Leigh Breakell Fernandez

Investigating Word Order Processing using Pupillometry and Event-Related Potentials

This work was supported by the Erasmus Mundus Joint Doctoral Programme (EMJD) of the European Union joint International Doctorate for Experimental Approaches to Language and Brain (IDEALAB) of the Universities of Groningen (NL), Newcastle (UK), Potsdam (DE), Trento (IT), and Macquarie University, Sydney (AU).

Supervisors:
Prof. B. Höhle
Prof. L. Nickels
Dr. J. Brock

Contents

List of Figures

List of Tables

Chapter 1

Literature Review

This thesis focuses on the processing costs associated with understanding spoken sentences with different word orders and grammatical constructions. These types of constructions are investigated using a psychophysiological (i.e. the physiological measures that underlie psychological processing) measure known as pupillometry, and the last study employs electroencephalography (EEG). Although the unifying goal of the thesis is to examine processing costs using psychophysical measures, the scope of the thesis is quite broad, investigating the processing of several different syntactic constructions and across different populations - children, native speakers of German, native speakers of English, and highly proficient second language speakers (L2) of English. In this chapter we provide a review of the literature on the primary methodology in this thesis, pupillometry which focuses on what the pupil can reveal about language processing. This is followed by a brief overview of the remainder of the thesis.

1.1 Pupillometry

Thinking is biological work and as such, thinking consumes resources; these resources are from a limited pool and this consumption is measurable. Behavioral measures (e.g. response time, etc.) are not enough to characterize this type of resource consumption given that brain activity does not directly map onto these types of measures; rather, physiological methods are more appropriate tools to measure cognitive workload given that they can map onto brain activity (Just, Carpenter, & Miyake, 2003). One such method involves measuring the small opening in the center of the eye (the pupil). The diameter of the pupil is inherently variable and ranges from approximately one to eight millimeters. Pupil change is controlled by a set of two opposing muscles in the iris, and the change in the diameter of the pupil primarily occurs as a result of changes in light levels and as a result of accommodation (when the two eyes work together to focus). However, smaller changes in the pupil also occur and are believed to reflect cognitive processing and mental activity (Andreassi, 2007).

These small movements of the pupil (typically less than .05mm) are the basis of cognitive pupillometry, which is the study of pupil movements as a measure of cognitive processing (Beatty and Lucero-Wagoner, 2000). Generally there are two types of pupillary movement: phasic, which have a short latency and are believed to reflect cognitive processing, and tonic which have a much longer latency and are believed to reflect arousal and can show changes in fatigue and alertness (Beatty, 1982). Pupil diameter change is easily measured using an eye tracker, and in this thesis infrared eye trackers are used. Typically, these types of eye trackers reflect infrared light into the retina and the number of pixels that fill the retina are recorded by the eye tracker (which takes into account the distance between the eye and the camera) and a measurement of pupil diameter is calculated (Klinger, 2010). Pupillometry is an interesting and relatively novel method that provides insight into cognitive resource consumption. Additionally, pupillometry provides a non-invasive and relatively cheap online measure which is also appropriate for use with vulnerable populations. More research using pupillometry is needed to further validate the measure, and for these reasons it is the method of choice throughout most of this thesis.

While pupillometry research can be seen as far back as the 1800s (see Beatty & Lucero-Wagoner, 2000), it was not until the 1960s that pupillometry became a more common tool for psychophysiologists. Below some of the classical and seminal pupillometry research is outlined. The research is then evaluated against a set of criteria, set by Daniel Kahneman, which enables pupillometry to be established as an effective tool for measuring cognitive processing.

1.2 Classic Pupillometry Studies

Some of the earliest studies using pupillometry investigated the relationship between the affect of the stimulus and pupil size. For example, Hess and Polt (1960) examined pupil response in men and women when viewing a series of pictures: a baby, mother and baby, nude male, nude female, and a landscape. They found that women showed a larger pupil response to the first four pictures listed compared to male participants, while males showed a larger pupil response to the latter two pictures. Similarly, Janisse (1973) reviewed research investigating pupil size and affect, and reported that the most consistent finding was that the pupil size was related to affect of a stimulus, and the greater the affect the larger the pupil size (whether positive or negative).

However, of more relevance to the research reported in this thesis are those studies that started to relate cognitive load with pupil size. For example, Hess and Polt (1964) investigated pupillary changes while participants were solving math problems. The authors found a correlation between problem difficulty and pupil size, with the more difficult problems evoking a larger pupil size. They also found that the pupil reached maximum diameter just before the participants provided their answer and then decreased to a previously measured control size. These findings suggested that changes in pupil size provided a measure of the effort allocated during mental activity.

Similarly, Kahneman and Beatty (1966) investigated pupil changes in relation to short-term memory. They measured pupil diameter changes while participants retained strings of digits (between 3 or 7 digits), strings of monosyllabic words, or performed transformation of digits. In terms of digit retention and recall, pupil diameter increased with every additional digit held in memory and decreased with every digit recalled. In addition, recall of digits (the easiest and most accurate task) evoked the smallest pupil diameter, followed by the word recall, and the largest pupil diameter was seen in the transformation condition. Interestingly, participants showed an anticipatory effect: when they were told that they would be a difficult trial coming up they showed a slight dilation of the pupil. Kahneman and Beatty argued that these findings provided evidence that pupillometry was an effective way to measure processing load or processing effort.

The examples above focused on active allocation of attention, however, in attempt to investigate pupillary change and nonverbal attention, Kahneman and Beatty (1967) investigated pitch discrimination. Participants judged whether a comparison tone was higher or lower in pitch than a previously heard tone; the tones varied in pitch, making some easier to discriminate and some more difficult to discriminate. The authors found that the pupil diameter increased with the difficulty of the discrimination, thus leading them to argue that pupil change was an effective measure of processing load in both active tasks (e.g. digit recall) and nonverbal tasks (e.g. pitch discrimination).

In another study investigating a nonverbal task, Beatty (1982) examined auditory vigilance and looked at both tonic and phasic pupillary changes; as stated above, phasic (short latency) pupillary changes are believed to reflect cognitive processing, while tonic (long latency) pupillary changes are believed to reflect arousal. Participants monitored auditory tones for a target tone over 3 trials taking 48 minutes. Beatty found that phasic pupillary amplitude decreased over time as did task accuracy, while tonic pupillary changes did not differ throughout the task; these data suggested that accuracy and phasic pupillary amplitude indicated effort during the task. Beatty also pointed out the similarities of the pupil changes to those seen with ERP components in previous vigilance tasks suggesting that underlying brain activation and pupillometry may be related.

Pupillometry has also been shown to be sensitive to individual differences; Ahern and Beatty (1979; 1981) divided participants into low and high intelligence (based on scores from a standardized test) and measured pupil diameter change during four tasks (multiplication, digit span, vocabulary, and sentence comprehension). The more difficult stimuli of each task evoked larger pupil sizes for both groups, and in terms of individual differences, they found that in 3 of the 4 tasks the low intelligence group showed larger pupil size than the high intelligence group. These findings suggest that the pupil is sensitive not only to most of the cognitive tests used but also to individual differences during these tests.

1.3 Kahneman's Criteria

Kahneman (1973) put forward 3 criteria that he suggested must be met in order for a physiological measure to be considered a valid measure of cognitive processing: 1) the measure should vary within a task as a result of difficulty; 2) the measure should be sensitive to tasks that employ different cognitive operations; and 3) the measure should be sensitive to individual differences. The work of Kahneman and Beatty (1966; 1967) and Hess and Polt (1964), described above, provides evidence that more difficult trials within a task evoke larger pupil changes, thus fulfilling Kahneman's within task variation criterion. Beatty (1982) showed that there is a continuum of evoked pupil size based on task, with more difficult tasks evoking larger pupil size compared to easier tasks (multiplication tasks evoked the largest pupil size, followed by memory, language, and perception tasks); these results fulfill the second criterion proposed by Kahneman. The sensitivity of the pupil to individual differences, the third of Kahneman's criteria, was seen in the work by Ahern and Beatty (1979, 1981) who showed a relationship between intelligence and pupil size.

The fulfillment of these criteria by pupillometry research provided a foundation for the use of pupillometry as a physiological measure of processing costs and set the stage for research utilizing a pupillometry paradigm. The focus of this thesis is language processing, consequently in the next section I will outline some of the research using pupillometry to investigate aspects of language processing.

1.4 Pupillometry and Language

The empirical research in this thesis focuses on sentence processing, particularly with respect to word order and ambiguity in a variety of groups (i.e. adult native speakers of English, adult native speakers of German, child native speakers of German, and adult proficient second language speakers (L2) of English). Hence, in this section I will start by reviewing research related to the empirical studies in this thesis. The review will start with research on sentence processing, focusing on ambiguity, complexity, and word order. This will be followed by a brief discussion of second language research and finally the use of pupillometry with children. This section will establish that pupillometry is an effective tool to study cognitive processing costs in a variety of linguistic tasks pertinent to the subsequent chapters, thus setting the stage for the research in this thesis that employs pupillometry.

1.4.1 Sentence Processing

1.4.1.1 Ambiguity

Pupillometry has been shown to be sensitivity to grammatical ambiguity, for example Schluroff et al. (1986) used a reading and pupillometry paradigm to investigate the effects of syntactic ambiguity when German speakers read (German) sentences, see Example (1):

1. Peter verfolgte den Mann mit dem Motorrad.
   Peter chased the man with the motorbike

This sentence can be interpreted in two ways, the first is the verb-oriented reading in which Peter used the motorbike to chase the man, or it can be a syntactically less complex[1], object-oriented reading in which Peter chased the man who was on the motorbike. The syntactically more complex verb-oriented reading evoked a larger pupil dilation than the less complex object-oriented reading both during the reading of the sentence and after processing the sentence.

Lexical ambiguities have also been investigated using a pupillometry paradigm; Ben-Nun (1986) compared pupil size between constructions with and without ambiguous words (*sentence* in the below examples), and also compared constructions that were biased towards one of the ambiguous words meaning either early or later in the construction (see Examples 2-5).

2. We met the famous attorney immediately after the *sentence*

[1] Syntactic complexity measures were based on The Yngve depth scale which indexes the average complexity of a sentence

3. Immediately after the *sentence* we met the famous attorney

4. We studied theoretical linguistics for purpose of analyzing the *sentence*

5. The long *sentence* was analyzed by a generative linguist

In (2) and (3), *sentence*, is referring to the punishment associated with breaking laws deemed by court while in (4) and (5) sentence is referring to a set of words typically conveying meaning. In (2) & (4), the initial context biases towards the appropriate interpretation of the ambiguous word occurs, while in (3) & (5) there is no bias and the word remains ambiguous until the end of construction. Ben-Nun found that ambiguous words (e.g. *sentence*) evoke larger pupil size than unambiguous words, suggesting that more processing resources are being consumed which, the author argued, may be a result of the retrieval of the two meanings. In terms of bias, both early and late bias sentence types evoked similar pupil size during sentence presentation, but post-sententially the late bias constructions evoked a larger pupil size compared to the early bias sentences. Ben-Nun argued that this post-sentential difference in the ambiguous condition indicated a choice decision, in which the participant has to decide between the two potential meanings.

In sum, pupillometry research into ambiguous constructions shows that the pupil is sensitive to ambiguity resolution during sentence processing. This provides a foundation for the use of pupillometry in Chapter 3 and 4 as means to investigate the comprehension of ambiguous and unambiguous constructions in German.

1.4.2 Syntax and Grammatical Complexity

Pupil dilation changes have also been shown to be a strong indicator of grammatical complexity, Schluroff (1982) found the pupil to be a better indicator of complexity than offline behavioral measures. Schluroff found that when participants listened to a more complex (based on The Yngve depth scale) aurally presented construction a larger pupil size was evoked compared to less complex sentences, and that this relationship was approximately linear. In contrast, the offline ratings of comprehensibility by the participants did not correlate with grammatical complexity (only with sentence length). Schluroff therefore argued that pupil diameter was a better indicator of grammatical complexity than offline ratings. Interestingly this study was conducted in English with native German speakers, and while the study was not investigating second language acquisition, it is one of the few studies using pupillometry with second language speakers (second language acquisition will be discussed below).

Just and Carpenter (1993) were interested in the cognitive resources consumed during the processing of object relative sentence constructions (Example 6) and subject relative sentence constructions (Example 7).

6. Object Relative: The reporter that the senator attacked admitted the error.

7. Subject Relative: The reporter that attacked the senator admitted the error.

Object relative constructions like those seen in (6) have been found to cause more processing difficulties compared to subject relative constructions like (7) (e.g. Ford, 1983; Just et al., 1996; King & Just 1991; Wanner & Mastros, 1978). This is for several reasons. First, object relative constructions have non-canonical word order - rather than the subject (*the senator*) appearing first in the sentence as is usual in English, the object of the relative clause appears first (*the reporter*). They also have atypical thematic roles, *the reporter* in Example (6) is both the patient of the relative clause verb (*attacked*) and the agent of the main clause verb (*admitted*). While in the subject relative clause, *the reporter*, is the agent of both verbs in the construction. Hence, object relative constructions are more difficult to process than the canonical subject relative constructions (7) (this will be discussed in more detail in Chapter 5).

Just and Carpenter (1993) investigated the processing of written object and subject relatives and found that object relative constructions evoked larger pupil responses. Peak pupil size occurred approximately 1.2 seconds after the onset of the critical verb in the main clause (*admitted*): the point that requires the most processing recourses in both conditions. Participants were also less accurate in answering comprehension questions about the object relative constructions. These results supported the hypothesis that object relatives consume more processing resources and are more difficult to comprehend than canonical subject relative constructions.

Another pupillometry study investigated the effects of age on processing of object and subject relative constructions; Piquado, Isaacowitz, and Wingfield (2010) investigated pupil change in older and younger adults during sentence retention and recall. As in Just and Carpenter (1993), participants heard subject relative (8) & (9) and object (10) & (11) constructions. However, Piquado et al. increased memory demands by adding additional modifiers (e.g. professional gambler, suspicious dealer and perfect card, rather than gambler, dealer and card; 9 & 11).

8. Subject relative: The gambler that signaled the dealer revealed the card.
9. Subject relative with modifiers: The professional gambler that signaled the suspicious dealer revealed the perfect card.
10. Object relative: The gambler that the dealer signaled revealed the card
11. Object relative with modifiers: The professional gambler that the suspicious dealer signaled revealed the perfect card.

Piquado et al. found that younger adults showed larger pupil sizes for the more complex object relative constructions compared to the less complex subject relative constructions. There was also an additive effect on the pupil size for sentences with modifiers. The older adult group did not display the same pupil pattern in terms of syntactic complexity, rather they only showed a larger pupil size with the longer sentences with modifiers. The authors postulated that older adults may not require the same resources for processing complex syntactic structures (compared to the younger group) given their experience with language.

Recently Demberg (2013) investigated a new way of analyzing pupillometry, known as the Index of Cognitive Activity. This is a micro-level analysis that counts the frequency of small rapid pupil dilations, these dilations are believed to be less susceptible to lighting and eye movements (as opposed to looking at overall pupil dilation at a macro-level). While performing a driving task, participants heard ambiguous subject relative constructions (Example 12) and object relative constructions (Example 13) in German, and were asked to answer questions about what they heard (similar linguistic constructions will be further explored in Chapters 3 & 4).

12. Ambiguous subject relative: Die Nachbarin, die einige der Mieter auf Schadensersatz verklagt hat, traf sich gestern mit Angelika.
    The neighbor, who sued some of the tenants for damages, met Angelika yesterday.

13. Ambiguous subject relative: Die Nachbarin, die einige der Mieter auf Schadensersatz verklagt haben, traf sich gestern mit Angelika.
    The neighbor, whom some of the tenants sued for damages, met Angelika yesterday.

Demberg found that the Index of Cognitive Activity (the frequency of the microdilations) increased during the ambiguity and decreased following the disambiguation. It was also higher in the object relative constructions compared to the subject relative clause constructions. Demberg also compared the overall (macro-level) pupil dilation to the Index of Cognitive Activity and found that both showed that the pupil contracted more quickly during comprehension of the less complex relative clause conditions than the object relative constructions. While the Index of Cognitive Activity seems like a promising method for analyzing pupil data, it appears that overall pupil dilation provides equivalent information on the processing of linguistic data (particularly with subject and object relative clauses), consequently, in this thesis, the more established measure, pupil change, is used.

Just and Carpenter (1993) were also interested in the processing involved with filler gap dependencies. These types of constructions assume that a word (i.e. filler) has moved from its original position ($_i$) and left a silent trace (t_i) of itself in that original position, and the parser must associate the moved element with the original site to assign grammatical roles to the filler (this will be discussed in depth in Chapter 2). They compared pupil change in constructions that contained filler gap dependencies with constructions that contained no filler gap dependencies to test processing resource allocation, examples can be seen below in (14-17)

14. Filler gap dependency (plausible): The confused police did not know which leader$_i$ the rioters followed t_i noisily down the street after the meeting.

15. No filler gap dependency (plausible): The confused police did not know whether the rioters followed the leader noisily down the street after the meeting.

16. Filler gap dependency (implausible): The confused police did not know which blanket$_i$ the rioters followed t_i noisily down the street after the meeting.

17. No filler gap dependency (implausible): The confused police did not know whether the rioters followed the blanket noisily down the street after the meeting.

Pupillary responses were larger with filler gap dependency constructions compared to those with no such dependency and, additionally, there was greater pupil change for the implausible conditions compared to the plausible. Just and Carpenter argued that these results provided evidence that there is a processing cost associated with constructions that involve a filler gap dependencies and with constructions that have implausibility, and that the pupil is sensitive to these constructions. In a similar study using pupillometry to investigate filler gap dependencies, Fernandez (2013) also found that the pupil was sensitive to filler gap dependencies and that it was sensitive to the linguistic constraints that dictate filler gap dependencies.

This section has briefly reviewed pupillometry research which has investigated the processing of filler gap dependencies with displaced elements, paying particular attention to research that looked at subject and object relative clauses. This research has shown that the pupil is sensitive to the difficulty that is associated with object-initial conditions compared to subject-initial conditions. Chapter 2 of this thesis, follows up this research, focusing on the difficulty associated with displaced elements in filler gap dependencies using pupillometry with native speakers of English and proficient second language speakers of English. Pupillometry is employed to investigate these constructions hoping to garner new insight into processing that may not be available using other methodologies.

1.4.3 Second Language Speakers (L2)

As mentioned previously, there have been a limited number of studies using pupillometry with second language speakers. While the study by Schluroff (1982), described above, was not directly looking at second language acquisition, nevertheless, it appears to be the first study that used pupillometry with second language speakers.

Hyönä, Tommola, and Alaja (1995) used pupillometry to examine second language processing more directly. They investigated the processing costs involved in listening and simultaneous interpretation between two languages (by native Finnish speakers trained in simultaneous English interpretation). The largest pupil sizes were found during a simultaneous interpretation task, pupil size decreased when participants were required to simply repeat what they heard, and the pupil was the smallest during passive listening. Additionally, when looking at the word level, they found larger dilations for English words compared to Finnish words (the native language of the participants), and larger pupil dilation was associated with the more difficult to translate words. Taken together, these data suggest that pupillometry is sensitive to the processing associated with translation between languages at both the global and word level.

In a more recent study investigating second language speakers and pupillometry, Schmidtke (2014) investigated word retrieval and picture matching in a visual world paradigm task (participants would hear a word while viewing 4 pictures, and would choose the picture that matched they word they heard) in monolingual speakers of English, early Spanish/English bilinguals (English learned before the age of 8), and late Spanish/English bilinguals (English learned after the age of 18). He tested whether pupillometry was sensitive to language experience, lexical frequency, and neighborhood density in English. He found larger pupil diameter for low frequency words, and also for words that had a high neighborhood density (many other words that share sounds with the target), suggesting that more cognitive effort is needed to retrieve these words. In terms of language experience, the bilingual groups displayed a delayed pupil response and a greater effect of neighborhood density compared to the native speakers (with the bilingual group displaying a larger pupil diameter than the native speakers). Frequency effects were more pronounced in the late bilinguals compared to the other two groups, and those bilingual speakers who showed a higher overall English proficiency displayed earlier and smaller pupillary responses to neighborhood density and frequency. The author argued that bilingual speakers show greater word retrieval effort due to decreased language experience, and that the pupil is a sensitive to this retrieval effort.

In sum, although there is limited research into second language comprehension using pupillometry, the studies outlined above have shown that the pupil is a sensitive indicator of the difficulties that accompany translation and word retrieval in a second language. This is capitalized on in Chapter 2 of this thesis where pupillometry is used to test theories of second language acquisition in a group of native German speakers who are proficient in English.

1.4.4 Child Research

Pupillometry has not only served as an effective way to measure processing in adults with typical language skills, it has also been effectively used to study processing in populations that may have difficulties in providing overt answers such as individuals with language impairment and children; here I will focus on some of the research that has been conducted using pupillometry with typically developing children.

Munsinger and Banks (1974) used pupillometry to investigate whether infants (1 year of age) and children (3 years of age) were able to discriminate different color wavelengths of light. They found that the children displayed similar pupillary responses to adults, and while they found that the infant group had somewhat more variable pupillary response compared to the adults, they argued that infant pupillary response was still a reliable measure. While this research is at the perceptual level it nevertheless highlights the fact that pupil response is measurable, sensitive, and reliable even at a young age.

More recently, Chatham, Frank, and Munakata (2009) used pupillometry and a continuous performance task (a target-nontarget discrimination test) to investigate cognitive control in 8-year-old and 3.5-year-old children. They found that, like adults, the older children were proactive (i.e., showed the ability to prepare for predictable upcoming events for example by inhibiting unrelated information and focusing on task related thoughts) in their cognitive control, while the 3.5-year-old group displayed reactive cognitive control (e.g. showed no evidence of preparing for upcoming events). The patterns for each group were found in reaction times, pupil change, and accuracy; these findings suggest that pupillometry is capable of providing insight into differences in cognitive control across different ages.

Kuipers & Thierry (2013) investigated allocation of attention and sensitivity to visual semantic integration in a study employing both pupillometry and electroencephalography (EEG) in the form of event-related potentials (ERP) with groups of monolingual and bilingual English speaking children (approximately three years of age). The children heard a word and saw a picture; the picture either matched the word or did not, and hence responses to these stimuli measured the child's ability to integrate the word meaning with the picture. They found no differences between the monolingual and bilingual children in terms of the ERP data, and found increased pupil dilation only in the bilingual group when the word and picture did not match. When comparing pupil dilation and ERP, they found a decreased N400 amplitude (the N400 is an ERP that is believed to index the difficulty associated with semantic integration), associated with an increased pupil diameter in the bilingual group, and the opposite pattern in the monolingual group. Kuipers & Thierry suggested that when the monolingual group invested more cognitive resources into the unexpected stimuli (increased pupil size) semantic integration was more difficult (a more negative N400); bilinguals on the other hand, invested more resources into the unexpected stimuli (increased pupil size), which in turn aided more efficient semantic integration (a less negative N400). This lead the authors to argue that bilingual children have an advantage over monolingual children, and that bilingual children are more flexible when it comes to mapping words onto objects.

Pupillometry with infants has also been used with the violation-of-expectation paradigm. This established paradigm has been used with preverbal infants to investigate whether a child understands cause and effect. Essentially this paradigm involves children attending to an event (measured with looking times), which is logically impossible (e.g. a blue train enters a tunnel and a red train comes out of the tunnel), and if the child spends longer looking at the event than they spend looking at a similar scene that contains no impossibilities (e.g. a blue train enters and a blue train comes out of the tunnel) it is assumed that the child has understood that their was a violation of their expectation. Jackson and Sirios (2009) examined pupil changes in addition to looking time measures in the violation-of-expectation paradigm with infants approximately 8 months old. Looking time measures revealed longer fixations when the expectations were violated, and pupil change revealed that a violation of expectation coupled with a novel stimulus evoked the largest change. This suggests pupil change can reveal insights into processing that may not be revealed in other measures, in infants as young as 8 months.

While this is by no means a comprehensive review of studies utilizing pupillometry with child populations, it highlights the usefulness of pupillometry for measuring cognitive processing outside of traditional participant groups (i.e. university students) in a range of tasks. It also demonstrates the sensitivity of pupillometry with very young infants, and shows that pupillometry can provide new insights into tasks over and above those provided by other measures (e.g. looking times). This finding motivated the use of pupillometry with a group of four-year-old children in Chapter 4 of this thesis.

1.5 Conclusions

This chapter has outlined previous research using pupillometry, and has highlighted the usefulness of changes in pupil diameter as a proxy for cognitive processing costs. Pupillometry has been used in a variety of contexts, and has been shown to be an effective way to investigate many aspects of cognition (e.g. Ahern & Beatty, 1979, 1981 (individual differences), 1982 (auditory vigilance); Hess and Polt, 1960 (affect), 1964 (solving mathematical problems); Janisse, 1973 (affect); Kahneman & Beatty, 1966 (memory recall), 1967 (pitch discrimination). The focus of this thesis is language comprehension, and as reported above, previous research has found the pupil to be sensitive to manipulations of complexity, ambiguity, and word order. Pupillometry has not only been used with typical adult populations, but has also been used with second language speakers and children. In the empirical studies in this thesis pupillometry is used to test aspects of syntactic comprehension (primarily with filler gap dependencies) in native speakers, second language speakers, and children as outlined below.

1.6 Thesis Overview

1.6.1 Chapter 2

The first empirical chapter (Chapter 2) focuses on filler gap dependences and how native speakers of English and skilled second language speakers of English process these difficult syntactic constructions. It reports an empirical study that uses pupillometry to test whether second language speakers are able to process a type of filler gap dependency, known as an intermediate gap, in the same way as native speakers, and in doing so tests the Shallow Structure Hypothesis (Clahsen and Felser, 2006).

