of memory and language*, *31*(6), 785–806.

Özge, D., Küntay, A., & Snedeker, J. (2019). Why wait for the verb? Turkish speaking children use case markers for incremental language comprehension. *Cognition*, *183*, 152–180.

Özge, D., Münster, K., Knoeferle, P., Küntay, A., & Snedeker, J. (2016). Predictive Use of German Case Markers in German Children. In *Proceedings of the 40th Annual Boston University Conference on Language Development* (pp. 291–303).

O'Grady, W., & Cho, S. W. (2001). First language acquisition. *Contemporary linguistics: An introduction*, 409–448.

Peter, M., Chang, F., Pine, J. M., Blything, R., & Rowland, C. F. (2015). When and how do children develop knowledge of verb argument structure? Evidence from verb bias effects in a structural priming task. *Journal of Memory and Language*, *81*, 1–15.

Popella, M. (2005). *Vergleichsstudie zum Erwerb der Kasusmarkierung von Erstklässlern in Grund- und Sprachheilsschulen* (Diplomarbeit). Universiät zu Köln.

Rayner, K., Sereno, S. C., Morris, R. K., Schmauder, A. R., & Clifton Jr, C. (1989). Eye movements and on-line language comprehension processes. *Language and Cognitive Processes*, *4*(3-4), SI21–SI49.

Richardson, D. C., & Kirkham, N. Z. (2004). Multimodal events and moving locations: eye movements of adults and 6-month-olds reveal dynamic spatial indexing. *Journal of Experimental Psychology: General*, *133*(1), 46.

Richardson, D. C., & Spivey, M. J. (2000). Representation, space and Hollywood Squares: Looking at things that aren't there anymore. *Cognition*, *76*(3), 269–295.

Richardson, D. C., & Spivey, M. J. (2004). Eye tracking: Research areas and applications. *Encyclopedia of biomaterials and biomedical engineering*, 573–582.

Rothweiler, M. (2015). Spracherwerb. In J. Meibauer, U. Demske, J. Geilfuß-Wolfgang, J. Pafel, K. H. Ramers, & M. Rothweiler (Eds.), *Einführung in die germanistische Linguistik* (pp. 255–297). Berlin: Springer.

Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Current directions in psychological science*, *12*(4), 110–114.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928.

Sauermann, A., Höhle, B., Chen, A., & Järvikivi, J. (2011). Intonational marking of focus in different word orders in German children. In M. W. et al. (Ed.), *28th west coast conference on formal linguistics [wccfl xxviii]* (pp. 313–322). Somerville, MA.

Schafer, A. J., Speer, S. R., Warren, P., & White, S. D. (2000). Intonational disambiguation in sentence production and comprehension. *Journal of psycholinguistic research*, *29*(2), 169–182.

Scheepers, C., & Crocker, M. W. (2004). Constituent order priming from reading to listening: A visual-world study. In M. Carreiras & C. Clifton Jr. (Eds.), *The on-line study of sentence comprehension: Eyetracking, ERP, and beyond* (pp. 167–185). New York: Psychology Press.

Schipke, C. S., Friederici, A. D., & Oberecker, R. (2011). Brain responses to case-marking violations in German preschool children. *Neuroreport*, *22*(16), 850–854.

Schulz, P. (2007). Erstspracherwerb Deutsch: Sprachliche Fähigkeiten von eins bis zehn. *Diagnostik am Schulanfang*, *3*, 67–86.

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, *71*(2), 109–147.

Sekerina, I. A., & Trueswell, J. C. (2012). Interactive processing of contrastive expressions by Russian children. *First language*, *32*(1-2), 63–87.

Sinclair, A., Sinclair, H., & De Marcelus, O. (1971). Young children's comprehension and production of passive sentences. *Archives de Psychologie*, *161–164*(41), 1–22.

Skopeteas, S., & Fanselow, G. (2009). Effects of givenness and constraints on free word order. *Information structure: Theoretical, typological, and experimental perspectives*, 307–331.

Snedeker, J., & Trueswell, J. C. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and language*, *48*(1), 103–130.

Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive psychology*, *49*(3), 238–299.

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*(4), 447–481.

Spivey-Knowlton, M. J., & Sedivy, J. C. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition*, *55*(3), 227–267.

Staudte, M., Crocker, M. W., Heloir, A., & Kipp, M. (2014). The influence of speaker gaze on listener comprehension: Contrasting visual versus intentional accounts. *Cognition*, *133*(1), 317–328.

Stites, M. C., & Federmeier, K. D. (2015). Subsequent to suppression: Downstream comprehension consequences of noun/verb ambiguity in natural reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(5), 1497–1515.

Sturt, P., Keller, F., & Dubey, A. (2010). Syntactic priming in comprehension: Parallelism effects with and without coordination. *Journal of Memory and Language*, *62*(4), 333–351.

Szagun, G. (2004). Learning by ear: on the acquisition of case and gender marking by German-speaking children with normal hearing and with cochlear implants. *Journal of child language*, *31*(1), 1–30.

Tabachnick, B. G., & Fidell, L. S. (2007). *Experimental designs using ANOVA*. Belmont, CA: Thomson/Brooks/Cole.

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

Tomasello, M. (1992). *First verbs: A case study of early grammatical development*. Cambridge, England: Cambridge University Press.

Tracy, R. (1986). The acquisition of case morphology in German. *Linguistics*, *24*(1), 47–78.