1.6.2 Chapter 3 & 4

The third and fourth chapters focus on the processing costs associated with different word orders in German sentences, as well as the processing costs that can arise as a result of the inherent ambiguity of some of these word orders. The empirical studies examine the processing of these constructions using pupillometry with adults (Chapter 3) and children (Chapter 4) aiming to test competing theories and to validate pupillometry as an effective method for testing the processing of word order in both adults and children.

1.6.3 Chapter 5

Chapter 5 focuses on subject and object relative clauses and the role that animacy plays in processing these constructions in English; the empirical study uses electroencephalography (EEG) to investigate the processing of these constructions. EEG measures brain activity through electrodes on the scalp, and in Chapter 5 these measurements are averaged together (and the random noise averaged out) giving us event related potentials (ERP), which provide information about brain activity in response to a stimulus (in this case object relative clause constructions). There seems to be a link between brain activity and pupil response (see for example, Aston-Jones & Cohen, 2005), and more research is needed investigating pupil change in relation to brain activity. Chapter 5 aims to test competing processing theories of object relative constructions while manipulating animacy with the aim of a future study that will correlate ERP and pupillometry.

1.7 References

Ahern, S. K., & Beatty, J. (1979). Pupillary responses during information processing vary with Scholastic Aptitude Test scores. *Science, 205*(4412), 1289-1292.

Ahern, S.K., & Beatty, J. (1981). Physiological evidence that demand for processing capacity varies with intelligence. In M.P. Friedman, J.P. Das, & N. O'Conner (Eds.), *Intelligence and learning* (pp. 201-216). New York: Plenum Press.

Andreassi, J.L. (2007). *Psychophysiology: Human Behavior and Physiological Response.* Mahwah, New Jersey : Lawrence Erlbaum Associates, Inc.

Beatty, J. (1982). Phasic not tonic pupillary responses vary with auditory vigilance Performance. *Psychophysiology, 19*, 167-172.

Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. In J.T. Cacioppo, L.G. Tassinary, & G.G. Berntson (Eds.), *Handbook of Pscychophysiology*, (pp. 142-162). USA: Cambridge University Press.

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics, 27*, 3-42.

Chatham, C. H., Frank, M.J., & Munakata, Y. (2009). Pupillometric and behavioral markers of a developmental shift in temporal dynamics of cognitive control. *Proceedings of the National Academy of Sciences of the United States of America, 106*, 5529-5553.

Demberg, V. (2013). Pupillometry: the Index of Cognitive Activity in a dual-task study In M. Knauff, M., Pauen, N., Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (p. 2154-2159). Austin TX: Cognitive Science Society.

Fernandez, L., B. (2013). *Pupillometry and Filler Gap Dependencies.* (Unpublished masters of philosophy dissertation). Northumbria University, Newcastle upon Tyne, England.

Ford, M. (1983). A method for obtaining measures of local parsing complexity throughout sentences. *Journal of Verbal Learning and Verbal Behavior 22*, 203218.

Hess, E. H., & Polt, J. M. (1960). Pupil size as related to interest value of visual stimuli. *Science, 132*, 349-350.

Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science, 143*, 1190-1192.

Hyönä, J., Tommola, J., & Alaja, A-M. (1995). Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. *The Quarterly Journal of Experimental Psychology, 48A*, 598-612.

Janisse, M.P. (1973). Pupil size and affect: a critical review of the literature since 1960. *The Canadian Psychologist, 14(4)*, 311-329.

Just, M.A., & Carpenter, P.A. (1993). The intensity dimension of thought: Pupillometric indices of sentence processing. *Theoretical Issues in Ergenomics Science, 47(2)*, 310-339.

Just, M. A., Carpenter, P. A., Keller, T. A., Eddy, W. F., & Thulborn, K. R. (1996). Brain activation modulated by sentence comprehension. *Science, 274*, 114-116.

Just, M.A., & Carpenter, P.A., & Miyake, A. (2003). Neuroindices of cognitive workload: Neuroimaging, pupillometric, and event-realted studies of brain work. *Canadian Journal of Experimental Psychology, 4*, 56-88.

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science, 154*, 1583-1585.

Kahneman, D., & Beatty, J. (1967). Pupillary response in a pitch discrimination task. *Perception & Psychophysics, 2*, 101-105.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language, 30*, 580-602.

Klingner, J. (2010). *Measure cognitive load during visual tasks by combining pupillometry and eye tracking.* (Doctoral dissertation, Stanford University, 2010).

Munsinger, H. & Banks, M.S. (1974). Pupillometry of visual sensitivity among infants, young children, and adults. *Developmental Psychology, 10*, 677-682.

Piquado, T., Isaacowitz, D., & Wingfield, A. (2010). Pupillometry as a measure of cognitive effort in younger and older adults. *Psychophysiology, 47*, 560-569.

Schluroff, M. (1982). Pupil responses to grammatical complexity of sentences. *Brain and Language, 17*, 133-145.

Schluroff, M., Zimmerman, T.E., Freeman, R.B., Hofmeister, K., Lorscheid, T., & Weber, A. (1986). Pupillary responses to syntactic ambiguity of sentences. *Brain and Language, 27*, 322-344.

Schmidtke, J. (2014). Second language experience modulates word retrieval effort in bilinguals: evidence from pupillometry. *Frontiers in Psychology, 5*, 137.

Wanner, E.,& Maratsos, M. (1978). An ATN approach in comprehension. In: M. Halle, J. Bresnan, & G. Miller (Eds.), *Linguistic Theory and Psychological Reality* (pp. 119-161). Cambridge, MA: MIT Press.

Chapter 2

Intermediate Gap processing & Pupillometry in L2 Speakers of English

Abstract

According to the Shallow Structure Hypothesis (SSH), second language (L2) speakers, unlike native speakers, build shallow syntactic representations during sentence processing. In order to test the SSH, this study investigated the processing of a syntactic movement in both native speakers of English and proficient L2 speakers of English using pupillometry to measure processing cost; of particular interest were constructions where movement resulted in an intermediate gap between clauses. Pupil diameter was recorded during auditory presentation of complex syntactic constructions; movement was manipulated (such that some conditions contained movement while others did not), as well as movement type (either causing an intermediate gap or not). Grammaticality judgments revealed no differences between the two groups, suggesting both were capable of comprehending these constructions. Pupil change slope measurements revealed no differences at the intermediate gap site, but showed similar facilitation during processing of the second gap site by both native and L2 speakers. This suggests that, contrary to the predictions of the SSH, L2 speakers are capable of constructing rich syntactic representations during the processing of intermediate gap constructions.

2.1 Introduction

Typical (canonical) English word order is Subject-Verb-Object with the first noun phrase (NP) in a construction being the subject of the verb and the second NP being the object. For example, in (1) below, the subject *John* occurs before the verb *brought*, which is followed by the object *cookies*.

1. John bought cookies yesterday after school.

However word order is not always enough to establish thematic roles (i.e. who is doing what to whom): elements within a sentence like (1) can be transformed into a question by replacing the object with *what* as seen in the Example (2) below. However, it is also possible to form a question by moving *what* to the front of the question (3).

2. John bought what yesterday after school?
3. What$_i$ did John buy t_i yesterday after school?

Constructions like (3) are known as filler gap dependencies, and it is generally assumed that these types of constructions are formed via movement operations (Chomsky, 1965): the filler (what$_i$) has moved from its position after the verb leaving a gap with a trace (t_i) at the movement site. The trace is phonologically silent but syntactically relevant as it is through the trace that grammatical properties are assigned to the filler *what*. The term filler gap dependency arises because correct interpretation of the role of the filler is dependent on the gap site. Some theorists argue for a traceless theory of grammar in which the filler is reactivated upon reaching the subcategorizing verb, Pickering and Barry (1991) call this the Direct Association Hypothesis (DAH). Nevertheless, there is a growing body of evidence supporting a grammar that contains filler gap dependencies (e.g. Clahsen & Featherson, 1999; Hestvik, Maxfield, Schwartz, & Shafer, 2007; Lee, 2004).

Most research into filler gap dependencies has focused on a speaker's native language (L1), however recent studies have also considered how second language learners process these types of dependencies in their second language (L2). Given their complexity, filler gap dependencies are an ideal construction to test theories of L2 language processing: by investigating the language processing of L2 learners we can determine whether they are able to acquire the same complex linguistic skills as native speakers. Such investigation can also provide a more comprehensive overview of how L2 speakers store moved elements and assign roles during processing, which in turn will lead to more accurate theories of second language acquisition and processing. In the current study we examine how proficient L2 leaners of English and native speakers process a type of filler-gap dependency known as an intermediate gap. We use a novel psychophysiological measure called pupillometry, in order to test a theory of L2 processing known as the Shallow-Structure Hypothesis. Before the empirical study is presented, we first outline the Shallow-Structure Hypothesis (SSH). This is followed by a discussion of the research comparing native and L2 processing of filler gap dependencies, particularly focusing on intermediate gap constructions.

2.1.1 Shallow Structure Hypothesis

In attempt to form an empirically based model of L2 language processing, Clahsen and Felser (2006) reviewed a series of studies comparing adult native speakers, child native speakers, and L2 language learners. In terms of syntax, Clahsen and Felser suggested that L2 speakers rely more on pragmatic and lexical-semantic information for sentence processing, as opposed to native speakers who rely on syntactic information. While L2 comprehension (in offline studies) appeared native-like, online processing diverged from that of native speakers, and these differences could not be explained by working memory restrictions, influence from their native language, incomplete acquisition of the L2 grammar, nor from slower processing speeds.

In light of these observations, Clahsen and Felser (2006) put forward the Shallow Structure Hypothesis (SSH) to explain the grammatical processing of L2 learners. According to the SSH, unlike native speakers, L2 learners do not compute full syntactic representations during comprehension, but rather construct shallow representations. This in turn leads to greater reliance on non-syntactic information (i.e. the lexical properties of the subcategorizing verb); L2 speakers are able to use lexical, pragmatic, and real word knowledge during comprehension in their non-native language, however they are only able to build shallow syntactic representations. Many of the studies investigating the SSH (including the current study) have focused on a type of filler gap dependency that contains an intermediate gap.

2.1.2 Intermediate Gaps

As noted above, a gap in a filler gap dependency involves an element moving from its canonical position and leaving a trace of itself (see Example 3), in contrast an intermediate gap is involved if the dependency spans across more than one clause. For example, in (4) the dependency of *who*$_i$ spans across the clause (*the consultant claimed*) and hence an intermediate gap *t'*$_i$ occurs at the boundary between this clause and the next (*the new proposal had pleased (the manager).*

4. The manager who$_i$ the consultant claimed *t'*$_i$ that the new proposal had pleased t_i will hire five workers tomorrow.

The longer the distance between a filler and a gap, the more processing resources needed; however, when an intermediate gap is present, it breaks up the filler-gap dependency into two shorter dependencies (*who*$_i$ to *t'*$_i$, and *t'*$_i$ to t_i) which aids comprehension and relieves working memory, thus facilitating the later integration of filler and gap (t_i). Gibson and Warren (1999) tested filler gap dependencies with an intermediate gap site (4) and compared these constructions to sentences without intermediate gaps like that shown in (5).

5. The manager who$_i$ the consultant's claim about the new proposal had pleased t_i will hire five workers tomorrow.
6. The consultant claimed that the new proposal had pleased the manager who will hire five workers tomorrow.
7. The consultant's claim about the new proposal had pleased the manager who will hire five workers tomorrow.

In (4) the moved wh-element is extracted across a verb phrase (VP) (*the consultant claimed*) thus signaling a new cyclical domain (an intermediate gap is present), while in (5) the extracted wh-element is extracted across a noun phrase (NP, *the consultant's claim*) and no intermediate gap is present. Additionally, the authors included baseline conditions that included no movement (6 & 7). Using a self paced reading task, Gibson and Warren found that intermediate gap sites did, indeed, facilitate filler and gap integration at the second gap site (following *pleased*) as shown by shorter total reading times for this segment (*had pleased*). At the intermediate gap site (e.g. in 4) they found longer reading times compared to the corresponding segment in the non-movement condition (6) suggesting that the participants did reactivate the filler at this site.

Marinis, Roberts, Felser, and Clahsen (2005) investigated how constructions with intermediate gap sites were processed by L2 learners of English, testing native speakers of Greek and German (languages that employ wh-movement), as well as native speakers of Chinese and Japanese (languages that do not employ wh-movement). If L2 learners were employing shallow representations of syntactic structures (as proposed by Clahsen & Felser, 2006) then Marinis et al. hypothesized that L2 learners would not make use of intermediate gap sites. The authors used the same movement stimuli as Gibson and Warren (1999) but attempted to improve the non-movement stimuli such that the non-movement conditions had the same number of words between the onset and the embedded verb making comparisons easier (as seen in 8-11).

8. Movement, verb phrase (filler gap, intermediate gap) - The nurse who$_i$ the doctor argued t'_i that the patient had angered t_i is refusing to work late.

9. Movement, noun phrase (filler gap)- The nurse who$_i$ the doctor's argument about the rude patient had angered t_i is refusing to work late.

10. Non-movement, verb phrase - The nurse thought the doctor argued that the rude patient had angered the staff in the hospital.

11. Non-movement, noun phrase - The nurse though the doctor's argument about the rude patient had angered the staff at the hospital.

Marinis et al. recorded reading times and comprehension accuracy using a non-cumulative moving-window procedure (Just, Carpenter & Woolley, 1982). Both L1 and L2 speakers of English showed similarly high comprehension accuracy (no differences were found between any of the L2 groups or the L1 group). There was also evidence of a slowing in reading time at the gap in the movement conditions (following *angered* in 8 and 9) for both groups, suggesting that the filler was being associated with the subcategorizing verb. However, while native speakers showed a slowing of reading at the intermediate gap site in the VP movement condition (following *argued* in 8) compared to the corresponding segment in the non-movement condition (following *argued* in 10), the L2 speakers showed no such slowing. Importantly at this site, there was an interaction between phrase type and movement, with the VP movement condition (which contains an intermediate gap) eliciting faster reading times than in the NP movement (which contains a gap but no intermediate gap), and with no difference between the two non-movement conditions (with no filler gap dependencies). The authors argued that the native speakers' slowed reading time at the intermediate gap site in VP movement sentences was evidence of filler reactivation which in turn led to faster reading at the following gap site. While the L2 group did show evidence of associating the filler with the subcategorizing verb in terms of longer reading times for movement compared to non-movement conditions, they did not show any reading time differences at final gap site between the VP and NP movement conditions, and they did not show any interaction between phrase type and movement. This suggests that the intermediate gap in VP movement conditions did not benefit the L2 learners in integrating the filler with the gap (e.g., after *angered* in 8).

Overall, these results led Marinis et al. to argue that L2 learners (even those whose native language contains similar syntax) employ a lexically driven gap filling strategy (in which the filler is linked to the lexical subcategorizer), not a syntactically driven strategy (in which the filler is linked to the gap and indirectly linked to the subcategorizer). L2 learners appeared not to be forming filler gap dependencies like native speakers, rather they seemed to be using a direct association in which they form a dependency between the filler and lexical subcategorizer (Pickering & Barry, 1991), in line with the SSH.

However, Dekydtspotter, Schwartz, and Sprouse (2006) expressed some concerns regarding the SSH; they argued that delayed or slowed processing in L2 speakers may make comparisons between native and L2 language processing problematic. For example, L2 research focusing on critical segments may be reflecting different moments in processing for L1 and L2 speakers. Dekydtspotter et al. (2006) re-analyzed the data from the Marinis et al. (2006) study, examining the segment following the intermediate gap to see whether there were any delayed effects in the L2 groups. They found, in this segment, slowed reading time in the VP (intermediate gap) movement condition for the Japanese and German native speakers. This suggests that these groups <u>are</u> reactivating the filler at the intermediate gap site similarly to native speakers, however the reactivation is slightly delayed. Interestingly, the Japanese language does not display this type of syntax (wh- movement), while German does, suggesting that the ability to use intermediate gaps is not dictated by the speaker's native language.

Dekydtspotter et al. (2006) argued, therefore, that the SSH is unable to explain the results from the Marinis et al. (2006) study, given that there seems to be a delayed computation of immediate gap sites (for at least some of the L2 learners). They also argued, for the native Chinese and Greek speakers, where no evidence for delayed facilitation was found, that the lack of evidence did not warrant the argument for shallow processing in these groups. This example highlights that observed differences in reading times at a critical segment may not in fact reflect disparate processing systems between native and L2 learners. It also highlights the need for further controlled and well designed studies to explore the underlying comprehension systems of L2 speakers.

More recently, Pliatsikas and Marinis (2013) investigated the role that previous naturalistic language exposure plays in L2 syntactic processing using the intermediate gap stimuli from the Marinis et al. (2006) study. A group of native English speakers and two groups of Greek L2 learners of English participated in their study (one group had only classroom exposure, and the other had approximately 9 years of naturalistic English exposure, but both groups had the same level of English proficiency). They found evidence that only those who learned L2 in a naturalistic setting showed evidence of making use of intermediate gaps, however the processing of the intermediate gap seemed to be delayed (as evidenced by increased reading time following the intermediate gap). Nevertheless, the intermediate gap ultimately facilitated the reintegration of the filler at the subcategorizing verb (as evidenced by shorter reading times at the final gap). These findings suggested that with adequate naturalistic exposure, L2 learners of English are capable of employing native-like syntactic processing contrary to the SSH.

To summarize, research has investigated the processing of filler gap dependency constructions with intermediate gaps sites in native speakers and second language learners. The SSH was highlighted as an influential theory that, during comprehension, assumes L2 learners construct shallow representations in which full syntactic representations are not constructed, which in turn causes an over reliance on non-syntactic information (unlike native speakers). While some data supports the SSH, several studies have found results that contradict the SSH, particularly in the processing of intermediate gaps (e.g. Dekydtspotter, et al., 2006; Pliatsikas & Marinis, 2013). Given these contradictions, it is important to further test the SSH by investigating the processing of intermediate gaps, in order to form a more accurate and comprehensive theory of second language acquisition and processing.

2.1.3 Study Aims

In the current study, pupillometry was employed to investigate the processing of intermediate gap constructions with native and L2 speakers of English using the same stimuli as Marinis et al. (2006). The aim was to test the SSH and to provide further insight into second language acquisition and processing.

Pupillometry is a psychophysiological measure that involves recording pupil diameter change in response to a stimulus. Change in pupil diameter is believed to reflect the resources expended during mental activity with larger pupil size reflecting greater use of cognitive resources (e.g. Hess & Polt, 1960; Kahneman & Beatty, 1966). Pupillometry has been shown to be an effective method for measuring cognitive processing load in a variety of linguistic tasks: from simple word recall (e.g. Kahneman, 1973; Kahneman & Wright, 1971) to sentence processing (e.g. Fernandez, 2013; Just & Carpenter, 1993). Both Fernandez (2013) and Just and Carpenter (1993) showed that pupil size was an index of processing difficulties associated with forming filler gap dependencies. Investigating the processing of these constructions using pupillometry provides us with a new methodology that is a more direct measure of filler gap dependency formation than self-paced reading. Consequently, here we extend this novel methodology to intermediate gap constructions where it gives a unique psychophysiological measure of the difficulty associated with filler gap dependencies in these constructions. Additionally, we used pupillometry with aurally presented stimuli, which is more natural than self-paced reading (as far as we are aware this is the only study investigating intermediate gaps to have used aurally presented stimuli).

We hypothesized that there would be an increased slope of pupil change at the intermediate gap site in the VP movement condition (following *argued* in Example 8: The nurse who$_i$ the doctor argued t'_i that the patient had angered t_i is refusing to work late) compared to the corresponding segments in the two non-movement conditions, and the NP movement condition. If participants are sensitive to the intermediate gap site, a steeper slope of pupil change at this site for the VP movement condition is predicted, as there should be an increase in processing load upon reactivation of the filler at the intermediate gap. This is the site in the VP movement condition where the dependency is broken up. There should be no such reactivation for the other conditions, given that there is no intermediate gap in the NP movement condition, and there are no gaps in the non-movement conditions. If, as predicted by the SSH, L2 speakers are not sensitive to the intermediate gap, no differences should be seen between the NP and VP movement conditions for this group.

At the second critical segment (following *angered*), we hypothesized there would be a larger decrease in slope of pupil change for the movement conditions (where there is a filler gap dependency) compared to the non-movement conditions. This is because processing load will be reduced as the parser associates the filler and the gap, thus reducing memory load given that the parser no longer has to hold the filler in working memory (Fernandez, 2013; Just & Carpenter, 1993). Importantly, we hypothesized a decreased slope of pupil change at this segment for the VP movement condition compared to the NP movement condition, given that, as stated above, the intermediate gap site in the VP condition should have facilitated the subsequent gap processing. If the L2 group was not sensitive to the intermediate gap, as predicted by the SSH, there should be no differences between the two movement conditions.

2.2 Methods

2.2.1 Participants

2.2.1.1 L2 Speakers

Thirty students recruited from the University of Potsdam in Germany participated for payment or as part of their undergraduate course requirements. All participants were native speakers of German and had normal or corrected to normal vision. All participants took the Oxford Online Placement Test to measure their English proficiency, and only those who were at a C1 or C2 level based on the Common European Framework of Reference for Languages were included in the experiment (individuals who score in these levels are considered proficient users; The Common European Framework, 2001). Of the thirty participants, eight participants did not meet the proficiency criteria; therefore twenty-two participants (17 female) were included in the analysis.

L2 speakers were selected to have limited naturalistic exposure to English, in order to test those L2 speakers who are most likely to make use of shallow processing (Pliatsikas & Marinis, 2013). Only one of the 22 included participants had spent significant time in an English speaking country, for 10 months as an adult. This was the only participant that had even limited adult naturalistic exposure to English, all other participants had no naturalistic English exposure.

2.2.1.2 English Speakers

Fourteen participants (8 female) recruited from the University of Potsdam and from the Berlin area participated for payment. All participants were native speakers of English and had normal or corrected to normal vision.

2.2.2 Stimuli

There were four sentence types, as seen in Table 2.1. This task consisted of 50 trials: four practice trials, twenty critical trials (five of each sentence type), and twenty-six fillers (eight of which were grammatical and eighteen of which were ungrammatical). The sentences were recorded by a female native speaker of American English in the phonetics lab at the University of Potsdam with a sampling rate of 44100 Hz and saved as .wav. The recordings were randomly placed into four versions and rotated in a Latin square design; the practice trials were the same across the four versions. Critical items were counterbalanced across participants so each participant was presented with one condition of each item, and order of the critical items was randomized across version.

Table 2.1: Example sentences for each of the four experimental conditions

	Verb Phrase	Noun Phrase
Movement	The nurse who$_i$ the doctor argued t'_i that the patient had angered t_i is refusing to work late.	The nurse who$_i$ the doctor's argument about the rude patient had angered t_i is refusing to work late.
Non-movemen	The nurse thought the doctor argued that the rude patient had angered the staff in the hospital.	The nurse though the doctor's argument about the rude patient had angered the staff at the hospital.

The critical items were identical to those used by Marinis et al. (2005). There were four critical conditions contrasting movement and phrase type (see Table 2.1); in the movement conditions the first NP (*the nurse*) preceded a relative clause, this relative clause began with a wh-pronoun, which was the object of the verb (*angered*). In the movement VP conditions there was an intermediate bridge verb (e.g. *argued*) between the filler (who_i) and the gap site (t_i), this verb type allowed wh-movement thus forming an intermediate gap (t'_i). This was not the case in the movement NP conditions, and therefore there was no intermediate gap site between *who* and the gap site following *angered*. The non-movement conditions were formed from the movement items by removing the filler gap dependency and adding an additional level of embedding to avoid differences in the structural complexity between the movement and non-movement conditions. Given that the distance between the filler and the gap differed in the two movement conditions, the non-movement conditions mirrored the differences and hence controlled for this confound during analysis.

2.2.3 Apparatus

Stimulus presentation was programmed using Tobii Studio software. Pupil diameter was recorded with a Tobii T120, sampling at at 120 Hz. Tracking was binocular but only the right eye was used for analysis. The eye tracker was built into a 17-inch thin film transistor liquid crystal color display monitor (1,280 x 1,024 pixel resolution). Participants sat approximately 60-65 cm away from the display and tracking was remote. Eyes were calibrated using a 9-point sequence.

2.2.4 Procedure

The participants were informed that a fixation point would appear on the screen for 2000 ms (to allow the pupil to adjust to screen luminance), after which the fixation point would turn into a fixation cross and a recorded sentence would be heard; the fixation cross remained on the screen for the entirety of sentence presentation and for an additional 2000 ms post recording offset. Participants were asked to focus on the fixation cross while it was on the screen, attend to the auditorily presented sentences, and to try to avoid blinking during sentence presentation. After the fixation cross disappeared from the screen a new screen appeared with a scale that instructed the participants to rate the grammatical acceptability of the sentence they had heard from a 1-7 (one being least grammatical /seven being most grammatical). Ratings were input by pressing the corresponding number key on a keyboard provided.

2.2.5 Analysis

The design was 2x2x2 (movement x phrase type x speaker group). Movement refers to whether the item contained a moved constituent. Phrase type refers to whether there was a NP between the filler and gap or a VP between filler and gap (or corresponding segment in the non-movement conditions). The VP movement condition was that which allowed for an intermediate gap. Both variables were manipulated within subject. Speaker group refers to whether the participant was a native speaker of English or a L2 speaker of English (native speaker of German).

Blinks were filtered out by replacing the missing values with linear interpolation, and the pupil was normalized by participant and by item, so all items began at zero. The pupil takes 1.2 seconds on average to reach its maximum diameter (Just & Carpenter, 1993), consequently, pupil change was analyzed across a 1.2 second window at two critical segments as explained above: 1) starting at the onset of the first verb (*argued*) in the VP conditions and at the corresponding word in the NP conditions (*argument*), and 2) following the onset of the second verb (*angered*) across the four conditions. Although there has been debate regarding L2 learners having slower processing, given the relatively long time window being analyzed, we anticipate any processing time differences between the two groups will be encompassed within this 1.2 second window, we therefore do not analyze the following segment (cf. Dekydtspotter et al., 2006).

Before analysis, data was screened for outliers; any data point three or more standard deviations from the mean of the conditions was replaced with the mean of that condition. All trials were averaged within each condition resulting in 4 vectors for each participant (one for each of the 4 conditions). These vectors were submitted to a simple regression where pupil size served as the dependent variable and time served as the independent variable. The slope of pupil change over time (i.e. the unstandardized regression coefficient) was the main dependent variable in this study. We also calculated the peak amplitude (largest size that the pupil reached during the critical time windows), and the peak latency (the time by which the peak amplitude was reached) for each vector; these measures are typical in pupillometry research (e.g. Schmidtke, 2014; Zekveld, Kramer, & Festen, 2012).

Grammaticality rating, pupil change slope (unstandardized regression coefficient), peak amplitude, and peak latency were each analyzed using linear mixed effect models using R (R Core Team, 2012) and *lme4* (Bates, Maechler & Bolker, 2012) and Tukey's contrasts from the *multcomp* package (Horthorn et al., 2013). All of which had the fixed effects of movement (movement/non-movement), phrase type (NP/VP), and speaker (native/L2), and random effects of participants; the random effect of participants was maximally specified with random slopes for movement, phrase type, and speaker. Models were compared such that a full model was contrasted with a reduced model to test the factor in question; nonsignificant factors were excluded from the model by contrasting each main effect and interaction, starting with speaker, followed by movement, phrase type, and finally the interactions. Model comparison was based on log likelihood ratio tests.

2.3 Results

2.3.1 Grammaticality Ratings

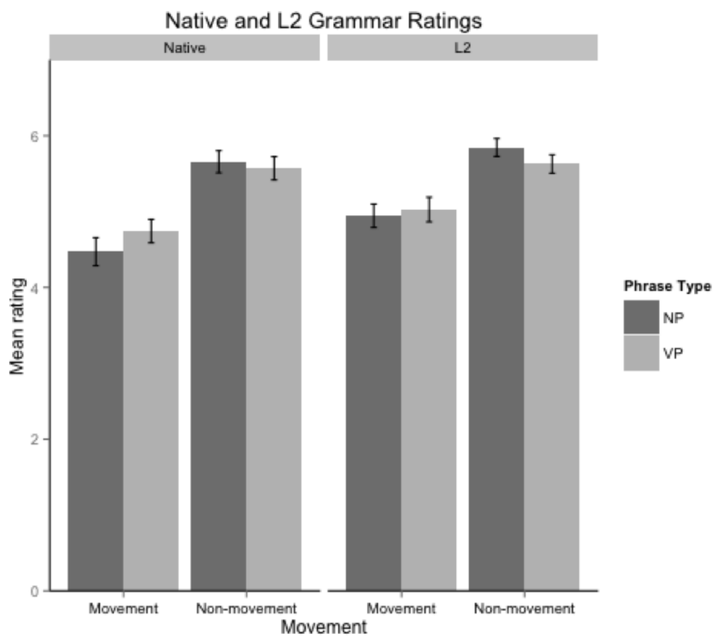

Figure 2.1: Mean grammaticality judgment for Native and L2 speakers (error bars in all graphs represent the standard error of the mean)

Grammaticality ratings of sentences were taken as a measure of ease of comprehension; speaker group did not significantly affect grammaticality ratings ($X^2(1)$= 2.47, p=0.11) but there was a main effect of movement ($X^2(1)$= 7.17, p<0.001) and a main effect of phrase type ($X^2(1)$=26.95, p<0.001): non-movement conditions evoked higher grammaticality ratings than movement conditions (Mean rating: non-movement 5.69 (SD=1.25), movement 4.83 (SD=1.57), and NP conditions (Mean rating: 5.27 (SD: 1.50) higher ratings than VP conditions (Mean rating: 5.26 (SD: 1.47), refer to Figure 2.1. There were no significant interactions ($X^2(1)$= 1.45, p=.23), comparing models with and without the interaction terms. This shows that native English speakers and L2 speakers rated sentence grammaticality similarly and there was no difference in their pattern across sentences.

2.3.2 First Verb

Analysis of this segment investigated whether there was an effect of the intermediate gap on pupil change slope, compared to the other conditions without an intermediate gap.

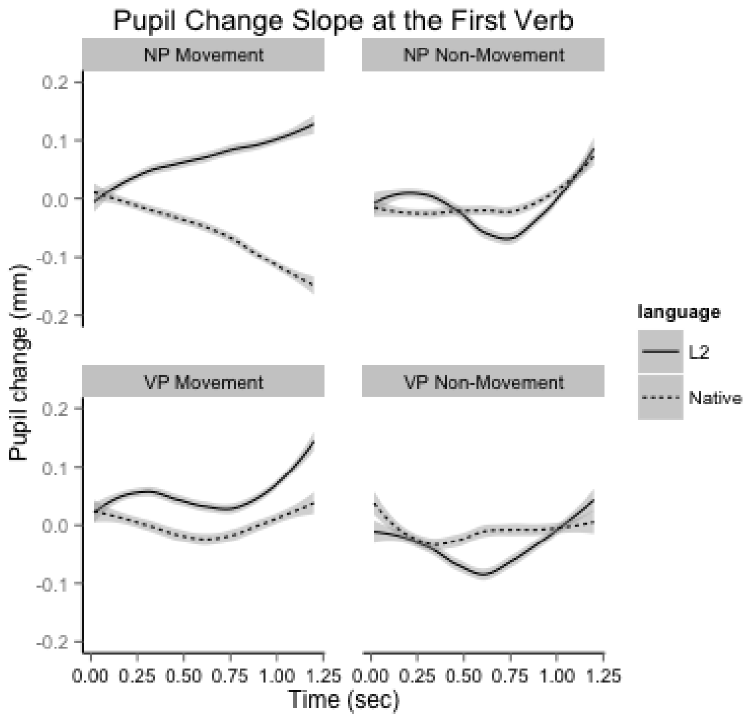

Figure 2.2: Pupil change from the onset of the first verb: Smoothed pupil change slope across time, the grey highlighting indicates standard error.

At the first verb there were no significant main effects on pupil change slope (speaker group: $X^2(1)= 1.08$, p=0.29; movement: $X^2(1)= 1.69$, p=0.20; phrase type: $X^2(3)= 0.06$, p=.79). There were no interactions ($X^2(3)= 4.85$, p=.18) on pupil change slope when comparing a model with and without an interaction (see Figure 2.2).

Peak amplitude also showed no significant main effects (speaker group $X^2(4)= 1.28$, p=0.86, movement $X^2(4)= 1.12$, p=0.89, phrase type $X^2(4)=$ 1.62, p=0.44 nor interactions $X^2(3)= 2.55$ p=0.47; see Figure 2.3). Additionally, there were no significant main effects nor interactions on peak latency (speaker group $X^2(4)= 1.27$, p=0.86, or phrase type $X^2(4)= 3.17$, p=0.53); comparing models with and without interactions ($X^2(4)= 2.65$, p=0.45; refer to Figure 2.4).

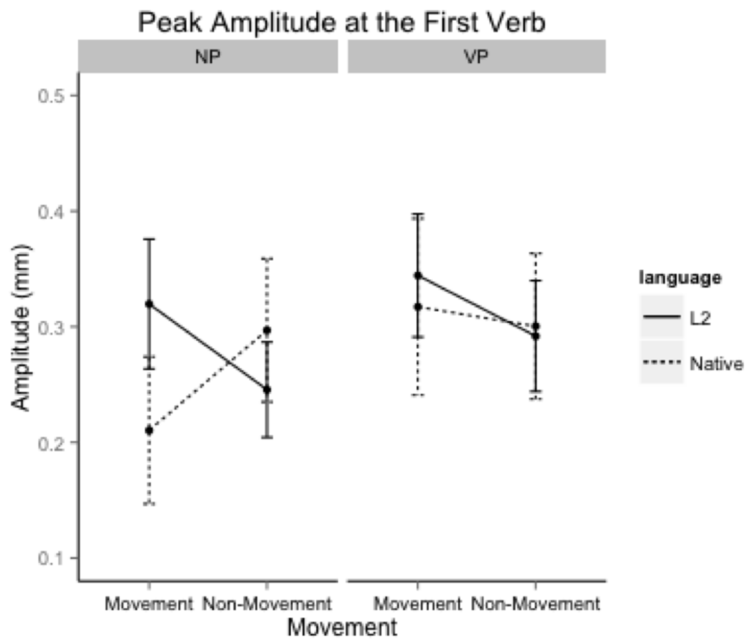

Figure 2.3: Peak amplitude at the first verb (bars represent standard error)

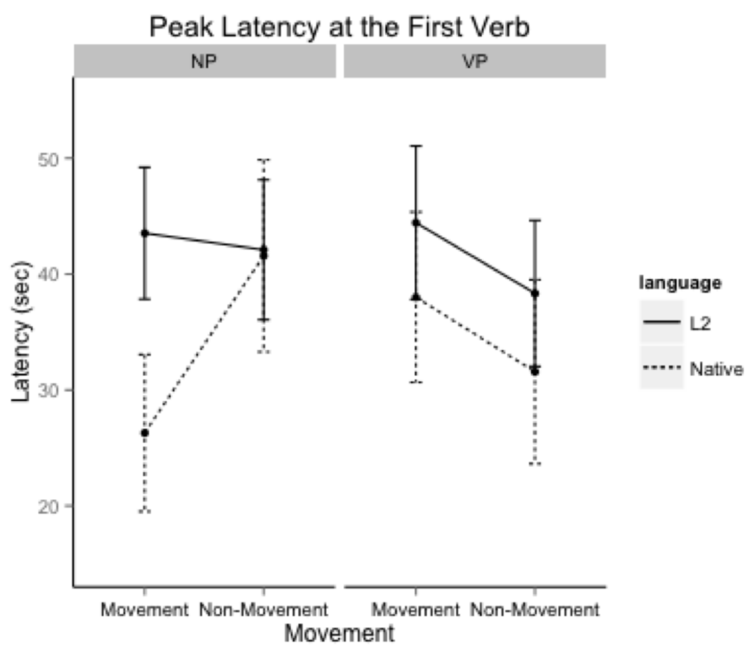

Figure 2.4: Peak latency at the first verb (bars represent standard error)

This indicates that the segment containing the intermediate gap and the corresponding segments (in the NP movement and the two non- movement conditions) did not evoke a measurable change in the pupil change slope, peak amplitude, or peak latency in either native speakers or L2 speakers.

2.3.3 Second Verb

This segment allowed investigation of whether there was an effect of the gap site on pupil size for both the VP and NP movement conditions, compared to the other conditions that did not have a moved element (and therefore no gap site).

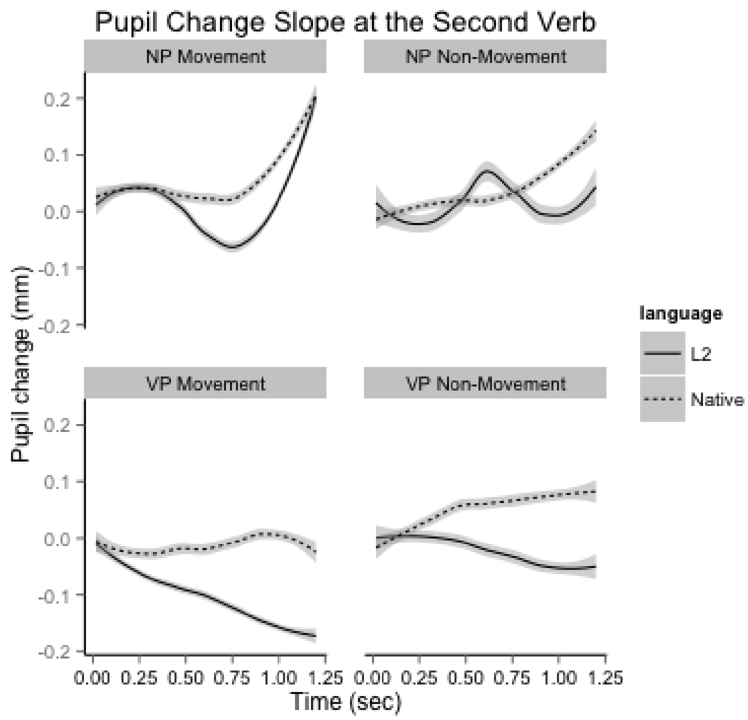

Figure 2.5: Pupil change from the onset of the second verb: Smoothed pupil change slope across time, the grey highlighting indicates standard error).

At the second verb, there were no main effects on peak amplitude (refer to Figure 2.6) (speaker group: $X^2(1)= 0.01$, p=0.93; movement: $X^2(1)= 0.01$, p=0.92; phrase type: $X^2(1)= 0.005$, p=0.94), nor interactions when comparing models with and without interactions ($X^2(3)= 0.59$, p=0.89). Also, there were no main effects on peak latency (speaker group: $X^2(1)= 1.27$, p=0.86; movement: $X^2(1)= 0.21$, p=0.64; phrase type: $X^2(1)= 0.13$, p=0.72), nor interactions when comparing models with and without interactions ($X^2(3)$= 1.81, p=0.61), refer to Figure 2.7.

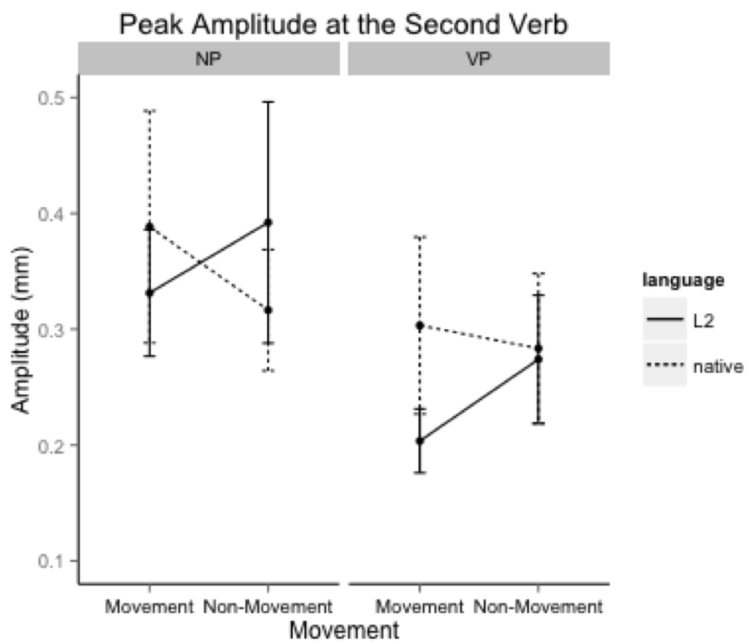

Figure 2.6: Peak amplitude at the second verb (bars represent standard error)

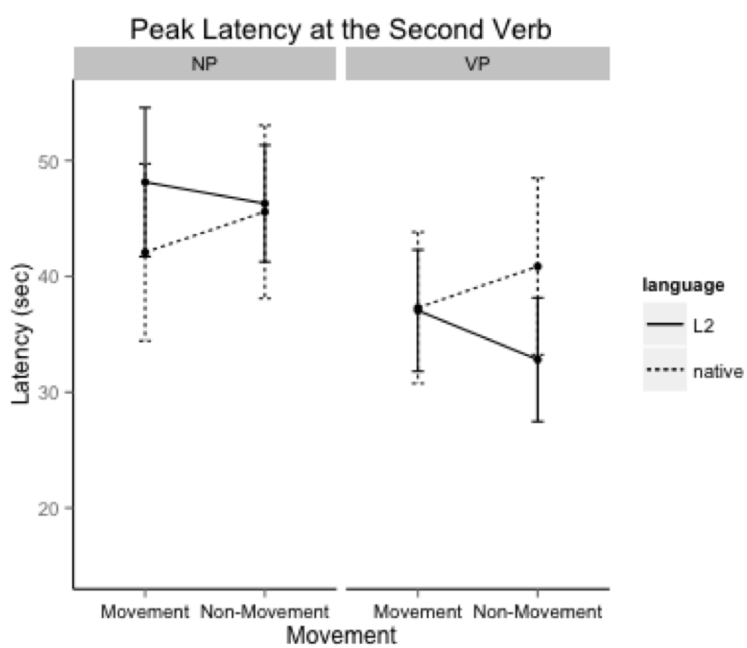

Figure 2.7: Peak latency at the second verb (bars represent standard error)

However, analysis of pupil change slope showed main effects of speaker, movement, and phrase type on pupil change slope ($X^2(1)= 3.77$, $p=0.05$) compared to a model with one ($X^2(1)= 156.94$, $p<0.001$) or two main effects ($X^2(1)= 9.30$, $p<0.01$); see Figure 2.5. Native speakers showed a significantly greater reduction in pupil change slope than the L2 speakers, the NP condition evoked a significantly greater increase in pupil change slope than the VP condition, and the non-movement conditions evoked a significantly greater increase in pupil change slope than the movement conditions.

Critically, however, there was a significant interaction between speaker, movement, and phrase type ($X^2(4)= 90.54$, $p<0.001$). Given this interaction, the movement and non-movement conditions were analyzed separately. Pupil change slope at the second verb was submitted to linear mixed effect models for the movement and the non-movement conditions separately; both of which had the fixed effects of speaker group (native/ L2), and phrase type (NP/VP) with random effects of participants; the random effect of participant was maximally specified with random slopes for speaker group and phrase type.

For the movement condition the main effect of speaker group approached significance ($X^2(1)= 3.61$, $p=0.08$) with a more positive change in pupil slope for the native speakers. There was also a significant main effect of phrase type ($X^2(1)= 0.005$, $p=0.05$) with a relatively increased change in pupil slope evoked by the NP condition. There was no interaction of speaker group and phrase type ($X^2(1)=0.21$, $p=0.64$). For the non-movement condition there was main effect of speaker group ($X^2(1)= 4.04$, $p<0.05$), with the native group having a greater increase in pupil change slope. There was no main effect of phrase type ($X^2(1)= 0.18$, $p=0.66$), nor an interaction ($X^2(1)=0.31$, $p=0.57$).

In summary, at the second segment, native and L2 speakers showed significantly different changes in slope in both movement and non-movement conditions. However, both groups showed the same patterns in terms of phrase type with no difference in the non-movement condition, and in the movement condition a greater increase in pupil slope for the NP movement condition than the VP movement condition (where, if anything there was a decrease in pupil size). This would suggest that there was some effect of the intermediate gap on the actual gap site in the VP movement condition, which required fewer processing costs when releasing the filler in the gap site compared to the movement NP condition in which no intermediate gap site was present. However, while there was a three way interaction between speaker, movement, and phrase type, investigation of the movement and non-movement conditions independently found no evidence of a two way interaction between speaker group and phrase type in either condition: there was no evidence that L2 speakers were processing movement VP sentences (the critical 'intermediate gap' condition) any differently to native speakers.

2.3.3.1 Overall summary

Both groups rated movement conditions lower in grammaticality than the non-movement conditions, with no differences between NP and VP phrases or between speaker groups. There were no significant effects across conditions in terms of any measure of pupil response at the first critical segment (following the intermediate gap in the VP movement condition and the corresponding segment in the other conditions). At the second segment (the site of the final gap site and the corresponding segments), importantly, there was an interaction between language, phrase type, and movement, critically, the movement NP condition evoked larger pupil change than the movement VP condition. The native and L2 group differed in their pupil change slope for both the movement and non-movement conditions (with the native group displaying a steeper pupil change slope), but showed similar patterns with the NP movement condition evoking a greater increase in pupil change slope than the VP movement condition (suggesting facilitation from the intermediate gap).

2.4 Discussion

Clahsen and Felser (2006) proposed the Shallow Structure Hypothesis (SSH) which states that L2 speakers do not form full syntactic representations during processing, rather they over rely on non-syntactic representations. While there is some research that has supported this account (e.g. Clahsen & Felser, 2006; Marinis et al., 2005; Williams et al., 2001), other research, using intermediate gaps, has challenged this claim that L2 speakers use shallow syntactic processing (e.g. Dekydtspotter, et al., 2006; Pliatsikas & Marinis, 2013). This study further tested the SSH by investigating the processing of intermediate gap constructions with a group of native speakers and a group of L2 speakers of English, using pupillometry, a novel methodology, as a measure of processing cost.

The analysis focused on pupil change slope, peak amplitude, and peak latency during two critical time windows in 4 construction types (examples are reiterated in 12-15)

12. Movement, verb phrase (filler gap, intermediate gap) - The nurse who$_i$ the doctor argued *t'*$_i$ that the patient had angered t_i is refusing to work late.

13. Movement, noun phrase (filler gap)- The nurse who$_i$ the doctor's argument about the rude patient had angered t_i is refusing to work late.

14. Non-movement, verb phrase - The nurse thought the doctor argued that the rude patient had angered the staff in the hospital.

15. Non-movement, noun phrase - The nurse though the doctor's argument about the rude patient had angered the staff at the hospital.

The two critical time windows were at the site of the intermediate gap in the movement VP condition (following *argued* in 12), and the corresponding segments in the other three conditions (following *argument* in 13 and *argued* in 14 & 15), and at the gap site in the movement conditions and the corresponding segments in the non-movement conditions (following *angered*). Behaviorally, there were no differences in grammaticality ratings between the native and L2 speakers; for both speaker groups non-movement conditions were rated higher than movement conditions, and NP conditions higher than VP conditions. This suggests that both groups have similar views of the grammaticality of these constructions.

At the first critical segment (the intermediate gap site after the verb, and corresponding segments) we hypothesized an increased pupil change slope for the VP movement condition compared to the VP non-movement condition (given that the in the VP movement condition the parser should be reactivating the filler at the intermediate gap site which should cause an increase in pupil size). The SSH predicted that L2 speakers should not be sensitive to intermediate gaps while native speakers should be. Consequently, the native group should show sensitivity to the intermediate gap in the VP movement condition, while the L2 group should show no such sensitivity. However, unlike previous research using reading paradigms (e.g. Marinis et al., 2005; Gibson & Warren, 2004), and contrary to our predictions, pupillometry revealed no significant differences between conditions at the first critical segment. It is possible that the cognitive costs associated with processing an intermediate gap site are not great enough to cause a measurable pupil change. Clearly more research is needed to further investigate this issue.

At the second critical segment (the gap site and the corresponding segments), we hypothesized relatively less increase in pupil diameter (a less positive pupil change slope) for the movement conditions than the non-movement conditions indexing the release of the filler in working memory at the gap site, and our data was in line with these predictions. This suggests that when forming the dependency between the filler and gap in the movement conditions (and releasing the moved element from the working memory), processing costs are relieved, and this is reflected in the pupil response.

Additionally at the second segment we hypothesized a larger decrease in pupil change slope for the VP movement condition compared to the NP movement condition indexing the facilitation of the intermediate gap. The SSH predicts there would be no differences between the NP and VP movement conditions for the L2 speakers at the second critical segment, given that the L2 speakers would not have taken advantage of the previous intermediate gap due to the shallow syntactic processing.

At the second critical segment, there were no effects of phrase type on either speaker group in processing of sentences without movement. However, in the movement condition there was a significantly greater reduction in pupil diameter for the VP movement condition compared to the NP movement condition, suggesting that the processing of the gap site in the VP condition required less cognitive effort overall. The VP movement conditions contained an intermediate gap which research has shown facilitates the processing of the gap site in this segment (e.g., Gibson & Warren, 2004; Marinis et al., 2005), and pupil diameter appears to reflect this facilitation. In previous research, difficulties at the first segment in the movement VP conditions have been interpreted as suggesting that the filler is being reactivated at the intermediate gap. It is this reactivation that is argued to lead to the facilitation when forming the filler gap dependency at the gap site (second segment, following *angered*). Interestingly, however, we did not detect difficulties at the intermediate gap site (first segment), but still found facilitation at the gap at the second segment after *angered*. This seems to support the likelihood that pupil change is not a sensitive enough index to detect the difficulties at the intermediate gap.

Additionally, data from this segment showed some indication that native speakers display an increased pupil change slope across both movement and non-movement conditions compared to L2 speakers (a marginally significant effect). This suggests that the native speakers may be investing more resources during processing. It is also possible that for the L2 group, in the movement condition, holding the filler is more difficult, and thus release of the filler at the gap site causes a steeper decrease in pupil slope (given that more resources were needed to hold the filler in working memory). Nevertheless, both groups show similar processing patterns across phrase types. Like native speakers, L2 speakers seem to be using fewer processing resources in the movement VP condition than the NP condition. This is striking and goes against the predictions made by the SSH, which argues that L2 speakers underuse syntactic information during parsing. In their study Marinis et al. (2006) found no evidence that L2 speakers were making use of intermediate gaps. The results of the current study have interesting implications for theories of second language acquisition and processing, and particularly the SSH. Given the similar pattern displayed by both speaker groups, the data in the current study contradict the SSH. The data from the second segment shows that both native and L2 speakers are facilitated by the intermediate gaps in the movement VP conditions; this suggests that L2 speakers are making rich syntactic representations during processing, and hence appears to contradict the SSH (e.g. Clahsen & Felser, 2006; Marinis et al., 2005.; Williams et al., 2001).