Truckenbrodt, H. (1995). *Phonological phrases–their relation to syntax, focus, and prominance* (PhD Thesis). Massachusetts Institute of Technology.

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, *73*(2), 89–134.

Ulrich, T., Penke, M., Berg, M., Lüdtke, U., & Motsch, H. (2016). Der dativerwerb. Forschungsergebnisse und ihre therapeutischen Konsequenzen. *Logos*, *24*(3), 176–190.

Weber, A., Grice, M., & Crocker, M. W. (2006). The role of prosody in the interpretation of structural ambiguities: A study of anticipatory eye movements. *Cognition*, *99*(2), B63–B72.

Weighall, A. R. (2008). The kindergarten path effect revisited: Children's use of context in processing structural ambiguities. *Journal of Experimental Child Psychology*, *99*(2), 75–95.

Weighall, A. R., & Altmann, G. T. (2011). The role of working memory and contextual constraints in children's processing of relative clauses. *Journal of child language*, *38*(3), 579–605.

Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J., & MacDonald, M. C. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive psychology*, *58*(2), 250–271.

Wexler, K. (1999). Innateness of language. *The MIT encyclopedia of the cognitive sciences*, 408–409.

Zhang, L., & Knoeferle, P. (2012). Visual context effects on thematic role assignment in children versus adults: Evidence from eye tracking in German. In *Proceedings of the 38th Annual Meeting of the Cognitive Science Society* (pp. 2593–2598).

CONSENT FORM AUDIO FILES

I, Katharina Hagenfeld, herby give consent for Julia Marina Kröger to use the audio files I recorded for:

- Data collection and data analyses
- Presentation at conferences
- Publication of results
- Follow-up studies

(Katharina Hagenfeld)

Julia Marina Kröger

Curriculum Vitæ

Personal Data

Date of birth	23.02.1988
Place of birth	Paderborn
Nationality	German

Experience

2018 – now	**Commercial executive**, *Paderborn*, Stollwerk Beratungen.
2014 – 2017	**Scientific staff member**, *Bielefeld University*, Faculty for Linguistics and Literature.
Oct – Dec 2014	**Scientific staff member**, *Paderborn University*, Faculty for Cultural Studies.
2012 – 2014	**Student assistant**, *Paderborn University*, Faculty for Cultural Studies, Dr. Anke Lenzing, Prof. Dr. Manfred Pienemann.

Teaching

2016 – 2017	**Eye Tracking in Psycholinguistics: A Practical Introduction**, *Bielefeld University*, together with Thomas Kluth and Eva M. Nunnemann.
2014 – 2015	**Bilingualism in Education**, *Paderborn University*.

Education

2014 – 2019	**Ph.D. student**, *Bielefeld University*, Faculty for Linguistics and Literature, Supervisor: Prof. Dr. Pia Knoeferle. Thesis Title: 'Real Time Thematic Role Assignment in Children and Adults: The Influence of Prosody, Case Marking and Visual Cues'
2013 – 2014	**M.A., Linguistics**, *Paderborn University*, Grade: 1.4. Thesis Title: 'Has, plays, makes: are frequently occurring inflected verbs stored as chunks?'
2010 – 2013	**B.A., English and American Literature and Culture and English Linguistics**, *Paderborn University*, Grade: 2.3. Thesis Title: '[mæs or mes] - Learner difficulties in the pronunciation of English'
2008 – 2010	**Teacher Training, English and Mathematics**, *Paderborn University*.
2007 – 2008	**Teacher Training, Spanish and Philosophy**, *Paderborn University*.

Ferrariweg 2 – 33102 Paderborn
0174 / 51 53 699 • *jkroeger@cit-ec.uni-bielefeld.de*

2007 **A-Levels (Abitur)**, *Gesamtschule Elsen*, Paderborn, Grade: 3,1.

Internships

2013 **Insitute for Buisness Linguistics**, *Paderborn*, University internship.

2013 **DMRC Paderborn University**, *Paderborn*, University internship.

2012 **Copyservice Paderborn University**, *Paderborn*, Economic internship.

2009 **Lise-Meitner Realschule**, *Paderborn*, Teacher trainee internship (English).

2008 **Gesamtschule Elsen**, *Paderborn*, Teacher training: orientation.

2005 **A. Regan and Colleagues Veteranry surgery**, *Bolton: UK*, Internship abroad.

2003 **Disctrict Court Paderborn**, *Paderborn*, School Internship.

Extra Curricular Activities

2019 **Reviewer**, Abstract review for the Embodied and Situated Language Processing and The Attentive Listener in the Visual World.

2017 **Reviewer**, Abstract review for the 39th Annual Meeting of the Cognitive Sciene Society.

2016 **Organising Commitee**, *Berlin*, Workshop on the Role of Pragmatic Factors in Child Language Processing.

2014 **Organising Commitee**, *Paderborn*, 14th International Symposium on Processability Approaches to Language Acquisition PALA.

2013 **Diagnostic Language Tests**, *Wriezen and Bad Saarow*, The influence of lexical retrieval on grammatical processing in connection with Processability Theory using rapid profile (in cooperation with Katharina Hagenfeld and Birgit Goehrmann).

Languages

German, *native*.

English, *professional*.

Spanish, *intermediate*.

Portuguese, *beginner*.

Swedish, *beginner*.

Russian, *beginner*.

German Sign Language, *beginner*.

Ferrariweg 2 – 33102 Paderborn
0174 / 51 53 699 • *jkroeger@cit-ec.uni-bielefeld.de*