The data reported here are consistent with earlier research that also found that L2 speakers are capable of making use of intermediate gaps (e.g. Dekydtspotter, et al., 2006; Pliatsikas & Marinis, 2013), but using a novel methodology that provides an index of processing costs and using auditory presentation. Dekydtspotter et al. (2006), found evidence that some L2 learners were indeed making use of intermediate gaps (in the movement VP condition) but that there was evidence for a processing delay. However, we found no difference in peak latency. Hence, while we found evidence for intermediate gap use by L2 speakers, it seems that there was no processing delay. Given that this study is, the best of our knowledge, the first to investigate intermediate gaps using auditory presentation, it is possible that the presentation mode played a role in our results. Hence, the lack of delay in the L2 group, could be due to the auditory presentation style. For example, it is possible that some suprasegmental features play a role (e.g. intonation, pausing, etc.) and may affect the way that the constructions are processed; more research needs to be done to further investigate processing differences across time between native and L2 speakers using pupillometry with auditory stimulus presentation.

2.5 Conclusion

The SSH argues that L2 speakers do not form full syntactic representations, but rather rely on non-syntactic information during comprehension, causing them to form shallow representations during processing (Clahsen & Felser, 2006). In this study the SSH was tested by investigating pupil response, as a measuring of processing costs, during the auditory presentation of sentences with filler gap dependencies and with or without an intermediate gap (and corresponding non-movement control sentences) by native and L2 speakers of English.

While there were significant differences between the two speaker groups, critically, when analyzing the movement condition separately, there was no interaction between speaker group and phrase type: consequently we do not have evidence that L2 speakers underuse syntactic information. Indeed, we found evidence supporting a position that L2 speakers do make use of the intermediate gap, and form filler gap dependencies. This is in contrast to the data from Marinis et al. (2005), but in accordance with a growing body of research that has found L2 speakers are capable of making use of intermediate gaps (e.g. Dekydtspotter, et al., 2006; Pliatsikas & Marinis, 2013). Overall, it seems that the L2 speaker parsing mechanism is capable of more in-depth processing than previously argued. Theories of second language acquisition and processing must take this into account, and provide a more comprehensive model of the rich and complex processing of which L2 speakers are capable.

Finally, it is important to note that all previous research has used a reading task, this is the first study to investigate the processing of intermediate gaps using auditory presentation. Consequently, it is possible that the differences in the results between this study and those in the past (e.g., Clahsen & Felser, 2006; Dekydtspotter, et al., 2006; Pliatsikas & Marinis, 2013) reflect the different nature of processing in the two modalities. We would argue that auditory presentation provides a more natural reflection of language processing and is more likely to reflect the automatic processes used by listeners.

Further research to replicate and extend the findings presented here, using converging methodologies is clearly warranted.

2.6 References

Bates, D.M., Maechler, M., & Bolker, B. (2012). *lme4: Linear mixed-effects models using S4 classes*, R package version 0.999999-0.

Chomsky, N. (1965). *Aspects of the Theory of Syntax.* MA: MIT Press.

Clahsen, H., & Featherson, S. (1999). Antecedent priming at trace positions: Evidence from German scrambling. *Journal of Psycholinguistic Research, 28*, 415-437.

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics, 27*, 3-42.

Dekydtspotter, L., Schwartz, B.D., & Sprouse, R.A. (2006). The comparative fallacy in L2 processing research. In M.G. O'Brien, C. Shea, & Archibald, J. (Eds.) *Proceedings of the 8th Generative Approaches to Second Language Acquisition Conference* (pp. 33-40). Summerville, MA: Cascadilla Proceedings Project.

Fernandez, L., B. (2013). *Pupillometry and Filler Gap Dependencies.* (Unpublished masters of philosophy dissertation). Northumbria University, Newcastle upon Tyne, England.

Gibson, E., & Warren, T. (1999). *The psychological reality of intermediate linguistic structure in long-distance dependencies.* Unpublished manuscript, MIT, Cambridge, MA.

Hestvik, A., Maxfield, N., Schwartz, R., & Shafer, V. (2007). Brain responses to filled gap. *Brain and Language*, 301-316.

Hothorn, T., Bretz, F., Westfall, P., Heilberger, R.M, & Schuetzenmeister, A. (2013). *Simultaneous Inference in General Parametric Models.* R Foundation for Statistical Computing Vienna, Austria.

Just, M.A., & Carpenter, P.A. (1993). The intensity dimension of thought: Pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology, 47(2)*, 310-339.

Kahneman, D. (1973). *Attention and Effort.* Englewood Cliffs, N.J.: Prentice-Hall.

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science, 154*, 1583-1585.

Kahneman, D., & Wright, P. (1971). Changes of pupil size and rehearsal strategies in a short-term memory task. *Quarterly Journal of Experimental Psychology, 23*, 187-196.

Lee, M.-W. (2004). Another look at the role of empty categories in sentence processing (and grammar). *Journal of Psycholinguistic Research, 33*, 51-73.

Marinis, T., Roberts, L., Felser, C., & Clahsen, H. (2005). Gaps in second language sentence processing. *Studies in Second Language Acquisition, 27*, 53-78.

Pickering, M., & Barry, G. (1991). Sentence processing without empty categories. *Language and Cognitive Processess, 6*, 229-259.

Pliatsikas, C., & Marinis, T. (2013). Processing empty categories in second language: When naturalistic exposure fills the (intermediate) gap. *Bilingualism: Language and Cognition, 16*, 167-182.

R Core Team (2012). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria.

Schmidtke, J. (2014). Second language experience modulates word retrieval effort in bilinguals: evidence from pupillometry. *Frontiers in Psychology, 5*, 137.

Williams, J., Mbius, P., & Kim, C. (2001). Native and non-native processing of English wh-questions: Parsing strategies and plausibility constraints. *Applied Psycholinguistics, 22*, 509-540.

Zekveld, A.A., Kramer, S.E., Festen, J.M. (2010). Pupil response as an indication of effortful listening: The influence of sentence intelligibility. *Ear and Hearing, 31*, 480-490.

Chapter 3

In the previous chapter, pupillometry was used to examine one kind of linguistic parameter - intermediate gap processing in sentences with movement. In this next chapter the scope is extended and pupillometry is used to investigate the processing of word order and case marking in adult German speakers. This chapter tests two competing theories which attempt to explain the difficulties that are associated with object-first constructions in German. These two theories are based on previous research that has employed event related potentials (ERPs) to examine processing of these constructions. Here we employed pupillometry with two aims: first to provide further empirical data on the processing of these structures; second, we aimed to use these data to establish that pupillometry could be an effective tool for investigating word order and case marking in German. This was not only to further validate pupillometry as a methodology, but also as vital preliminary research for the study reported in Chapter 4 which investigates these same constructions with 4-year-old native German speaking children.

Investigating the Processing of German Word Order in Adults

Abstract

This study investigates the processing of canonical and non-canonical sentence word order in German by measuring participants' pupil response. The relative order of the sentence subject and the sentence object is relatively free in German with both orders being possible. However, compared to the subject-first (SVO) order, object-first (OVS) orders evoke more processing costs in adults and are harder for young children to understand. This is one of the first studies to use pupillometry to investigate word order and the effect of case marking in German. By confirming that pupillometry is sensitive to word order and case marking, this study further builds the evidence that pupillometry is an effective tool for testing a range of linguistic constructions. Pupil change slope change slope and comprehension were measured during presentation of oral sentences which manipulated word order (SVO, OVS) and whether or not the role of the first noun phrase was ambiguous or not depending on its case marking. We found no differences in pupil response at the first noun phrase regardless of word order. This suggests that previous reports of differences between the OVS and SVO sentences were unlikely to reflect working memory differences but rather may be due to the unexpectedness of an object-first construction. At the disambiguating 2nd noun phrase, unambiguous SVO constructions revealed a shallower pupil change slope compared to ambiguous SVO, unambiguous OVS, and ambiguous OVS sentences. This suggested that difficulties arose when the OVS order

became apparent, and the parser had to revise the original and preferred SVO interpretation. Additionally, across the whole sentence there was larger pupil change for OVS constructions compared to SVO, with the ambiguous conditions evoking a larger peak pupil size than the unambiguous. Overall this study shows that pupillometry is sensitive to the processing required in the comprehension of different word orders in German, and provides us insight into the difficulties associated with these different word orders.

3.1 Introduction

A key aspect of sentence meaning is expressing who is doing what to whom, and different languages use different means to mark these roles. In English, word order is the essential cue, with the subject (as the agent of an action) typically appearing before the object. However, this is not the case across all languages. For example, in German, a case-marked language, the subject and object in the simple transitive are marked by the use of case which mainly is cued by the article form: the subject of the sentence is marked by articles in nominative case, and the object by articles in accusative case.

Table 3.1: Example sentences, contrasting SVO and OVS word orders disambiguated by case.

SVO	Der Affe fängt gleich den Frosch
	The monkey$_{\mathrm{nom}}$ chases the frog$_{\mathrm{acc}}$
	The monkey chases the frog
OVS	Den Frosch fängt gleich der Affe
	The frog$_{\mathrm{acc}}$ chases the monkey$_{\mathrm{nom}}$
	The monkey chases the frog

In Table 3.1 both the SVO and OVS constructions have a noun with the article in nominative ($_{\mathrm{nom}}$) case (*der*) which indicates that this noun phrase (NP; *der Affe* (the monkey)) is the subject of the verb (*fängt*). In contrast, the article in the accusative ($_{\mathrm{acc}}$) case (*den*) indicates that this NP (*den Frosch*) is the object of the verb. As this type of case marking makes the role of the NP explicit, it renders word order less important and both the SVO and OVS have the same meaning (the monkey chases the frog) despite the difference in word order.

However, in German, nominative and accusative cases are only unambiguously marked for nouns with masculine gender (der$_{nom}$ vs. den$_{acc}$). In contrast, for nouns with feminine or neuter gender, the article does not disambiguate these two cases; their nominative form is identical to their accusative form. In Table 3.2 we see both SVO and OVS constructions, which differ in whether the first NP is ambiguous or unambiguous in its role in the event depending on the article.

Table 3.2: Example sentences with case marking and word order manipulation

	Unambiguous NP1	Ambiguous NP1
SVO	**Der Uhu** kitzelt gleich das Meerschwein	Das Meerschwein kitzelt gleich **den Uhu**
	The owl $_{nom}$ tickles the guinea pig$_{acc}$	The guinea pig$_{nom}$ tickles the owl$_{acc}$
	The owl tickles the guinea pig	The owl tickles the guinea pig
OVS	**Den Uhu** kitzelt gleich das Meerschwein	Das Meerschwein kitzelt gleich **der Uhu**
	The owl$_{acc}$ tickles the guinea pig$_{nom}$	The guinea pig$_{acc}$ tickles the owl$_{nom}$
	The guinea pig tickles the owl	The guinea pig tickles the owl

In the unambiguous constructions, the first NP is unambiguously case marked with the masculine article. This article is nominative (*der*) in the SVO sentence, revealing SVO word order at the onset. Similarly, for the unambiguous OVS construction, OVS order is revealed at the onset of the first NP, as (*den*) is the unambiguous accusative form. In contrast, in the ambiguous constructions, the article of first NP takes the same form whether nominative and accusative (neuter form: *das*; feminine: *die*). However, in these sentences, upon reaching the second NP (*den Uhu*), the roles of the NPs become clear, given that the masculine form is unambiguously marked (as accusative for SVO, and nominative for OVS).

While German has this relatively free word order, comprehenders still have the preference to interpret the first NP as the subject of the sentence. This preference has been established using different experimental paradigms: self-paced reading (Schlesewsky, Fanselow, Kliegl & Krems, 2000; Schriefers, Friederici, & Kühn, 1995), speeded grammaticality judgments (Bader & Meng, 1999), event related potentials (ERP; Friederici, Hahne, & Mecklinger, 1996; Friederici, Steinhauer, Mecklinger, & Meyer, 1998; Frisch, Schlesewsky, Saddy, & Alpermann, 2002; Matzke, Mai, Nager, Rsseler, & Münte, 2002; Rösler, Pechmann, Streb, Röder,& Hennighausen, 1998; Schipke, 2012), and functional magnetic resonance imaging (Bahlman, Rodriguez-Fornells, Rotte, & Münte, 2007). For example, Schlesewsky et al. (2000) reported 4 self-paced reading studies and 2 offline studies investigating ambiguous first NP constructions. The data from these 6 studies showed that there was a preference for a SVO interpretation with ambiguous constructions, additionally the authors found that when the parser had to reanalyze to an OVS interpretation (e.g. when encountering an ambiguous first NP OVS construction) it was a costly endeavour in terms of reading times.

Bader and Meng (1999) found similar results using a grammaticality judgment task (participants were presented a sentence word by word and then asked to judge the grammaticality of the sentence as quickly as possible) to investigate the processing of four types of subject/object ambiguous sentences (NP-scrambled sentences, pronoun movement sentences, relative clause sentences, and embedded questions). They found that OVS sentences were judged less accurately than SVO sentences, and that the reaction time for correctly answered trials was greater with OVS sentences than for SVO sentences across all four sentence types - indicating a SVO interpretation bias.

3.1.1 ERP Research

ERP research has also shown differences in brain activity when processing non-canonical constructions compared to the preferred canonical constructions in the form of a Left Anterior Negativity (LAN) (e.g. Matzke et al., 2002; Rösler, et al., 1998; Schlesewsky, Bornkessel, & Frisch, 2003). The LAN is an event related potential that occurs approximately 300 ms after stimulus onset and is believed to index the processing costs involved in holding a filler in working memory, and the working memory requirements for assigning thematic roles (Kluender & Kutas, 1993a; 1993b). Rösler et al. (1998) found that when a sentence began with an unambiguous object, ERP patterns at the second NP showed a more negative amplitude (at 300-450 ms) compared to sentences that began with an unambiguous subject. The authors argued this was similar to a LAN component, and indicative of an increased load on working memory. Rösler et al. also found when participants were given a comprehension question (probing who was doing what to whom), they needed less time to answer for SVO constructions compared to OVS constructions.

Another study also found more negative LAN components for unambiguous OVS sentences compared to unambiguous SVO sentences, however this was at both the first NP and at the second NP (Matzke et al., 2002). Matzke et al. attributed this LAN at the first NP to the need to maintain the NP in working memory in the (non-canonical) OVS sentences; the LAN was only seen for unambiguous OVS structures suggesting that the parser took an SVO interpretation as default, in line with research suggesting a SVO preference in German. The authors speculated that the LAN at the second NP in the unambiguous OVS constructions might be related to the working memory involved in storage and retrieval of the object, prior to integration in the event representation.

Additionally, for ambiguous OVS sentences, Matzke et al. found a greater positivity during the P600 time window at the disambiguating second NP compared to unambiguous OVS sentences. The P600 is an event related potential that occurs approximately 600 seconds after stimulus onset with a late positive-going deflection. It is believed to index analysis/reanalysis of syntactic information (Osterhout et al., 1994). The authors interpreted this P600 as an indicator that the parser must revise the initial (preferred) SVO interpretation.

Additionally, ERP research has shown a P600 (e.g. Friederici et al., 1998; Frisch, Schleswsky, Saddy, & Alpermann, 2002; Matzke et al., 2002) when the parser reached the unambiguous second NP in OVS constructions where the first NP was ambiguous. For example, positivity at the second NP (after an ambiguous first NP) has been found when that construction proved to be OVS compared to when the ambiguous construction proved to be in the preferred SVO order (Frisch, et al., 2002) during the P600 time window. These authors argued that the P600 indicates the parser's need to revise the initial, and preferred, SVO interpretation into the OVS word order when the 2nd NP unambiguously confirms the correct interpretation.

As noted above, Rösler et al. (1998) and Matzke et al. (2002) argued that the LAN for the OVS sentences was related to a working memory load associated with non-canonical structures. However, Schlesewsky, Bornkessel, and Frisch (2003) argued this seemed unlikely and that the LAN was more linguistic in nature. They suggested that the LAN arises when there is a mismatch of expectations. In other words, when an object precedes a subject, it is a violation of the expectation of the SVO canonical order. Schlesewksy et al. attempted to tease these two ideas (violation of canonicity and working memory constraints) apart by using constructions with an added object pronoun before the subject as seen inExample (1).

1. Dann hat ihm der Leher den Roman Geben.
   Then has him$_{\text{dat}}$ the$_{\text{nom}}$ teacher the$_{\text{acc}}$ novel given.
   Then the teacher gave him the novel.

Pronouns in German can be moved in front of a subject without difficulty and without giving rise to a non-canonical construction. Schlesewksy et al. argued that this phenomenon allowed them to test these two competing theories as the moved object pronoun must be held in working memory but as the sentence is still canonical there is no violation of canonicity expectation. Hence, if a LAN was evoked by the moved object pronoun then it could be interpreted as working memory related (holding the displaced element), if no such LAN was found then the effects found in previous research may instead represent a canonicity violation. This was the pattern observed: there was no LAN for the object pronoun when it was moved to a position in front of the subject, supporting the position that the LAN response was a result of a violation of canonicity expectations.

3.1.2 Pupillometry Research

Given the robust findings that there is more difficulty associated with OVS word order in German, in the current study we are interested in measuring the processing costs associated with different constructions in German using a different methodology; a psychophysiological measure known as pupillometry. This aimed to build on the research of Schlesewksy et al. and further test the competing theories explaining this difficulty: violation of canonicity based and working memory based theories.

Pupillometry involves measuring the change in the diameter of the pupil (the small opening in the centre of the eye). It has been shown to be an effective measure of cognitive processing load following the seminal work of researchers such as Eckhard Hess and Daniel Kahneman, (e.g. Hess & Polt, 1960; Kahneman & Beatty, 1966). Pupillometry has also been used with a variety of linguistic tasks such as: sentence and word recall (Kahneman, 1973; Kahneman & Wright, 1971), letter encoding (Beatty & Wagoner, 1978), syntactic ambiguity resolution (Ben-Nun, 1986; Engelhardt, Ferreira, & Patsenko, 2009; Gutirrez & Shapiro, 2010; Just & Carpenter, 1993; Piquado, Isaacowitz, & Wingfield, 2010; Schluroff, 1981; Schluroff et al., 1986), discourse (Zellin, Pannenkamp, Toepel, & van der Meer, 2011), translation

(Hyönä, Tommola, & Alaja, 1995), and prosody and syntactic ambiguity resolution (Engelhardt, et al., 2009). Importantly for the current study, Kahneman and Beatty (1986) found that pupil change reflected working memory demands: when participants were asked to hold and recall digits in working memory, pupil diameter increased with each additional digit and similarly pupil diameter decreased as each digit was recalled. Just and Carpenter (1993) and Fernandez (2013) used pupillometry as a measure of working memory and found that the pupil was sensitive to the processing costs associated with non-canonical constructions in English.

Pupillometry has been shown to have a similar sensitivity to that of Event Related Potentials (ERPs) and to correlate with underlying brain activity (Friedman, Harkerem, Sutton, & Fleiss, 1973; Gutirrez & Shapiro, 2010; Kuipers & Thierry, 2011), but it is still unclear how the brain modulates pupil dilation. It is possible that there is an indirect link between pupil change and locus coeruleus (LC); LC is a neuromodulatory brain system that is the only source of noradrenaline or norepinephrine (NE) in the forebrain. It is believed to play an important role when it comes to behavioural performance and shifting of attention (Sara, 2009). In her review Sara writes that previous neuroimaging studies suggest, ”an essential role of the LC noradrenaline system is to promote or even orchestrate dynamic interactions among networks involved in cognition” (pg 221). While the breadth of this topic is outside the scope of this chapter, there may be a relationship between the LC-NE system and pupil change (see for example, Aston-Jones & Cohen, 2005), suggesting that phasic pupil change may be an indirect measure of LC-NE activity (and therefore of cognitive activity, e.g, Hess & Polt, 1964; Kahneman & Beatty, 1966).

The only pupillometry study that has investigated German word order, to my knowledge, was conducted by Demberg (2013); the study was not a linguistic study, per se, rather it was testing the usefulness of the Index of Cognitive Activity (ICA) using a dual task paradigm. The ICA is a micro-level pupillometry analysis that counts the frequency of small rapid pupil dilations which are believed to be less susceptible to lighting and eye-movements than pupil dilation at a macro level. Participants heard locally ambiguous subject relative clauses (canonical) (see Example 2) and object relative clauses (non-canonical) (see Example 3) in German while performing a driving task, and were asked questions about what they had heard.

2. Die Nachbarin, die einige der Mieter auf Schadensersatz verklagt hat, traf sich gestern mit Angelika.
   The neighbor, who sued some of the tenants for damages, met Angelika yesterday.
3. Die Nachbarin, die einige der Mieter auf Schadensersatz verklagt haben, traf sich gestern mit Angelika.
   The neighbour, whom some of the tenants sued for damages, met Angelika yesterday.

In these examples word order is ambiguous until reading the finite verb (hat/haben) of the relative clause.

Demberg investigated the ICA and overall pupil dilation (the change in pupil diameter size) during this dual task. When comparing the overall pupil diameter measurement and the ICA to each other, she found overall differences between the two methods only when it came to the driving task. Thus it also appears that overall pupil dilation and ICA provide similar information on the processing of linguistic data (particularly canonical and non-canonical sentences). In the linguistic task, Demberg found that both overall pupil dilation and the ICA showed that the pupil decreased more quickly during the canonical subject relative constructions compared to the non-canonical object relative constructions. These data suggest the parser invests more cognitive resources as the sentence unfolds during the processing of non-canonical sentences compared to the canonical sentences regardless of analysis type.

In sum, previous research has shown differences in processing between canonical and non-canonical constructions and differences in processing at an ambiguous first NP, and at the 2nd disambiguating NP in German with different methodologies. However only one study employed a pupillometry paradigm with these type of constructions, but in the context of a dual task. In this study we use pupillometry as a means to investigate these differences with the aim of establishing pupillometry as an effective tool to measure the processing associated with different word orders. By establishing pupillometry as an effective way to test word order it broadens the methodological spectrum, and pupillometry will be particularly useful with vulnerable populations (e.g. children and patients). Given that there is quite a large body of research investigating word order using ERP, and it is still unclear what the pupil reflects, testing these constructions using pupillometry may help us better understand pupil change in the context of brain activity. Additionally, we aim to test whether the difficulty associated with the processing of the first NP in unambiguously marked OVS constructions is due to working memory demands or canonicity violations.

3.1.3 Study Aims

Pupillometry will be used to investigate sentences similar to those in Table 3.2 (earlier). There are three relevant measures: pupil change slope (pupil diameter), peak pupil amplitude, and peak pupil latency. Pupil change slope is a dynamic measure of processing costs, while peak pupil amplitude and latency provide information about which conditions evoke the largest processing costs, and whether more time was needed to reach this peak of processing. We are interested in comparing SVO and OVS constructions focusing on processing at the first (ambiguous or unambiguous) NP, the second (unambiguous) NP, and across the whole sentence.

Given that pupillometry has been shown to be a measure of processing costs, and correlate with brain activity, if the LAN produced by an unambiguous first NP in an OVS construction is due to working memory (Matzke et al., 2002; Rösler, et al., 1998) we hypothesise this should be reflected in pupil change slope. Alternatively, if the LAN is due to canonicity expectation violation (Schlesewsky, et al., 2003) then no differences are predicted in pupil change slope between the unambiguous OVS and SVO sentences. The peak amplitude and latency will also reflect these hypotheses respectively.

When encountering the unambiguous second NP in the ambiguous SVO condition, it becomes apparent that the construction falls in line with the preferred SVO word order in German; in contrast, upon reaching the second NP in the ambiguous OVS constructions it becomes apparent that the sentence is in the OVS order and the parser will have to revise the original (preferred or default) SVO interpretation. Therefore, if at the second NP, the parser is revising the original interpretation, the cognitive effort involved should result in a larger change in pupil diameter for ambiguous OVS constructions compared to unambiguous SVO constructions (where order is apparent from onset) and the ambiguous SVO (where order is not apparent at the onset, but the second NP confirms the preferred interpretation of SVO order). This should also be reflected in peak amplitude, and peak latency.

Across the whole sentence a larger pupil slope, peak amplitude, and latency is predicted for the OVS constructions compared to the SVO given that the OVS constructions are more difficult to process than SVO constructions. Additionally, ambiguous constructions will evoke more processing costs than unambiguous constructions, given that the parser must hold the NPs in working memory until thematic roles can be assigned, and this will also be reflected with an increased pupil slope, larger peak amplitude, and latency.

3.2 Methods

3.2.1 Participants

Twenty-one students (16 female), recruited from University of Potsdam, participated for payment or as requirement for their undergraduate course. All participants were native speakers of German and had normal or corrected to normal vision.

3.2.2 Materials

Four sentence types were used in this experiment, as seen in Table 3.2 (earlier). There were a total of 24 critical trials in each condition. The stimuli were constructed in a 2x2 design, manipulating word order (Subject-Verb-Object vs Object-Verb-Subject), and ambiguity (ambiguous or unambiguous article of the first NP). The unambiguous stimuli have the first NP in the construction unambiguously marked with a masculine article, making word order apparent from the onset. The ambiguous constructions have an NP with an ambiguously marked feminine or neuter article, and a second NP with an unambiguously marked masculine article, making order apparent only upon reaching the second NP.

The critical sentences were simple transitive sentences containing two NPs and one verb. The referents of the NPs were always two different animals, and the verb was either *kitzeln* (to tickle) or *fangen* (to chase). The animals for one sentence were chosen in a way that they were equally (un)likely to engage in the action described in the sentence so that no plausibility information could influence sentence interpretation (e.g. predator chasing prey). The sentences were recorded by a female native speaker of German at a normal speaking rate in the phonetics lab at the University of Potsdam with a sampling rate of 22050 hz and saved as .wav files.

Participants heard a sentence, and subsequently were presented with an image depicting the action representing the verb in the sentence. This image was either correct or showed the reverse thematic roles. The images used in this study were still frames extracted from animated videos (used in Chapter 4). The images for the verb *kitzeln* depicted one animal tickling the other with a feather; the images for *fangen* depicted one animal in motion towards the other animal that appeared still. In half of the trials (n=12) the auditory sentence (e.g. The owl tickles the guinea pig) matched the image presented to the participant (i.e. participants would see an owl ticking a guinea pig), and in half of the trials (n=12) the subject and object role were reversed in the image (i.e. participants would see a guinea pig tickling an owl). This was counterbalanced and randomized across versions; each animal was only used once. The recordings and corresponding images were placed into 4 versions and rotated in a Latin square design. Each participant heard 24 trials, 10 with the verb tickle and 14 with the verb chase; 6 trials in each of the conditions.

3.2.3 Apparatus

Stimulus presentation was programmed using Tobii Studio software. Pupil diameter was recorded with a Tobii T120, sampling at 120 Hz. Tracking was binocular but only the right eye was used for analysis. The eye tracker was built into a 17-inch thin film transistor liquid crystal colour display monitor (1,280 x 1,024 pixel resolution). Participants sat approximately 60-65 cm away from the display and tracking was remote. Eyes were calibrated using a 9-point sequence.

3.2.4 Procedure

In each trial a fixation cross appeared on the screen for 2000ms prior to stimulus onset, for the duration of the stimulus, and for an additionally 2000ms post stimulus offset. The 2000ms prior to onset was to allow the pupil to adjust to screen luminance. Participants were asked to focus on the fixation cross when it was on the screen, attend to the aurally presented sentence, and to try to avoid blinking. We used this method (a blank screen with an auditory presentation) to ensure eye movements did not affect the pupil response. After the offset of the auditory stimuli the fixation cross remained on the screen for an additional 2000ms (to enable recording of any potential wrap up effects). Then the fixation cross was replaced with instructions that a picture would be presented and the participant should judge whether the image matched the sentence they had just heard, by pressing ”J” on the keyboard for ”Ja” (yes) if the recording and the image matched or ”N” for ”Nein” (no) if the recording and picture did not match. Participants were instructed to press the space bar when ready to see the picture, giving them the opportunity to rest their eyes after not blinking for the duration of the trial. Images were presented as comprehension probes so that the methods could be replicated with children (Chapter 4).

3.2.5 Analysis

Blinks were filtered out by replacing the missing values with linear interpolation, and the pupil was normalized by item for each participant, so all items started at zero. We analysed pupil change slope across the entirety of the sentence as well as two other critical time windows, beginning at the first NP and beginning at the second NP. A 1.2 second window starting at the onset of each NP was investigated as 1.2 seconds is the time it takes for the pupil to reach its maximum diameter (Just & Carpenter, 1993). Before analysis, data was screened for outliers; any data points three or more standard deviations from the mean were replaced with the mean of that condition. All trials were averaged together within each condition resulting in 4 vectors for each participant (one for each of the 4 conditions). These vectors were submitted to a simple regression where pupil size served as the dependent variable and time served as the independent variable. The slope of pupil diameter change over time (i.e. the unstandardized regression coefficient) was the main dependent variable in this study. We also calculated the peak pupil amplitude (the largest size that the pupil reached), and the peak pupil latency (the time that the peak pupil size was reached) for each vector; these are typical measures in pupillometry research (e.g. Schmidtke, 2014; Zekveld, Kramer, & Festen, 2010).

As the critical sentences varied slightly in terms of length (mean: 4.35 sec; sd: 0.25), for the analysis across the whole sentence, the constructions were cut so the sentences all had the same length. All sentences were cut to the length of the shortest sentence (3.83 seconds); this only affected the analysis across the whole sentence, given that the analyses across the two NPs were over a set time window of 1.2 seconds. Comprehension accuracy, the unstandardized regression coefficient (slope of pupil diameter change), peak amplitude, and peak latency were submitted to linear mixed effect models using R (R Core Team, 2012), the *lme4* package (Bates, Maechler, & Bolker, 2012), and Tukey's contrasts from the *multcomp* package (Horthorn et al, 2013). The linear mixed effect models included fixed effects of word order (SVO/OVS) and ambiguity (ambiguous/unambiguous), and random effects of participants; the random effect of participant was maximally specified with random slopes for word order, and ambiguity. Models were compared such that a full model was contrasted with a reduced model to test the factor in question; nonsignificant factors were excluded from the model by contrasting each main effect and interaction, starting with word order, followed by ambiguity, and finally the interactions. Model comparison was based on log likelihood ratio tests.

3.3 Results

3.3.1 Accuracy of Sentence-Picture Matching

Figure 3.1 shows the mean percentages of correct judgments (including correct rejections as well as correct acceptance) of the participants.

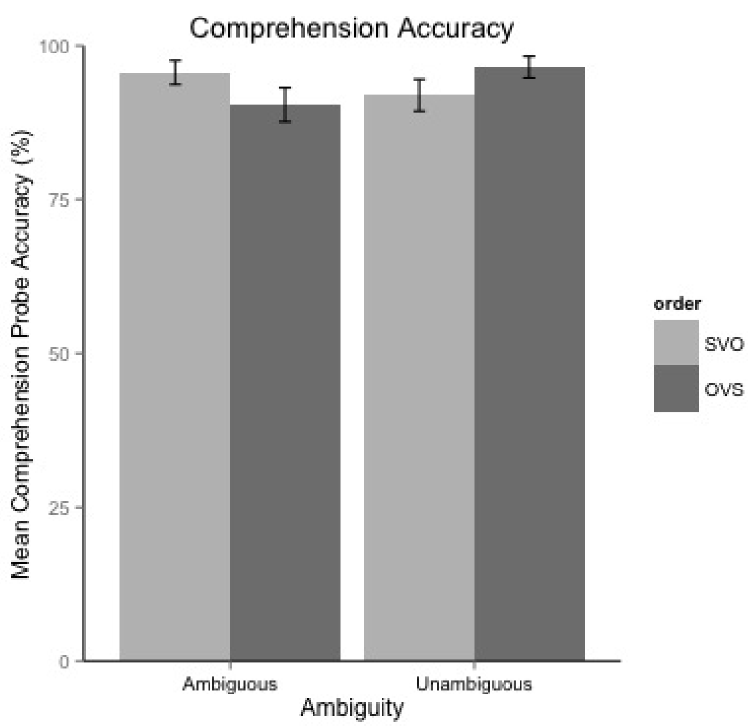

Figure 3.1: Sentence-picture matching comprehension accuracy (the error bar represents the standard error of the mean).

Performance was close to ceiling (unambiguous OVS= 96% correct, unambiguous SVO = 91%, ambiguous OVS = 90%, ambiguous SVO = 95%) and linear mixed effect modelling revealed that there was no significant effect of stimulus type on accuracy ($X^2(3)= 5.36$, $p=0.15$).

3.3.2 First NP

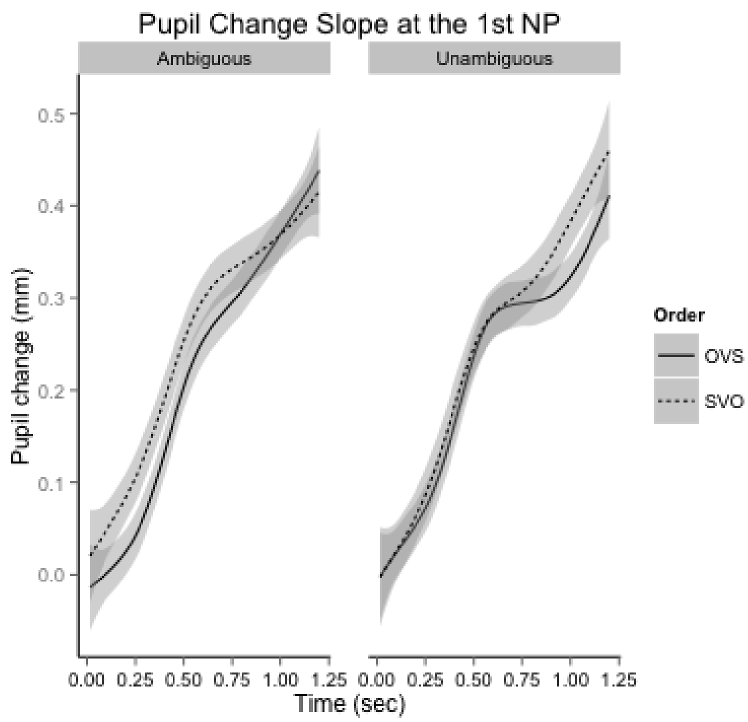

Figure 3.2: Pupil change from the onset of the first NP: Smoothed pupil change slope across time, the grey highlighting indicates standard error.

There were no significant effects of word order or ambiguity observed on pupil response at the first NP ($X^2(2)=0.91$, $p=0.63$), nor a significant interaction ($X^2(2)=1.66$, $p=0.64$), refer to Figure 3.2.

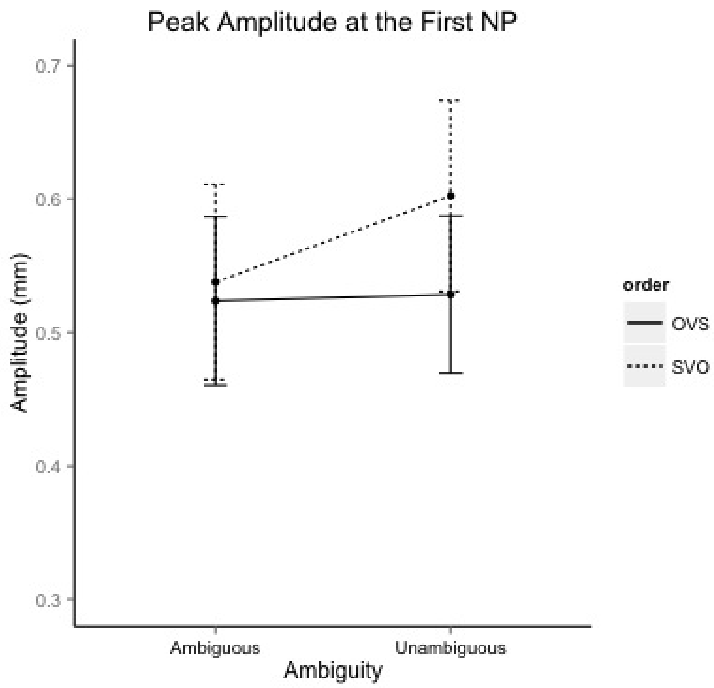

Figure 3.3: Peak amplitude from the onset of the First NP (bars represent standard error).

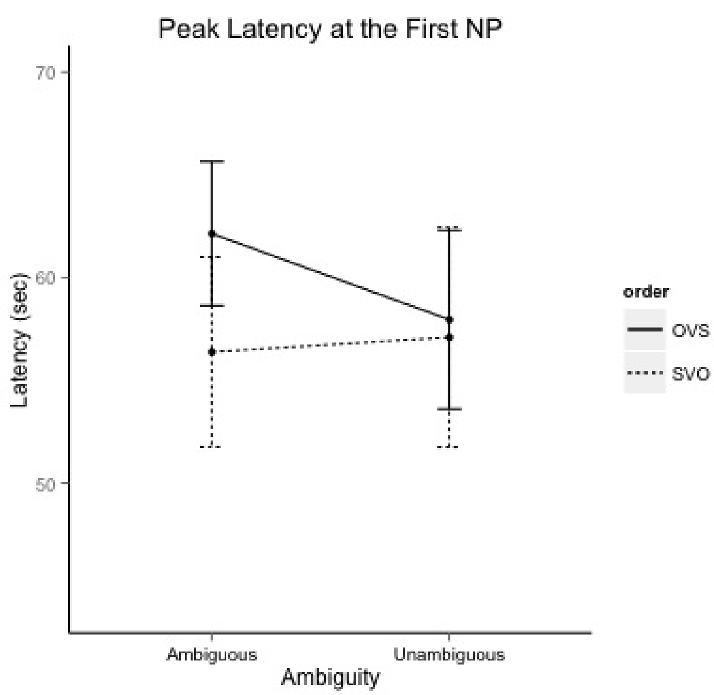

Figure 3.4: Peak latency from the onset of the first NP (bars represent standard error).

There were no main effects on peak amplitude ($X^2(2)$=0.83, p=0.65), nor a significant interaction ($X^2(3)$=1.19, p=0.7), refer to Figure 3.3. There were also no main effects on peak latency :($X^2(2)$=1.11, p=0.57), nor an (interaction $X^2(3)$=1.43, p=0.69), refer to Figure 3.4.

3.3.3 Second NP

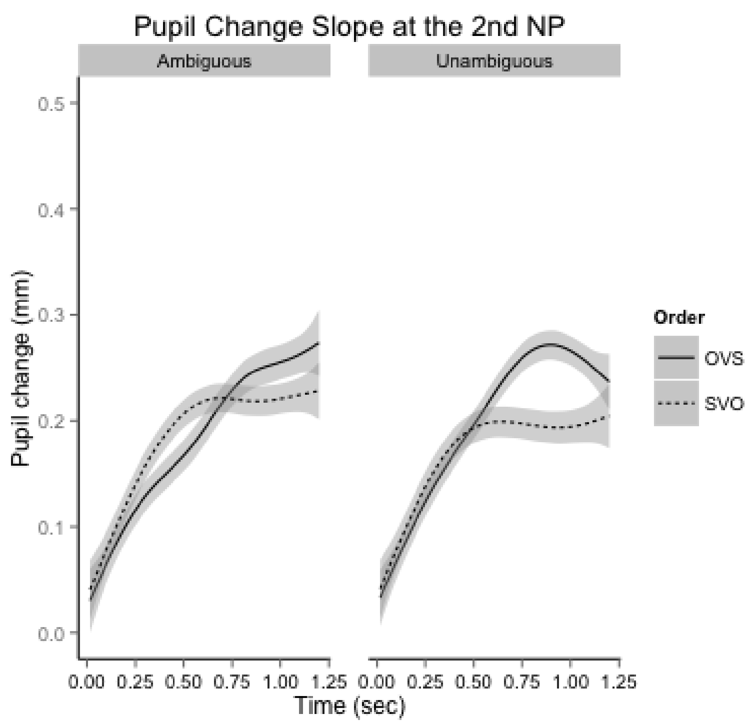

Figure 3.5: Pupil change from the onset of the second NP: Smoothed pupil change slope across time, the grey highlighting indicates standard error.

At the second NP there were no main effects on pupil change slope ($X^2(2)$= 1.03, p=0.59). However, there was an interaction between ambiguity and word order ($X^2(1)$= 4.12, p=0.04): the unambiguous SVO condition evoked a shallower pupil change slope than the ambiguous SVO, unambiguous OVS , and ambiguous OVS conditions (Tukey's contrasts: all *p's* <0.001), see Figure 3.5.

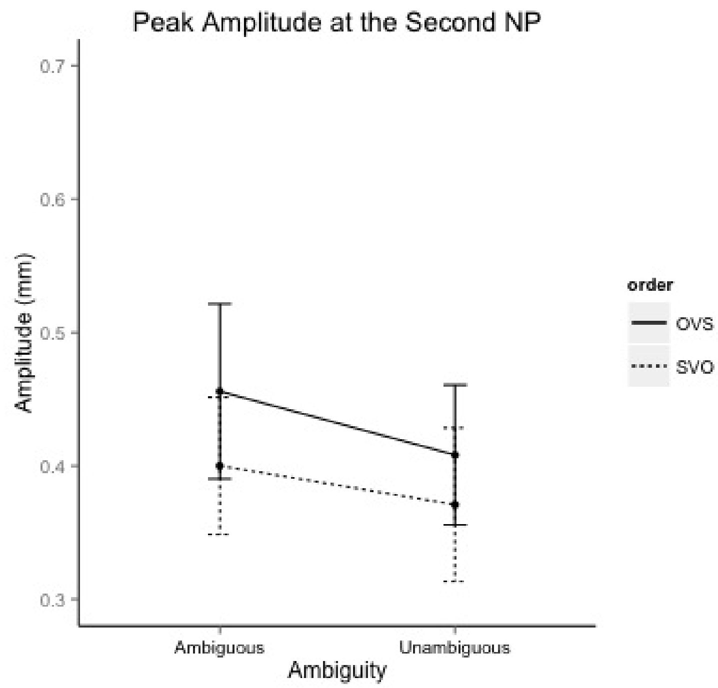

Figure 3.6: Peak amplitude from the onset of the second NP (bars represent standard error)

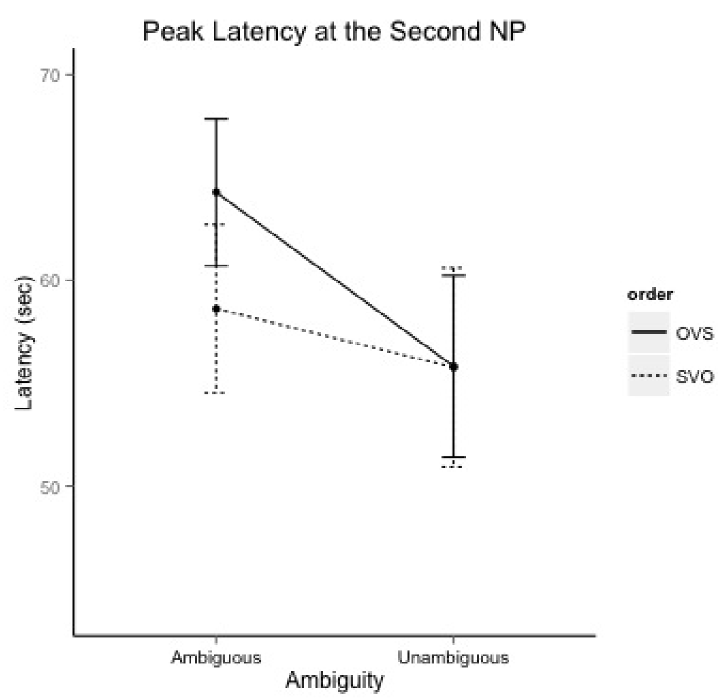

Figure 3.7: Peak latency at the second NP (bars represent standard error).

There were no significant effects on raw peak amplitude (see Figure 3.6; main effects $X^2(2)= 2.40$, p=0.30, interaction $X^2(3)= 2.88$, p=0.40, or on the raw peak latency (see Figure 3.7; main effects $X^2(2)= 2.01$, p=0.37, interaction $X^2(3)= 2.08$, p=0.56).

3.3.4 Across the Whole Sentence

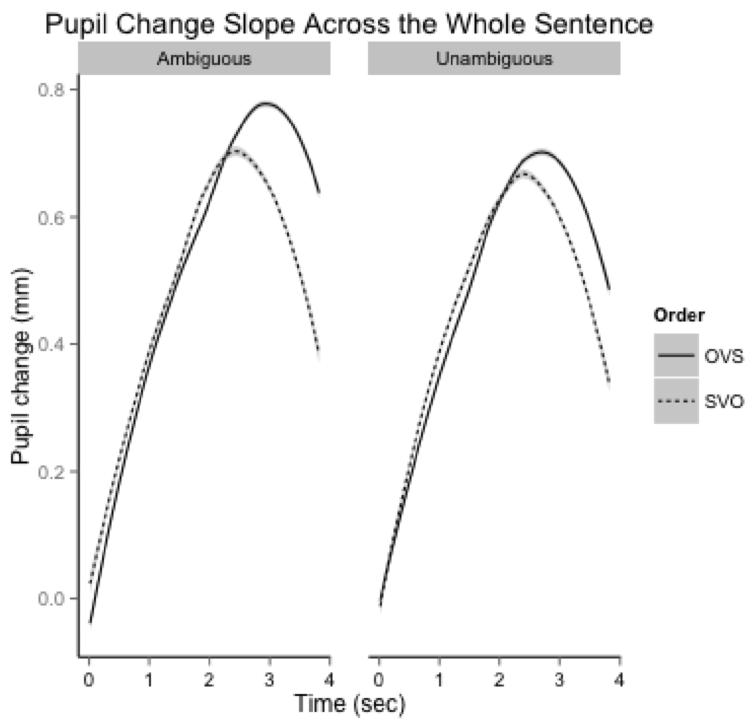

Figure 3.8: Pupil change across the whole sentence: Smoothed pupil change slope across time, the grey highlighting indicates standard error.

Across the whole sentence there was a main effect of word order ($X^2(1)=$ 9.51, p=0.002), with the OVS condition evoking a steeper pupil change slope than the SVO condition (see Figure 3.8). The main effect of ambiguity approached significance ($X^2(1)= 2.85$, p=0.09) - ambiguous sentences tended to evoke greater pupil change slope than unambiguous sentences. There was no interaction ($X^2(2)= 1.23$, p=0.27).

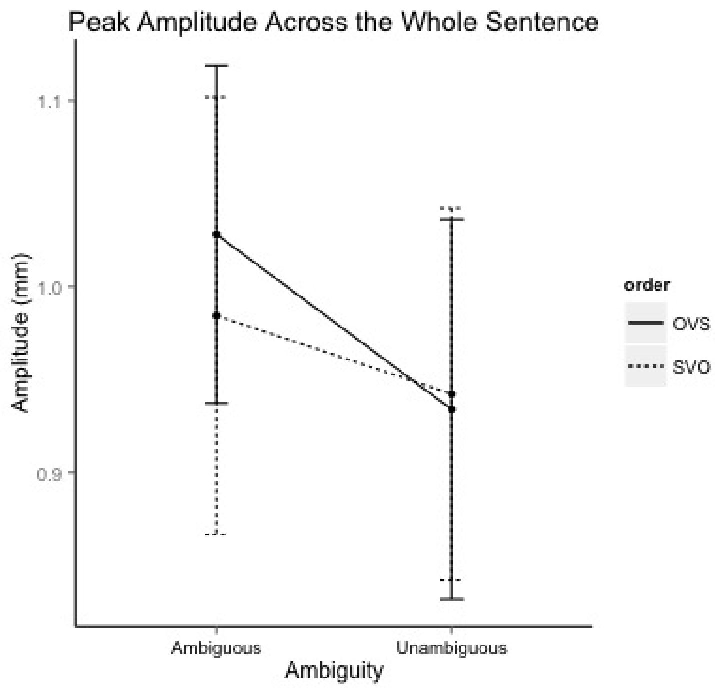

Figure 3.9: Peak amplitude across the whole sentence (bars represent standard error)

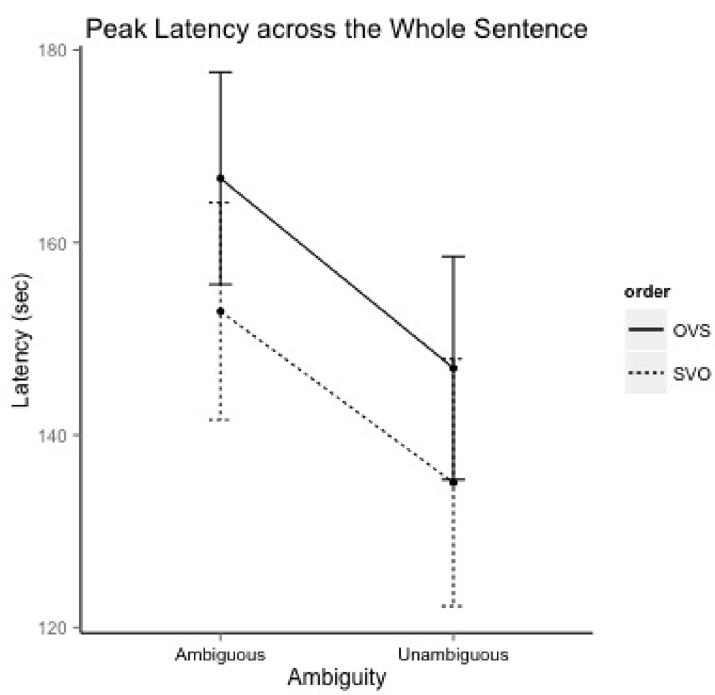

Figure 3.10: Peak latency across the whole sentence (bars represent standard error).

Raw peak amplitude across the whole sentence also revealed a main effect of ambiguity ($X^2(1)= 3.71$, $p=0.05$; see Figure 3.9): with the ambiguous conditions evoking a larger peak diameter than the unambiguous conditions. There was no main effect of word order ($X^2(1)= 1.72$, $p=0.19$) and no interaction ($X^2(2)=1.83$, $p=0.39$). Raw peak latency across the whole sentence revealed no main effects ($X^2(2)= 0.93$, $p=0.62$; see Figure 3.10) and no interaction ($X^2(2)= 1.06$ $p=0.78$).

3.4 Discussion

In this study we investigated German adults' processing of auditory sentences by tracking participant's pupil dilation. Sentences either comprised an unambiguously case marked first NP, which allowed for an immediate thematic role assignment, or an ambiguously case marked first NP where role assignment was only possible later when the unambiguously case marked second NP was processed. Word order was also manipulated. German has a relatively free word order, but there is a preference for the SVO construction (e.g. Bader & Meng, 1999; Matzke, et al., 2002; Rösler, et al., 1998; Schipke, 2012; Schlesewsky, et al, 2000; Schriefers, et al, 1995). This study used pupil change slope as an indicator of the processing costs associated with sentence comprehension.

Participants comprehended the different constructions with high accuracy and no difference between conditions. Analysis of the pupil response at the first NP found no differences between any of the conditions in terms of pupil change slope, peak pupil size, or peak pupil latency. As mentioned in the introduction, previous research using ERP had found a LAN at the first NP for unambiguous OVS constructions relative to unambiguous SVO constructions, and some researchers attributed this LAN to index the working memory costs evoked by holding the object in memory until reaching the verb (e.g. Matzke et al., 2002; Rösler, et al., 1998). As pupillometry is also argued to be a measure of working memory costs (e.g. Hess & Polt, 1960; Kahneman & Beatty, 1966), the lack of different pupil responses across conditions seems to contradict this interpretation of the LAN response. Instead it supports, an alternative account where the LAN is attributed to a violation of canonicity (Schlesewsky, et al., 2003): the parser anticipates an SVO word order and upon encountering an unambiguous OVS construction a LAN is evoked that indexes this violation of the expectation.

At the second NP, while there were no differences in peak amplitude or peak latency, we found there was an interaction between word order and ambiguity, with the unambiguous SVO sentences showing a shallower pupil change slope than the other conditions (unambiguous OVS, ambiguous OVS, and ambiguous SVO). Previous ERP research has found, at the second NP, a P600 for ambiguous first NP SVO constructions, which was argued to signify syntactic revision (Matzke et al., 2002). Consistent with this, in this study, the ambiguous first NP OVS constructions evoked a steeper pupil change slope compared to the unambiguous SVO constructions. This supports an interpretation of syntactic revision: the parser initially interpreted these sentences as SVO and then, upon reaching the disambiguating second NP made a reanalysis to revise the initial interpretation.

ERP research also found a LAN at the second NP for unambiguous OVS constructions, which has been suggested to reflect working memory demands (Matzke et al., 2002). However, we found no difference between the ambiguous and unambiguous OVS conditions, suggesting, and like our pupillometry evidence at the first NP, that the difficulty associated with ambiguous OVS constructions is not due to additional working memory constraints. In addition, ambiguous first NP SVO constructions evoked a larger change in pupil diameter than the unambiguous SVO constructions; while this was not predicted, it suggests that there is a measurable cost to processing ambiguous SVO constructions despite the fact they fall in the preferred SVO order. This perhaps suggests that the parser somehow flags the 'unsubstantiated' nature of the parse, which is only resolved once an unambiguous article is encountered. This may also account for the difference found between the ambiguous constructions and unambiguous SVO at the second NP.

The pupil slope across the entirety of the sentence showed that OVS conditions evoked a steeper pupil change slope compared to the SVO conditions suggesting that overall, OVS sentences require more processing resources than canonical OVS constructions. This data is in line with previous research that has found a preference for SVO constructions in German (e.g. Bader & Meng, 1999; Bahlman et al., 2007; Friedrecici & Mecklinger, 1996; Friederici et al., 1998; Frisch et al., 2002; Matzke et al., 2002; Rösler et al., 1998; Schipke, 2012; Schlesewsky et al., 2000; Schriefers et al., 1995). There was also a main effect of ambiguity approaching significance, with sentences that were ambiguous at the first NP evoking a steeper pupil change slope than unambiguous sentences. Coupled with the peak amplitude being larger for the ambiguous conditions compared to the unambiguous conditions, this suggests that OVS constructions and constructions that are ambiguous until late in the sentence require more processing resources than unambiguous and SVO constructions.

Previous research has debated the implications of the LAN that occurs in response to unambiguous OVS constructions; some authors argue it signifies increased working memory processing (Matzke et al., 2002; Rösler, et al., 1998), while others argue it signifies a violation of an expectation of canonicity (Schlesewsky, et al., 2003). The results presented here support the latter account where there are processing costs as a result of violation of canonicity and upon reaching the second NP having to revise or confirm (in the case of the ambiguous SVO construction) the original interpretation.

3.5 Conclusion

This study has examined constraints on processing of different word orders in German. We found evidence supporting a position where unambiguous SVO constructions are processed with relative ease, while ambiguous constructions and OVS constructions require extra resources, but these extra resources (which seem to be due to revision and reanalysis) are not apparent (at least using pupillometry) until encountering the second NP, or looking at the processing of the whole sentence.

In addition, the results of this study suggest that more cognitive resources are required for the processing of constructions that do not fall into canonical SVO constructions. While this is a well-established finding in German across several methodologies, this is one of the first studies to find this using pupillometry. Our data suggests that the difficulties associated with unambiguously non-canonical (OVS) constructions may be the result of a violation of canonicity, while the difficulty associated with non-canonical constructions that are ambiguous at the first NP (ambiguous OVS) is the result of revision and reanalysis. This is in contrast to previous ERP research that has argued the difficulty associated with unambiguous OVS constructions is working memory based (Matzke et al., 2002).

This is the first study investigating word order in German purely using pupillometry, and more research is needed to validate these findings further.

Comparing ERP and pupil change to these constructions in the same experiment would be particularly interesting, and may provide more insight into the relationship between pupillometry and cognitive and brain mechanisms. However, by confirming that pupillometry is an effective tool to measure findings that have been well established with other methodologies we begin to build a greater foundation for extending the use of pupillometry to testing of more linguistic manipulations. This also opens the door to conduct cognitive research with vulnerable individuals using pupil change slope, which can be collected inexpensively and non-invasively with eye tracking devices.

3.6 References

Bader, M., & Meng, M. (1999). Subject-Object ambiguities in German embedded clauses: An across-the-board comparison. *Journal of Psycholinguistic Research, 28*, 121-143.

Bates, D.M., Maechler, M., & Bolker, B. (2012). *lme4: Linear mixed-effects models using S4 classes*. R package version 0.999999-0.

Bahlmann, J., Rodriguez-Fornells, A., Rotte, M., & Mnte, T.F. (2007). An fMRI study of canonical and noncanonical word order in German. *Human Brain Mapping, 28(10)*, 940-949.

Beatty, J. & Wagoner, & B.L. (1978). Pupillometric signs of brain activation vary with level of cognitive processing. *Science, 199*, 1216-1218.

Ben-Nun, Y. (1986). The use of pupillometry in the study of on-line verbal processing: Evidence for depths of processing. *Brain and Language, 28*, 1-11.

Demberg, V. (2013). Pupillometry: the Index of Cognitive Activity in a dual-task study In M. Knauff, M., Pauen, N., Sebanz, & I. Wachsmuth (Eds.) *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (p. 2154-2159). Austin TX: Cognitive Science Society.

Englehardt, P.E., Ferrerira, F., & Patsenko, E.G. (2010). Pupillometry reveals processing load during spoken language comprehension. *The Quarterly Journal of Experimental Psychology, 63*, 639-645.

Friederici, A. D., Hahne, A., & Mecklinger, A. (1996). Temporal structure of syntactic parsing: Early and late event-related brain potential effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 1219-1248.

Friederici, A.D., Steinhauer, K., Mecklinger, A., & Meyer, M. (1998). Working memory constraints on syntactic ambiguity resolution as revealed by electrical brain responses. *Biological Psychology, 47*, 193-221.

Friedman, D., Harkerem, G., Suttton, S., & Fleiss, J (1973). Effect of uncertainty on pupillary dilation response and the vertex evoked potential. *Electroencephalography and Clinical Neurophysiology, 34*, 474-484.

Frisch, S., Schlesewsky, M., Saddy, D., & Alpermann, A. (2002). The P600 as an indicator of syntactic ambiguity. *Cognition, 85*, B83-92.

Gutirrez, R.S., & Shapiro, L.P. (2010.) *Measuring the Time-course of Sentence Processing with Pupillometry.* Poster presented at the 23rd annual meeting of the CUNY conference on human sentence processing. New York, New York.

Just, M.A., & Carpenter, P.A. (1993). The intensity dimension of thought: Pupillometric indices of sentence processing. *Theoretical Issues in Ergenomics Science, 47(2)*, 310-339.

Hess, E. H., & Polt, J. M. (1960). Pupil size as related to interest value of visual stimuli. *Science, 132*, 349-350.

Hothorn, T., Bretz, F., Westfall, P., Heilberger, R.M, & Schuetzenmeister, A. (2013). *Simultaneous Inference in General Parametric Models.* R Foundation for Statistical Computing Vienna, Austria.

Hyönä, J., Tommola, J., & Alaja, A-M. (1995). Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. *The Quarterly Journal of Experimental Psychology, 48A*, 598-612.

Kahneman, D. (1973). *Attention and Effort.* Englewood Cliffs, N.J.: Prentice-Hall.

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science, 154*, 1583-1585.

Kahneman, D., & Wright, P. (1971). Changes of pupil size and rehearsal strategies in a short-term memory task. *Quarterly Journal of Experimental Psychology, 23*, 187-196.

Kluender, R. and Kutas, M. (1993a). Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience, 5*, 196-214.

Kluender, R. and Kutas, M. (1993b). Subjacency as a processing phenomenon. *Language and Cognitive Processes, 8*, 573-633.

Kuipers, J.R., & Thierry, G. (2011). N400 amplitude reduction correlates with an increase in pupil size. *Frontiers in Human Neuroscience, 5*, 1-5.

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

Osterhout, L., & Holcomb, P. J., & Swinney, D. A. (1994).Brain potentials elicited by garden path sentences: Evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20*, 786-803.

Piquado, T., Isaacowitz, D., & Wingfield, A. (2010). Pupillometry as a measure of cognitive effort in younger and older adults. *Psychophysiology, 47*, 560-569.

R Core Team (2012). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria.

Rösler, F., Pechmann, T., Streb, J., Röder, B., & Hennighausen, E. (1998). Parsing of sentences in a language with varying word order: Word-by-word variations of processing demands are revealed by event-related brain potentials. *Journal of Memory and Language, 38*, 150-176.

Schipke, C. S. (2012). *Processing mechanisms of argument structure and case-marking in child development: Neural correlates and behavioral evidence* (Unpublished doctoral dissertation). Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig.

Schlesewsky, M., Bornkessel, I., & Frisch, S (2003). The neurophysicological basis of word order in German. *Brain & Language, 86*, 116-128.

Schlesewsky, M., Fanselow, G., Kliegl, R. & Krems, J. (2000). Preferences for grammatical functions in the processing of locally ambiguous wh-questions in German. In Hemforth, B. & Konieczny, L. (eds.) *Cognitive Parsing in German.* Dortrecht: Kluwer.

Schluroff, M. (1981). Pupil responses to grammatical complexity of sentences. *Brain and Language, 17*, 133-145.

Schluroff, M., Zimmerman, T.E., Freeman, R.B., Hofmeister, K., Lorscheid, T., & Weber, A. (1986). Pupillary responses to syntactic ambiguity of sentences. *Brain and Language, 27*, 322-344.

Schmidtke, J. (2014). Second language experience modulates word retrieval effort in bilinguals: evidence from pupillometry. *Frontiers in Psychology, 5*, 137.

Schriefers, H., Friederici, A.D., & Kühn K. (1995). The processing of locally ambiguous relative clauses in German. *Journal of Memory and Language, 34*, 227-246.

Zellin, M., Pannekamp, A., Toepel, U., & van der Meer (2011). In the eye of the listener: Pupil dialation elucidates discourse processing. *International Journal of Psychophysiology, 81*, 133-141.

Zekveld, A.A., Kramer, S.E., Festen, J.M. (2010). Pupil response as an indication of effortful listening: The influence of sentence intelligibility. *Ear and Hearing, 31*, 480-490.

Chapter 4

In Chapter 3 the processing of German word order was investigated using pupillometry. We found that pupillometry was an effective tool with which to measure the processing costs associated with different word orders in German, and in doing so broadened the methodological scope of pupillometry. In this chapter we will be using the same stimuli to test the processing of these constructions with native German speaking children, in aiming to provide evidence for processing differences between canonical subject-first and non-canonical object-first constructions.

Investigating the Processing of German Word Order in Children

Abstract

German has a relatively free word order with both subject-first and object-first constructions being possible. German also employs case marking to delineate the roles of each noun phrase. Despite this, there remains a preference for subject-first constructions, and object-first constructions cause children comprehension difficulty until approximately 7 years of age. There is evidence that younger German children mainly rely on word order during comprehension. In this study we manipulated word order and case marking to investigate sentence comprehension in a group of 4-year-old German speaking children, using a novel methodology, pupillometry, as a measure of cognitive cost. Participants performed poorly on behavioral measures of comprehension, perhaps due to the time between sentence presentation and the comprehension question. Pupillometry found that subject-first constructions evoked the greatest pupil change, indicating increased processing costs. This suggested that the children were investing their resources into those constructions that were most familiar to them. Direct comparison of pupil responses to these stimuli in adults and children showed that adults and children were displaying different patterns when processing at the first NP (for the two ambiguous conditions), at the second NP (for the two ambiguous and the unambiguous OVS condition), and across the whole sentence (for all but the unambiguous SVO construction). These data seem to suggest, again, that children were investing their resources into those constructions they could most efficiently comprehend.

These data support previous research suggesting that 4-year-old children mostly rely on word order during sentence processing, but that they can detect case marking cues, even if they don't process them in the same way as adults.

4.1 Introduction

Languages differ in the cues that help determine who is doing what to whom in a sentence; for example, some languages rely more on word order cues (e.g. English) while others rely more on case marking cues (e.g. German): grammatical cues that provide information about the roles of the noun phrases in the sentence (i.e. marking case such as nominative, accusative, etc). English typically uses a subject-verb-object (SVO) word order in transitive sentences, in which the subject (or agent) of the sentence occurs before the verb, and the object (or patient) occurs after the verb. German, also typically uses SVO word order, but, given the presence of case marking, there is more flexibility to use different word orders. This case marking typically occurs on the article, with the article of the subject noun phrase being marked with the nominative case and the object being marked with accusative case. Case making, however, can be ambiguous because only the masculine article changes from nominative to accusative (*der* vs. *den*), while the feminine and neuter articles have the same form for both cases. Consequently, in cases with feminine and neuter articles ambiguity can arise (at the first NP or across the whole sentence) given that the syntactic function of the NP is not explicitly marked.

When the first noun phrase (NP) is marked with a masculine article either in the nominative or the accusative case, it is then unambiguous whether the first NP is the subject (SVO; e.g. **Der Affe** fängt gleich die Maus [The $\text{monkey}_{\text{nom}}$ chases the $\text{mouse}_{\text{acc}}$] = The monkey chases the mouse) or the object (OVS; e.g. **Den Affen** fängt gleich die Maus [The $\text{monkey}_{\text{acc}}$ chases the $\text{mouse}_{\text{nom}}$]= The mouse chases the monkey) of the sentence. In contrast, when the first NP is feminine or neuter, the case marking is ambiguous, and it is not until encountering a second, masculine, NP that the role of the first NP becomes apparent (e.g. SVO: Die Maus fängt gleich **den Affen** [The $\text{mouse}_{\text{ambiguous}}$ tickles the $\text{monkey}_{\text{acc}}$] = The mouse tickles the monkey; OVS: Die Maus fängt gleich **der Affe** [The $\text{mouse}_{\text{ambiguous}}$ tickles the $\text{monkey}_{\text{nom}}$]= The monkey chases the mouse). When the second NP is also feminine or neuter the sentence remains ambiguous.

Children must learn these word order and case marking cues (among many others), in order to become competent in their native language. Bates and MacWhinney (1987, 1989) put forward the competition model to explain how children learn to use grammatical cues during acquisition; this model is based on cue validity. Cue validity is dependent on cue reliability (i.e. the frequency that the cue leads to correct interpretation) and cue availability (i.e. the frequency that the cue is provided in the materials to be processed). They argue that children should acquire those cues that have the highest validity first, and based on this, Bates and MacWhinney (1987) argued that constructions that have redundant cues should be the easiest to understand, particularly if these occur frequently within the language. This is known as the coalitions-as-prototypes model. For example, in German the unambiguous SVO constructions may be the easiest to comprehend given that they fall in typical SVO word order and are also marked unambiguously with case.

In German, case making is the most reliable cue, however it has low cue availability, given that only the masculine form is unambiguously marked for accusative and nominative case. The word order cue, on the other hand, has a high availability but has a low reliability, given the flexibility inherent in German word order. In this study we are interested which cues German speaking children use to establish who is doing what to whom when listening to a transitive sentence, in particular how they process case marking and cope with ambiguity. First, previous research investigating word order and case marking with children will be briefly reviewed.

4.1.1 Word order Processing in Children

Dittmar, Abbot-Smith, Lieven, & Tomasello (2008) investigated the role of word order and case marking in sentence comprehension in two groups of German children (2;7-year-olds and 4;10-year-olds) using a task in which the children had to act out the sentences using figurines. They used four sentence types: (1) SVO word order using a familiar verb in which both NPs were unambiguously case marked, (2) SVO word order using a novel verb in which both NPs were unambiguously case marked, (3) SVO word order using a novel verb with both NPs ambiguously case marked, and (4) OVS word order using a novel verb with both NPs unambiguously case marked. They found that the 2;7-year-old group was able to accurately comprehend the unambiguously SVO marked condition that had a familiar verb, while in all of the other conditions these children showed chance level performance. The 4;1-year-old children were able to comprehend the unambiguously marked SVO conditions for both the familiar and novel verbs, and also were able to comprehend the ambiguously marked SVO condition, but did not show above chance level performance with the unambiguously marked OVS condition. The authors argued that the task might have been too difficult for the 2;7-year-old group, as act out tasks incur high working memory demands. For the older group, it seemed that they were using word order cues, but had not yet acquired the ability to use case marking cues for interpretation.

Given the high demands of the act out task, in Dittmar et al. (2008), in their next experiment, 2;7, 4;1 and 7;3 year-old children saw two videos of an event involving two animals and heard a sentence with an unfamiliar verb; one video had the correct thematic roles, and one the reverse thematic roles. They found that the 2;7-year-old group showed above chance performance for unambiguously marked SVO condition, but in the other two conditions (SVO word order with both NPs ambiguously case marked and OVS word order with both NPs unambiguously case marked) they showed chance or below chance level performance. Unlike in the previous experiment, in this task the 2;7-year-old children showed above chance level performance for both familiar and novel verbs for the unambiguous SVO conditions. The 4;1-year-old children did not differ between the two tasks: they showed above chance performance for the unambiguously marked SVO condition and the ambiguously marked SVO condition, but displayed chance level performance for the unambiguous OVS condition. The 7;3-year-old children displayed above chance performance on all conditions.

These data suggest that there is a continuum in terms of the development of word order and case marking cue acquisition in German. In line with the coalitions-as-prototypes model, two-year-olds are able to comprehend the prototypical unambiguously marked SVO construction (when word order and case marking cues are redundant). Children over 4 begin to use word order cues during comprehension, but do not use case marking cues. Children over the age of seven, like adults, use both word order cues and case marking cues during comprehension. This suggests that constructions with redundant cues are important for early language acquisition, and it is not until older ages that children are capable of weighing the importance of grammatical cues during comprehension.

In a similar study Grünloh, Lieven, and Tomasello (2011) investigated the processing of OVS constructions while manipulating the prosody of the first NP with 4;10-year-old children and adults, using a pointing task on the computer screen (similar to the Dittmar et al. (2008) study). In this study OVS constructions were either unambiguously marked or ambiguously marked OVS constructions (with both NPs having ambiguous case marking). They presented these sentence types with neutral prosody or with the first NP having a prosodic cue that indicated a patient-first interpretation (Weber et al., 2006). They found that the children relied heavily on word order, but when case marking cues and prosody were both present the 4;10-year-olds were more accurate in comprehension than with word order cues alone. However, when presenting the same materials in a different task (with a context and puppets) the authors found that the 4;10-year-old group was able to use intonation to correctly interpret sentences without case marking. The authors speculate that intonation (in a more natural context) is often used for new information and this may have caused the children to pay closer attention to the prosodically salient element. Overall this suggests that, unlike the Dittmar et al. study, 4-year-olds are capable of using aspects of case marking, intonation, and that intonation cues can override word order bias for children as young as 4 years old.

Using ERP, Schipke et al. (2012) investigated the processing of unambiguous OVS and unambiguous SVO constructions with 3-year-olds, 4;6-year-olds, 6-year-olds, and adults. They investigated ERP components to the first NP and the second NP of these auditorily presented constructions. Behaviorally, the adults had no trouble with comprehending the sentences (as measured by sentence-picture matching responses), with ceiling level performance on all conditions. The adults displayed a negativity (which the authors called a 'topicalization negativity') at the first NP for the OVS constructions compared to the SVO, with no differences between conditions at the second NP. This suggested that there is some difficulty when encountering an object first construction (the topicalization negativity), but this is dealt with quickly and does not affect the subsequent processing of the OVS construction; this is in line with research suggesting that there is preference in German for initially interpreting sentences as subject-first constructions.

Across the three groups of children, Schipke et al. found a continuum for the development of processing of OVS and SVO constructions. Consistent with previous research, all three groups of children performed at chance level when it came to the behavioral interpretation of OVS constructions, but performed more accurately and above chance on the SVO constructions (3 yo: 64%; 4;6 yo: 89%; 6 yo: 96%); the only reliable difference was that the accuracy for OVS constructions was better for the 6-year-old group compared to the 3-year-old group. In terms of the ERP results, the 6-year-old group displayed a similar negativity to the first NP as displayed by the adults, but also showed a P600-like positivity at the 2nd NP. Schipke et al. argued that 6-year-olds process the first NP of the OVS construction similarly to adults, but the P600 at the second NP signifies difficulty in assigning the second NP the correct thematic role, which also manifests in the comprehension accuracy.

Schipke et al.'s 4;6-year-old group displayed no differences at the first NP, but did display a slight P600 like positivity at the second NP for the OVS compared to the SVO constructions. Hence, in this group, both word orders were being processed similarly at the first NP suggesting that regardless of case marking the 4;6-year-olds were using a word order strategy. The positivity at the second NP suggested that while word order seemed to play a more important role, this age group was still able to detect the case marking at the second NP. Schipke et al. argued that the P600 signified the difficulty associated with integrating the second NP and the costs associated with assigning thematic roles to this object-first construction. This suggests that that while children of 4;6 years have difficulty assigning thematic roles in object first constructions, they are aware of the case markings and are processing them in some form. The 3-year-old group displayed a late positivity at the first NP for the OVS constructions compared to the SVO, and no differences at the second NP, suggesting that this age group had some degree of sensitivity to unambiguous accusative marking at the first NP in OVS constructions, but it may have been that they processed this as though there was a morphosyntactic violation at the first NP compared to the SVO constructions.

Together, these data led Schipke et al. (2012) to argue that 3-year-olds can detect the difference between OVS and SVO constructions but use word order over case marking in sentence processing, the 4;6-year-olds recognize the case marking differences but still struggle with thematic role assignment, and 6-year-olds process case marking at the first NP similarly to adults, but thematic role assignment at the second NP remains effortful.

Other research investigating the processing of these types of constructions in adults has found that despite the SVO constructions being more common in German, adults are capable of comprehending these sentence types with high accuracy (e.g. Chapter 3, this thesis; Bader & Meng, 1999;

Schlesewsky, Fanselow, Kliegl & Krems, 2000; Schriefers, Friederici, & Kühn, 1995). However, there are still some processing difficulties associated with the less common OVS constructions and with ambiguous constructions (e.g. Matzke et al., 2002; Rösler, et al., 1998; Schlesewsky, Bornkessel, & Frisch, 2003). This had been suggested to be due to either a violation of canonicity (Chapter 3, this thesis; Schlesewsky et al., 2003), or due to working memory demands (Matzke et al., 2002; Rösler et al., 1998).

Research has found processing difficulties associated with late disambiguating constructions at the point in which the second NP reveals the dispreferred OVS word order compared to when it reveals the SVO order in fMRI, ERP (e.g., Friederici, Steinhauer, Mecklinger, & Meyer, 1998; Frisch, et al., 2002; Matzke, et al., 2002; Bahlman, Rodriguez-Fornells, Rotte, & Münte, 2007), and in the previous chapter a pupil paradigm. The difficulties associated with the ambiguous constructions are believed to index the processing costs that arise when the parser has to revise the original parsed (and preferred) SVO order into an OVS order (upon reaching the second disambiguating second NP).

4.1.2 Current Study

In the first study to use pupillometry (a measure of cognitive effort) with children, we extend the previous research using online processing with children investigating the processing of case marking and sentence order with 4-year-old children. This research also builds on the previous chapter, where pupillometry was shown to be sensitive to both word order and case marking cues in German-speaking adults. In Chapter 3, using stimuli based on the manipulations of word order and ambiguity presented in Table 2, across the whole sentence, a larger pupil change slope was found when processing OVS constructions compared to SVO and larger pupil change slope for ambiguous compared to unambiguous conditions (approaching significance), with ambiguous constructions evoking a larger peak pupil size than unambiguous. There were no differences at the first NP, but at the second NP, Chapter 3 found that the unambiguous SVO constructions evoked a smaller pupil change slope than both OVS conditions and the ambiguous SVO condition, which was attributed to a violation of the expectation of canonicity. The difference between the ambiguous OVS constructions compared to the unambiguous SVO condition was mostly likely due to reanalysis from the preferred SVO interpretation into the OVS order, and the difference between the ambiguous SVO and the unambiguous SVO construction was most likely due to the 'unsubstantiated' nature of the parse, that is not resolved until the second NP.

In the study presented below, four-year-old children were chosen given that previous research has found evidence for use of word order cues during processing and the development of case marking sensitivity during processing for this age group (e.g., Dittmar et al., 2008; Grünloh et al., 2011 ; Schipke et al., 2012). Based on previous research, and in line with coalitions-as-prototypes model, 4 year-olds should be able to comprehend SVO sentences with high accuracy, particularly when they are unambiguously case marked. However, they are likely to display chance level performance when it comes to the interpretation of OVS constructions (e.g., Dittmar et al., 2008; Grünloh et al., 2011 ; Bates & MacWhinney, 1987; 1989; Schipke et al., 2012).

It has been proposed that 4 year-olds primarily rely on word order and interpret the first NP as an agent, but that they are able to detect case marking inconsistency at the second NP, which causes some difficulties for the OVS constructions (Schipke et al., 2012). Based on this, no differences in pupil response are predicted between the SVO and OVS constructions for this age group at the first NP; and at the second NP there is predicted to be greater pupil diameter change for the OVS constructions compared to the SVO constructions. When it comes to the pupil diameter across the whole sentence, it is possible that the children may mimic the adults in the previous chapter and display a greater pupil diameter for the OVS constructions compared to the SVO and for the constructions that are ambiguous at the first NP compared to those that are unambiguous. This would signify that children the age of 4 years are sensitive and using case marking cues for both the ambiguous and unambiguous OVS constructions.

4.2 Methods

4.2.1 Participants

Thirty children participated in this study and were recruited from the greater Potsdam area of North East Germany. All participants were native monolingual speakers of German. Parents gave consent for participation. Nine participants were excluded due to missing data from movement, missing trials, etc.; twenty-one participants (10 female) were therefore included in the analysis with an average age of 4.47-years-old (sd=0.23). Participants' parents were paid 7.50 (Euro), and participants were given a toy for participating.

4.2.2 Materials

Four sentence types were used in this experiment, as seen in Table 4.1. There were a total of 24 critical trials in each condition.

Table 4.1: Example sentences with different word order and case marking

	Unmbiguous First NP	Ambiguous First NP
SVO	**Der Uhu** kitzelt gleich das Meerschwein	Das Meerschwein kitzelt gleich **den Uhu**
	The owl $_{nom}$ tickles the guinea pig$_{acc}$	The guinea pig$_{nom}$ tickles the owl$_{acc}$
	The owl tickles the guinea pig	The owl tickles the guinea pig
OVS	**Den Uhu** kitzelt gleich das Meerschwein	Das Meerschwein kitzelt gleich **der Uhu**
	The owl$_{acc}$ tickles the guinea pig$_{nom}$	The guinea pig$_{acc}$ tickles the owl$_{nom}$
	The guinea pig tickles the owl	The guinea pig tickles the owl

The stimuli were constructed in a 2x2 design, manipulated by word order (SVO/ OVS), and ambiguity (ambiguous/unambiguous). The SVO stimuli were in subject-verb-object word order, and the OVS stimuli were in object-verb-subject word order; the unambiguous stimuli had the first NP in the construction unambiguously case marked (with the masculine case), making the subject or object apparent at the first NP. The ambiguous constructions had an ambiguously case marked first NP (with either neuter or feminine case marking), and an unambiguously case marked second NP (with the masculine case marking) making the subject and object apparent only upon reaching the second NP.

The critical sentences were simple transitive sentences containing two NPs and one verb. The referents of the NPs were always two different animals, and the verb was either *kitzeln* (to tickle) or *fangen* (to chase). The animals for one sentence were chosen in a way that they were equally likely to engage in the action described in the sentence so that no plausibility information could enter the sentence interpretation (e.g. predator chasing prey).

The sentences were recorded by a female native speaker of German at a normal speaking rate in the phonetics lab at the University of Potsdam with a sampling rate of 22050 Hz and saved as .wav files.

The participants heard a sentence, and afterwards were presented with a video depicting the action that was just described (this video was either correct or showed the reverse thematic roles). The videos used in this study showed the two animals from the previous sentence; for the verb *kitzeln* (tickle) the video depicted one animal approaching the other and tickling them with a feather, and for *fangen* (chase) the image depicted one animal running towards the other animal that did not move. In half of the trials the sentence presented aurally to the participant (e.g. The owl tickles the guinea pig) matched the video presented to the participant (i.e. participants would see an owl ticking a guinea pig), and in half of the trials the recording presented aurally did not match the video presented to the participant, rather the agent and patient role were reversed (i.e. participants would see a guinea pig tickling an owl). This was counterbalanced and pseudo randomized across 8 versions and participants were randomly assigned to one of the eight pseudo-randomized versions. Each participant saw a video that accurately depicted the recording half of the time and saw a video that depicted the reverse object/subject relationship half of the time; each animal was only used once.

4.2.3 Apparatus

Stimulus presentation was programmed using Tobii Studio software. Pupil diameter was recorded with a Tobii 1750, sampling at 50 Hz. Tracking was binocular but only the right eye was used for analysis. The eye tracker was built into a 17-inch thin film transistor liquid crystal colour display monitor (1,280 x 1,024 pixel resolution). Participants sat approximately 60-65 cm away from the display and tracking was remote. Eyes were calibrated using a 5-point sequence.

4.2.4 Design and Procedure

Word order (SVO/OVS) and ambiguity (unambiguous/ambiguous) were manipulated within subjects. The 8 pseudo-randomized versions were broken into three blocks of eight trials, between each block there was an attention-grabbing clip with Snoopy, Elmo, or Kitty. Within each block there were never more than three trials in a row where the animal was moving on the same side of the screen, never more than three trials in a row in which the video and sentence matched or did not match, never more than two trials in a row with the same verb, and never more than two trials in a row with the same word order. Critical items were counterbalanced across each block, so that each participant saw only one item of each condition.

Participants either sat by themselves or in the lap of their caregiver, a video was shown with an animation of a girl explaining the directions for the experiment; they were going play a game, they would watch a video with two animals, but before they watched the video she (the girl) would guess what the animals were going to do in the video, and after they watched the video the child has to say whether she guessed correctly or not. When the participant was ready, a fixation cross appeared on the screen for 2000 ms (to allow the pupil to adjust to screen luminance), after which the audio recording of the sentence was presented; the fixation cross remained on the screen for the entirety of the sentence and for an additional 2000 ms after the end of the sentence presentation. Participants were asked to focus on the fixation cross, and to attend to the aurally presented sentence. After the fixation cross disappeared the participants saw the girl telling them to pay attention, the video depicting the action of the previous sentence would play, and the girl would return on the screen asking whether she was right. Participants answered the question orally which was recorded by the experimenter. We used this method (presenting a screen with just a cross while the auditory stimulus was presented) to ensure eye movements did not affect pupil change.

4.2.5 Analysis

Blinks were filtered out by replacing the missing values with linear interpolation, and the pupil was normalized by item for each participant, so all items started at zero. We did not look at pupil change during the video presentation, rather we analyzed pupil change across the entirety of the sentence presentation as well as two other critical windows: 1) the onset first NP and 2) the onset of the second NP. A 1.2 second window starting at the onset of each NP was investigated; while there is no research, to our knowledge, on the time it takes the pupil of a child to reach maximum diameter, in adults the pupil reaches maximum diameter in approximately 1.2 seconds (Just & Carpenter, 1993). Before analysis, data was screened for outliers; any data point three or more standard deviations from the mean were replaced with the mean of that condition. All trials were averaged together within each condition resulting in 4 vectors for each participant (one for each of the 4 conditions). These vectors were submitted to a simple regression where pupil size served as the dependent variable and time served as the independent variable. The slope of pupil size change over time (i.e. the unstandardized regression coefficient) was the main dependent variable in this study. We also calculated the peak amplitude size (which was the largest size that the pupil reached during the presentation of the whole sentence and during the two time windows), and the peak latency (which was the time in which the peak pupil size was reached) for each vector; these are typical measures in pupillometry research (e.g. Schmidtke, 2014; Zekveld, Kramer, & Festen, 2010).

Given that the critical sentences varied slightly in terms of length all pupil data were cut to encompass the pupil change during length of the shortest sentence (3.83 seconds) for the analysis across the whole construction (given that there was one unstandardized regression coefficient value for each time point, the trials had to be the same length). Comprehension accuracy, the unstandardized regression coefficient, peak amplitude, and peak latency were submitted to linear mixed effect models using R (R Core Team, 2012) and *lme4* (Bates, Maechler & Bolker, 2012), and Tukey's contrasts from the *multcomp* package (Horthorn et al, 2013). All of which had the fixed effects of word order (SVO/OVS) and ambiguity (ambiguous NP1/unambiguous NP1), and random effects of participants; the random effect of participant was maximally specified with random slopes for order, and ambiguity. Models were compared such that a full model was contrasted with a reduced model to test the factor in question; nonsignificant factors were excluded from the model by contrasting each main effect and interaction, starting with the order, followed by ambiguity, and then finally the interactions. Model comparison was based on log likelihood ratio tests.

4.3 Results

4.3.1 Sentence-Picture Matching Performance

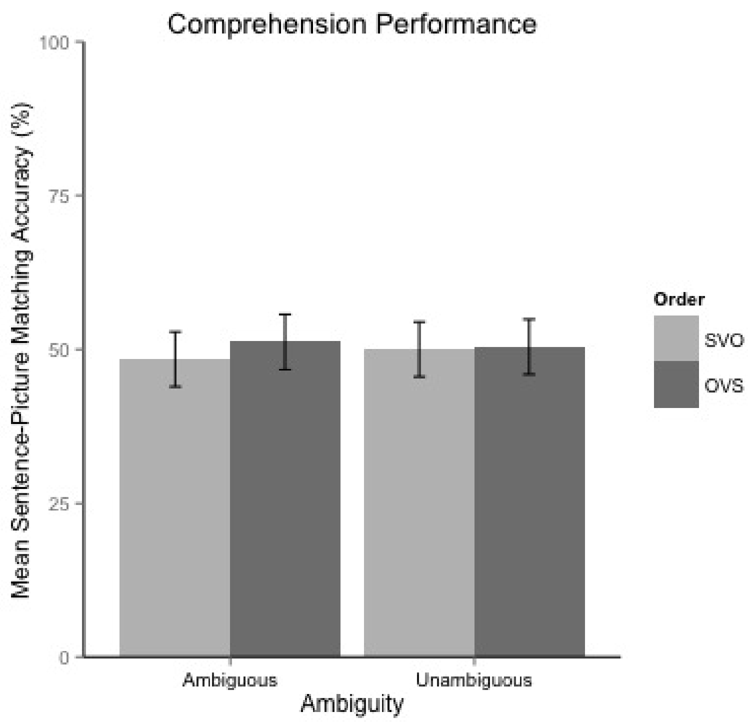

Figure 4.1: Accuracy of sentence to video judgements (the error bar represents the standard error of the mean).

Participants were asked to make a judgement on whether the sentence they heard matched the video on the computer screen. The mean percentages of correct judgments can be seen in Figure 4.1. Linear mixed effect modeling revealed no main effects ($X^2(2)$= 0.13, p=0.93) and no interaction ($X^2(3)$= 0.20, p=0.91). In none of the conditions did the participants score significantly above chance (50%; all $ps>$ 0.70).

Given the factors manipulated in this study appeared to have no effect on comprehension accuracy (which was uniformly low), we ran a post-hoc analysis of comprehension performance for the stimuli presented in the first half (12 items) of the experiment compared to the last half (12 items) of the experiment. This aimed to investigate the possibility that the task taxed participants such that they performed worse as the experiment went on, or performed better as they became more practiced at the task. Hence, we ran a linear mixed model with the fixed effects of word order (SVO/OVS) and ambiguity (ambiguous/unambiguous), presentation half (first/second) and random effects of participants; the random effect of participant was maximally specified with random slopes for order, ambiguity, and presentation half.

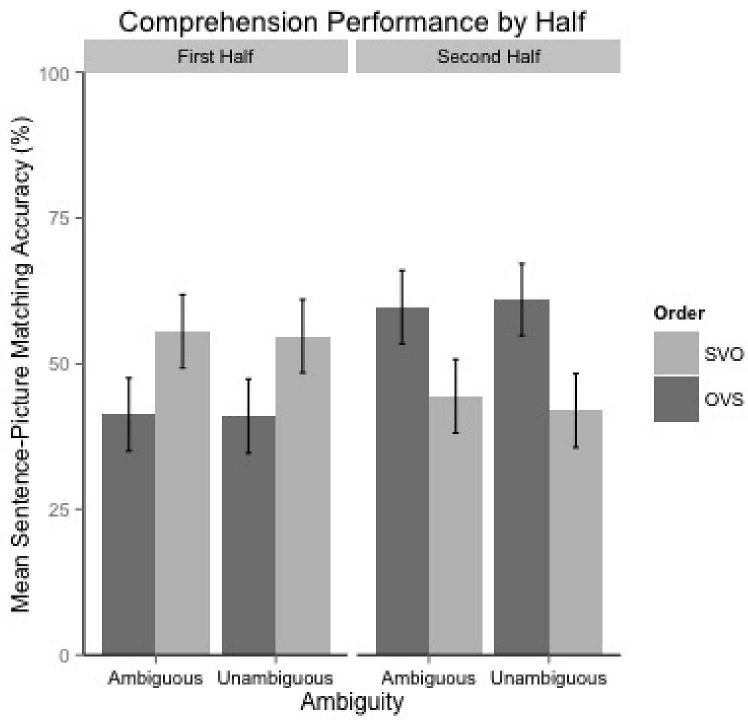

Figure 4.2: Sentence and picture matching comprehension accuracy by half (the error bar represents the standard error of the mean)

The mean percentages of correct judgments split by halves can be seen in Figure 4.2. Linear mixed effect modelling revealed no main effects (word order: $X^2(1)= 0.13$, p=0.72; ambiguity: $X^2(1)= 0.65$, p=0.42; presentation half: $X^2(1)= 0.64$, p=0.42). However, there was an interaction between word order and half ($X^2(1)= 13.08$, $p<0.01$). Tukey's contrasts revealed that performance on the OVS constructions was significantly more accurate in the second half compared to the SVO constructions in the second half (p=0.03), and performance on the OVS constructions was significantly more accurate in the second half compared to the first half (p=0.01).

4.3.2 First NP

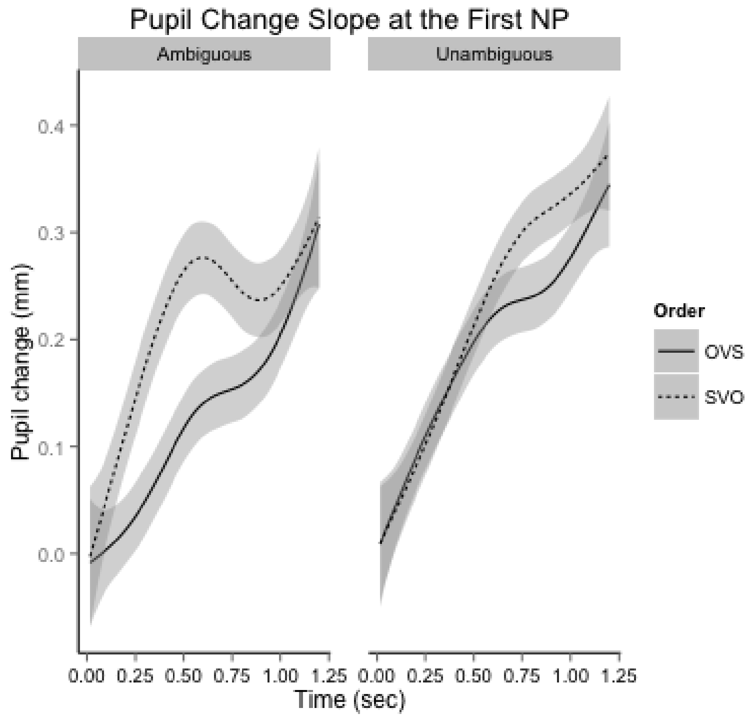

Figure 4.3: Pupil change from the onset of the first NP: Smoothed pupil change slope across time, the grey highlighting indicates standard error.

Linear mixed effect modeling of the pupil change slope at the first NP (Figure 4.3) revealed no main effects (word order: $X^2(1)= 1.50$, p=0.21; ambiguity: $X^2(1)= 0.38$, p=0.53). There was an interaction between ambiguity and word order ($X^2(3)= 19.99$, p<0.001) when comparing a model with and without an interaction. Tukey's contrasts revealed that SVO unambiguous, SVO ambiguous, and OVS unambiguous had a larger pupil diameter than the OVS ambiguous condition (all ps<0.001).

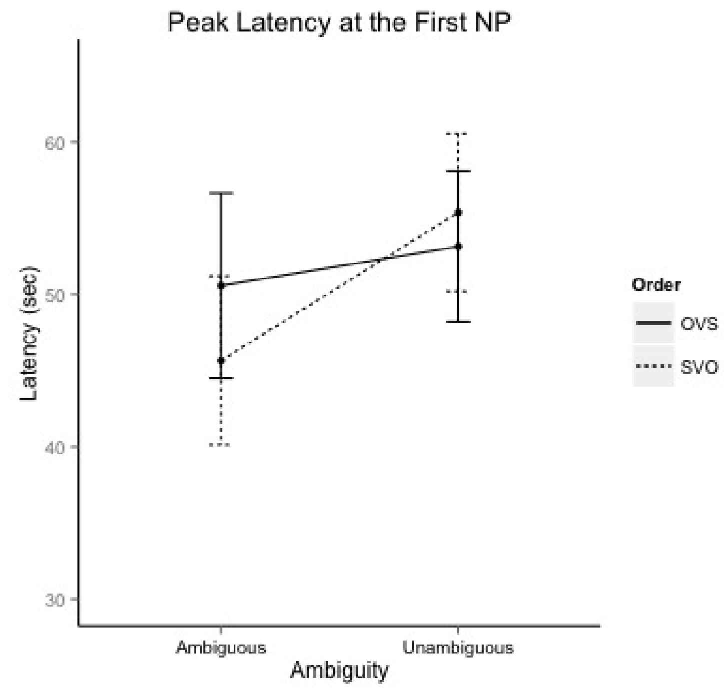

Figure 4.4: Peak amplitude from the onset of the first NP (bars represent standard error).

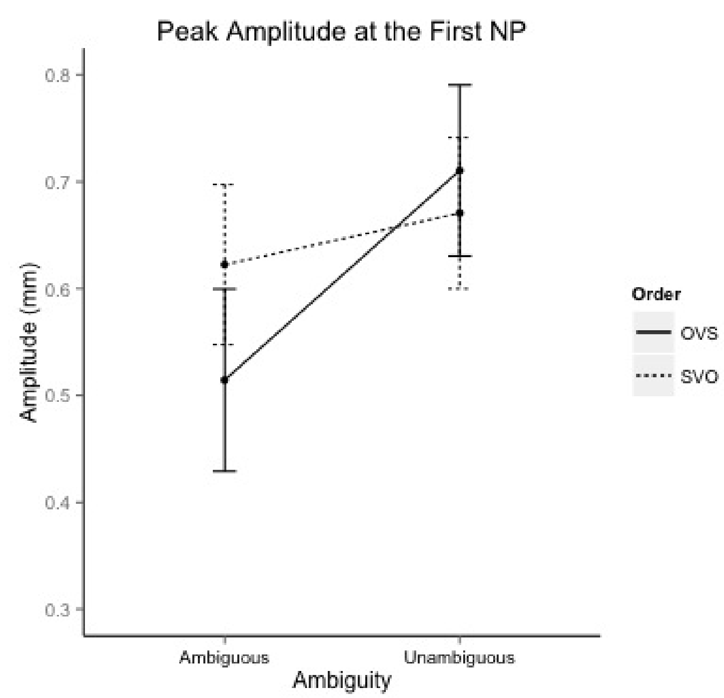

Figure 4.5: Peak latency from the onset of the first NP (bars represent standard error).

Raw peak amplitude revealed no main effects (word order: $X^2(1)= 0.79$, p=0.21; ambiguity: $X^2(1)= 1.47$, p=0.22) and no interaction ($X^2(2)= 2.06$, p=0.56) see Figure 4.4, and raw peak latency also revealed no main effects (word order: $X^2(2)= 0.22$, p=0.63; ambiguity $X^2(1)= 4.34$, p=0.22), see Figure 4.5.

4.3.3 Second NP

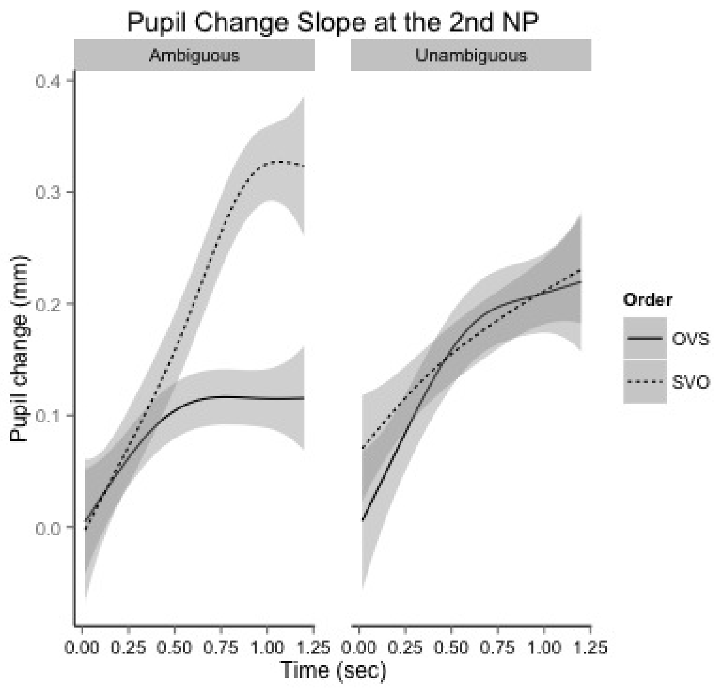

Figure 4.6: Pupil change from the onset of the second NP: Smoothed pupil change slope across time, the grey highlighting indicates standard error).

At the second NP (see Figure 4.6) there were no main effects of word order or ambiguity on pupil change slope (word order: $X^2(1)= 2.30$, p=0.13; ambiguity: $X^2(1)= 0.56$, p=0.45). However, once again there was an interaction of ambiguity and word order ($X^2(3)= 27.41$, $p<0.001$) when comparing a model with and without an interaction. Tukey's contrats revealed that there was a shallower pupil change slope for the ambiguous OVS condition compared to both unambiguous OVS and unambiguous SVO conditions. There was also a steeper pupil change slope for ambiguous SVO compared to the ambiguous OVS and unambiguous OVS conditions (all $ps<0.05$), but the difference did not reach significance with the ambiguous SVO (p=0.11). There was no significant difference between the unambiguous

SVO and OVS conditions (p=0.87).

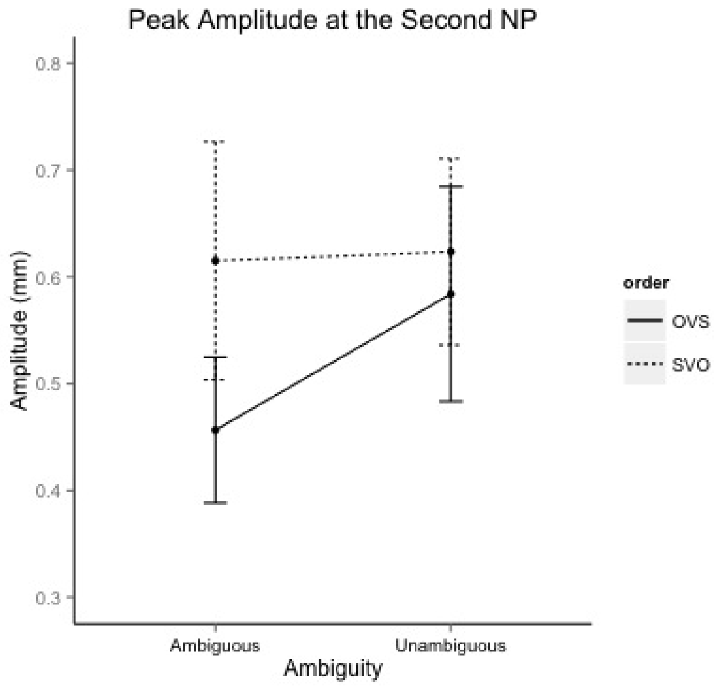

Figure 4.7: Peak amplitude from the onset of the second NP (bars represent standard error).

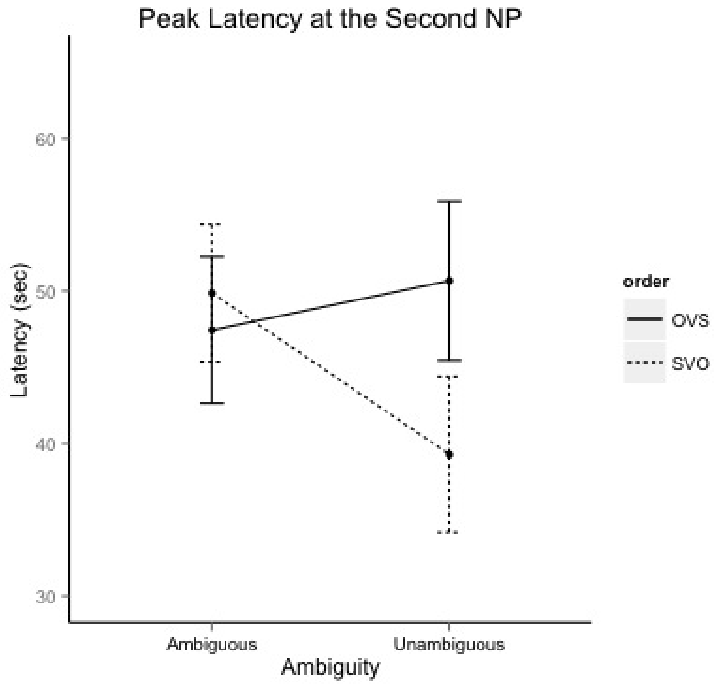

Figure 4.8: Peak latency from the onset of the second NP (bars represent standard error).

Raw peak amplitude revealed no main effects (word order: $X^2(1)= 0.92$, p=0.34, ambiguity: $X^2(1)= 0.62$, p=0.43) or interaction ($X^2(3)= 3.83$, p=0.28), see Figure 4.7. Raw peak latency also revealed no main effects (word order: $X^2(1)= 0.34$, p=0.27; ambiguity: $X^2(1)= 0.45$, p=0.55) or interaction ($X^2(1)= 2.19$, p=0.53), see Figure 4.8.

4.3.4 Whole Sentence

Figure 4.9: Pupil change across the whole sentence: Smoothed pupil change slope across time, the grey highlighting indicates standard error.

Linear mixed effect modeling of the change of pupil slope across the whole sentence revealed a main effect of order approaching significance ($X^2(1)=$ 2.84, p=0.09), with the SVO conditions having a greater pupil slope change than the OVS conditions (refer to Figure 4.9). There was no interaction ($X^2(3)=$ 1.32, p=0.72).

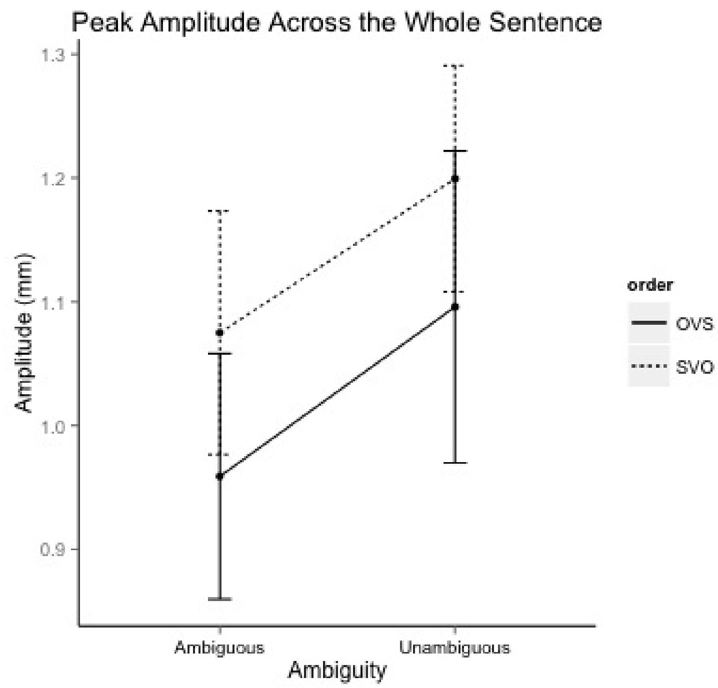

Figure 4.10: Peak amplitude across the whole sentence (bars represent standard error).

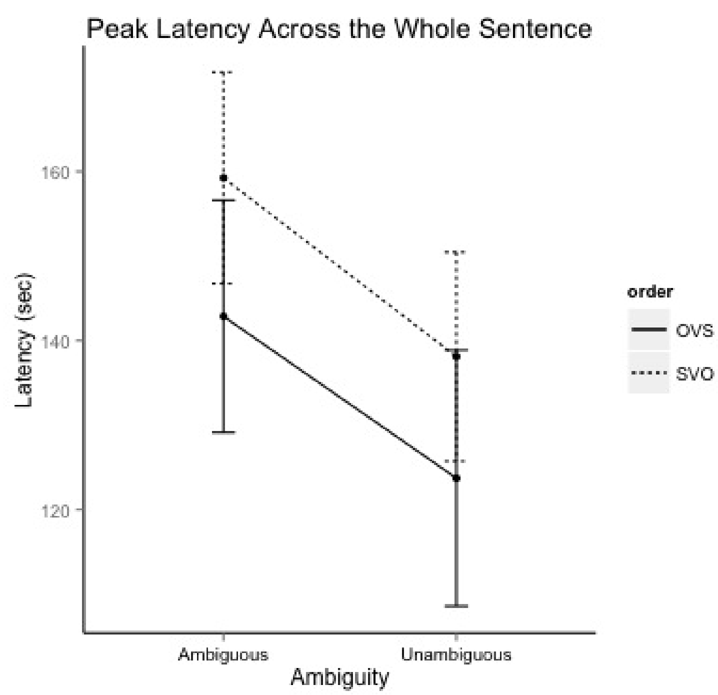

Figure 4.11: Peak latency across the whole sentence (bars represent standard error).

Raw peak amplitude revealed no main effects (refer to Figure 4.10, word order: $X^2(1)= 0.79$, p=0.21; ambiguity: $X^2(1)= 2.29$, p=0.13) or interactions ($X^2(3)= 1.65$, p=0.64), and raw peak latency also revealed no main effects (refer to Figure 4.11, word order: $X^2(1)= 1.37$, p=0.24; ambiguity: $X^2(1)=$ 1.95, p=0.16) or interactions ($X^2(3)= 3.65$, p=0.30).

4.3.5 Adult and Child Comparison

As we had a unique opportunity to directly compare pupil data to the same stimuli across two age groups, we compared the adult data from the previous chapter and the child data from the current study; to the best of our knowledge this is the only study to do have done such a comparison. This an exploratory analysis with the aim of ascertaining information on pupil change differences between adults and children.

4.3.6 Analysis

All participants from the two experiments that went into the analysis of the single groups were included here, linear mixed effects models included the fixed effects of group (adult/child) and condition (SVO unambiguous/SVO ambiguous/OVS unambiguous/OVS ambiguous), and random effects of participants; the random effect of participant was maximally specified with random slopes for group, and condition.

4.3.6.1 First NP

Linear mixed effect modeling of the pupil slope change at the first NP revealed a main effect of both group and condition ($X^2(3)= 103.98$, $p<=0.001$) compared to a model with one main effect, ($X^2(1)= 0.58$, p=0.33), and a significant interaction ($X^2(3)= 33.93$, $p<0.001$ (refer to Figure 4.12). Tukey's contrasts revealed that the two ambiguous conditions evoked a steeper pupil slope change in the adults compared to the children ($p<0.05$).

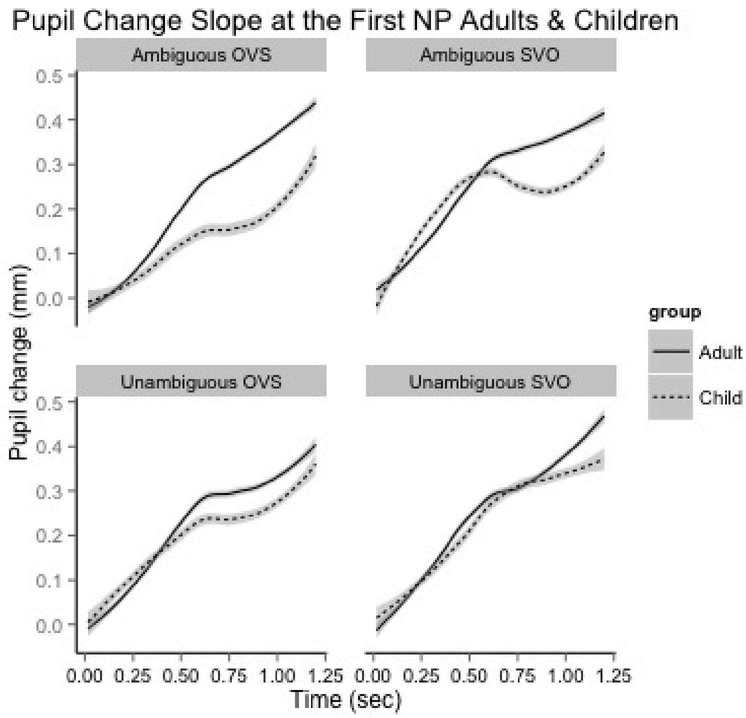

Figure 4.12: Adult and child smoothed pupil change slope at the first NP (the grey highlighting indicates standard error).

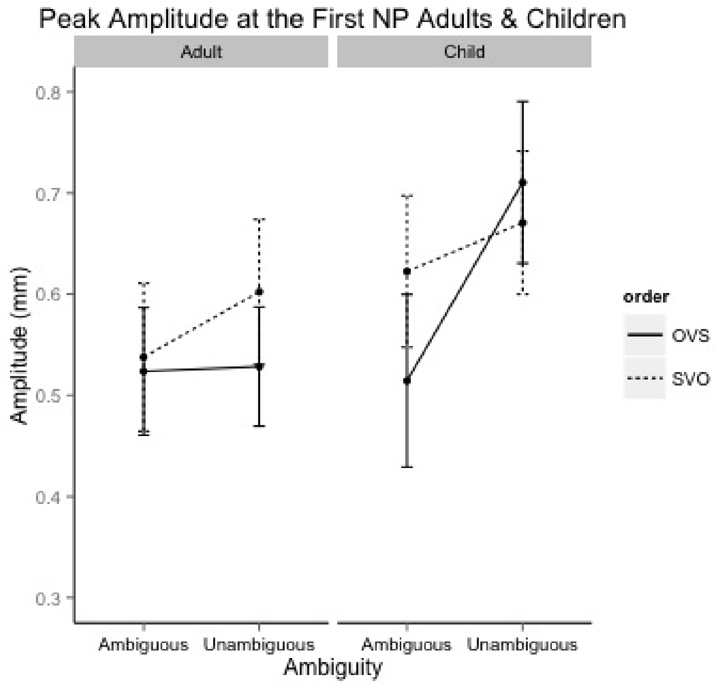

Figure 4.13: Adult and child peak latency from the onset of the first NP (bars represent standard error).

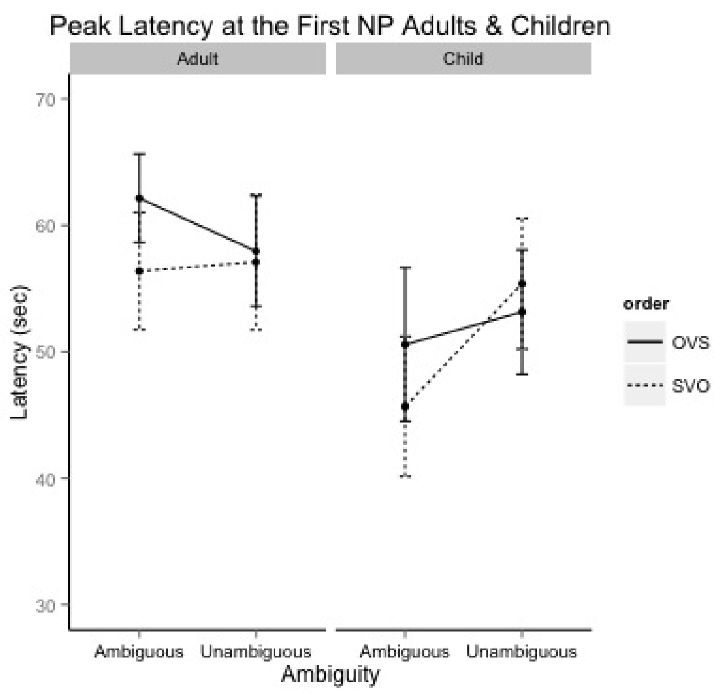

Figure 4.14: Adult and child peak latency from the onset of the first NP (bars represent standard error).

Raw peak amplitude revealed a main effect of group approaching significance ($X^2(1)= 13.04$, $p=0.08$) with the child group having larger peak amplitudes than the adult group, and no interactions ($X^2(3)= 1.61$, $p=0.66$), refer to Figure 4.13. Raw peak latency revealed no main effects (group: $X^2(1)= 1.69$, $p=0.22$; condition: $X^2(1)= 4.25$, $p=0.24$), and no interaction ($X^2(3)= 2.42$, $p=0.49$), refer to Figure 4.14.

4.3.7 Second NP

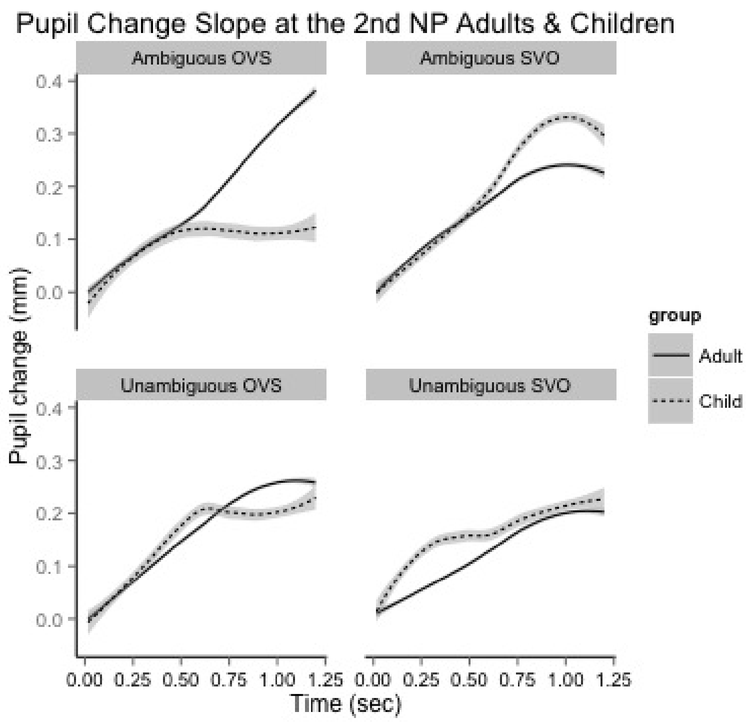

Figure 4.15: Adult and child smoothed pupil change slope at the second NP (the grey highlighting indicates standard error).

Linear mixed effect modeling of the pupil change slope at the second NP revealed a main effect of group and condition ($X^2(3)= 33.57$, $p<0.001$) compared to a model with one main effect, ($X^2(1)= 0.43$, $p=0.50$), and an interaction ($X^2(3)= 91.30$, $p< 0.001$) (refer to Figure 4.15). Tukey's contrasts revealed that the two unambiguous conditions evoked a steeper pupil change slope in the child group compared to the adult group, and revealed that the OVS ambiguous condition evoked a steeper pupil change slope in the adult compared to child group (all $ps<0.05$).

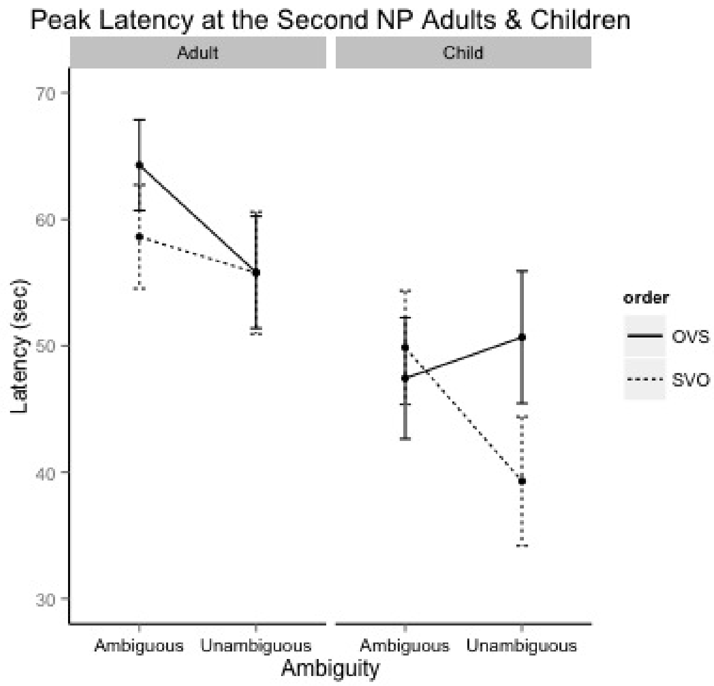

Figure 4.16: Adult and child peak amplitude from the onset of the second NP (bars represent standard error).

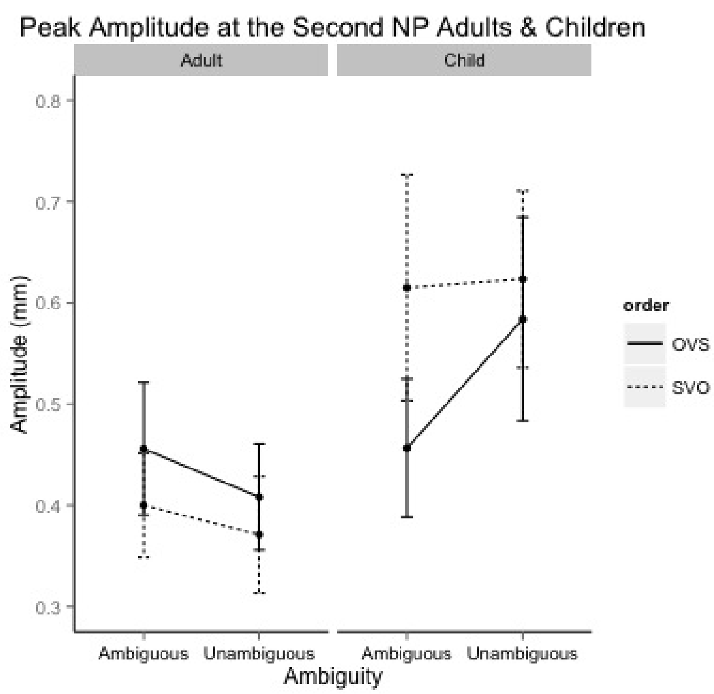

Figure 4.17: Adult and child peak latency from the onset of the second NP (bars represent standard error).

Raw peak amplitude revealed a main effect of group ($X^2(1)$= 10.05, p<0.01) with the child group having larger peak amplitude than the adult group, and no interactions ($X^2(3)$= 2.71, p=0.44), Figure 4.16. Raw peak latency revealed a main effect of group ($X^2(1)$= 6.23, p< 0.05), with the adult group having having a later peak latency than the child group, and there was no interaction ($X^2(3)$= 3.65, p=0.30) when comparing a model with and without an interaction, refer to Figure 4.17.

4.3.8 Whole Sentence

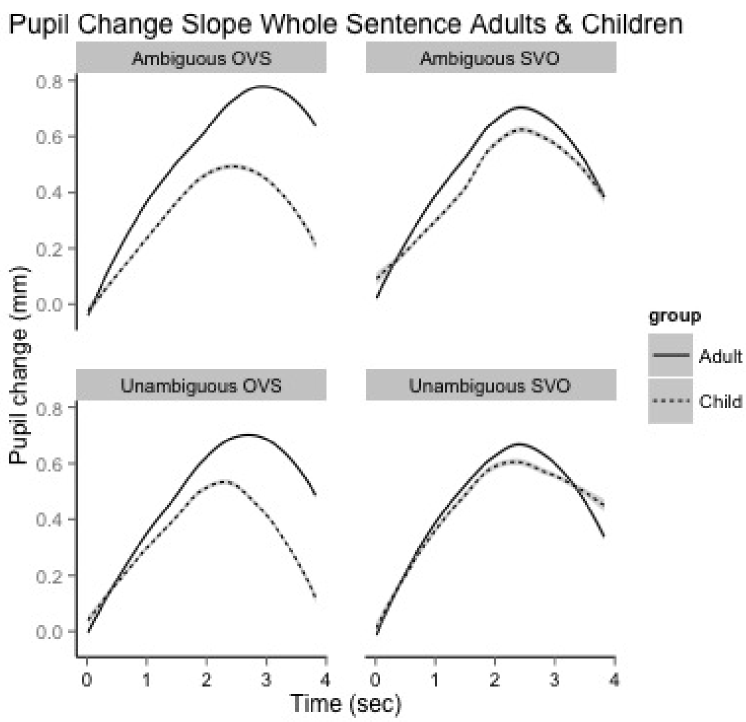

Figure 4.18: Adult and child smoothed pupil change slope across the whole sentence (the grey highlighting indicates standard error).

Linear mixed effect modeling of the pupil change slope across the whole sentence revealed a main effect of group and condition ($X^2(3)$= 75.32, p<0.001) compared to a model with only a main effect of group ($X^2(1)$=1.90, p=.09, and an interaction $X^2(3)$= 306.04, p< 0.001) (refer to Figure 4.18). Tukey's contrasts revealed that the SVO ambiguous, OVS unambiguous, and OVS ambiguous conditions evoked a steeper pupil slope in the adult group compared to the child group (all ps< 0.001).

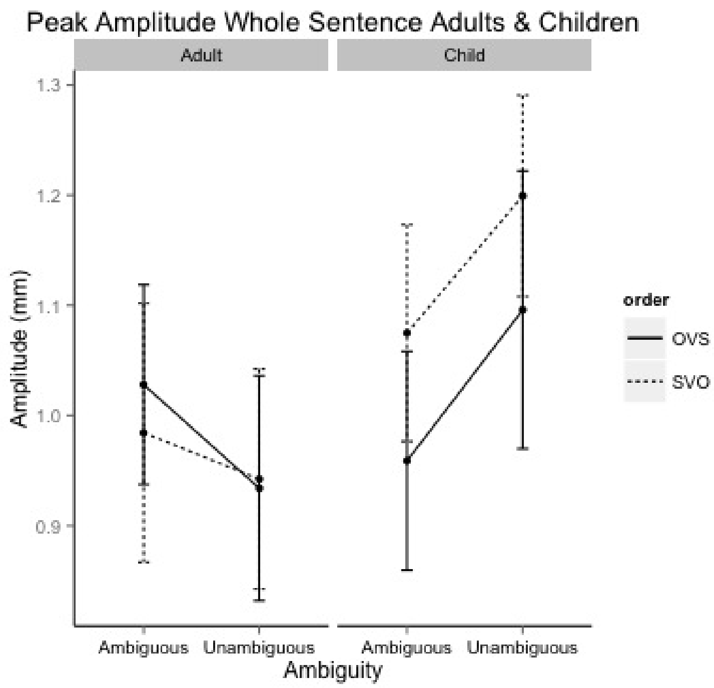

Figure 4.19: Adult and child peak amplitude across the whole sentence (bars represent standard error).

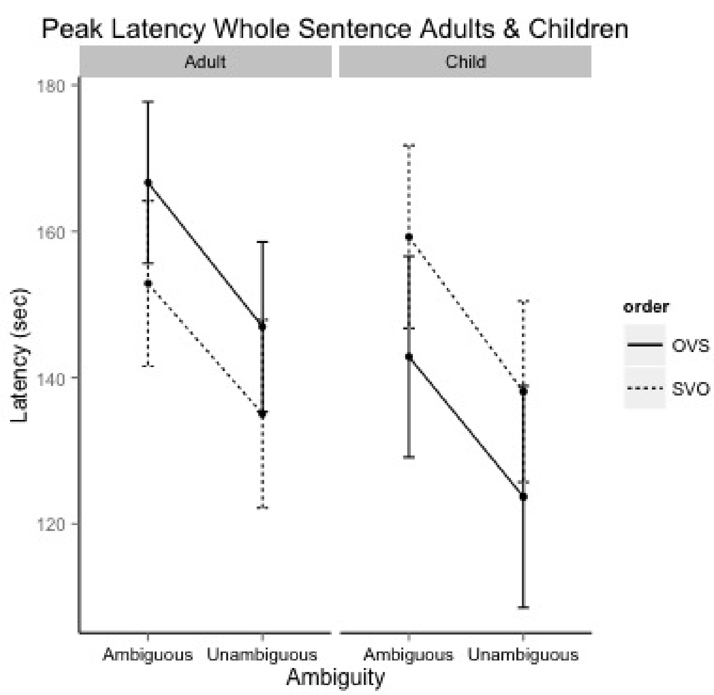

Figure 4.20: Adult and child peak latency across the whole sentence (bars represent standard error).

Raw peak amplitude revealed no main effects (group: $X^2(1)= 0.49$, $p=0.38$; condition: $X^2(3)= 5.06$, $p=0.13$) and no interaction ($X^2(3)= 3.06$, $p=0.38$), see Figure 4.19. Additionally, raw peak latency revealed no main effects (group: $X^2(1)= 0.81$, $p=0.36$; condition: $X^2(3)= 0.91$, $p=0.82$) and no interaction $X^2(3)= 4.10$, $p=0.25$), see Figure 4.20.

4.4 Discussion

In this study changes in pupil diameter were used as means of measuring of cognitive costs in the processing of word order in a group of 4;5-year-old German speaking children. The constructions were either in SVO order, the more frequent word order in German for which adults show superior processing (e.g. Chapter 3, this thesis; Bader & Meng, 1999; Schlesewsky, et al., 2000; Schriefers, et al., 1995), or in OVS order; additionally, the sentences were either case marked such that the order of the arguments in the sentence could be inferred from the first noun phrase (unambiguous first NP condition) or only unambiguously case marked at the second NP in the construction(ambiguous first NP condition). We used a novel child friendly methodology (pupillometry) to test the processing involved with these constructions, to determine whether pupillometry was sensitive to these linguistic manipulations, and whether it could reveal any new insights into the processing of these constructions.

Previous research had suggested that, around the age of 4 years, children are capable of interpreting SVO constructions with high accuracy, but show chance level performance for OVS constructions (e.g Bates & MacWhinney (1987;1989); Dittmar et al., 2008; Grünloh et al., 2011 ; Schipke et al., 2012.). Our behavioral results showed that the manipulations in the study did not affect performance across conditions; chance level performance was anticipated for the OVS constructions, but we also found chance level performance for the SVO constructions, which was contrary to previous research. Several studies have found that this age group, and even 2-year-olds (Dittmar et al., 2008) displayed high accuracy for unambiguous SVO constructions.

Given these unexpected findings a post-hoc analysis was run to see whether there were differences in performance during the first and second half of the items. It appeared as though the participants may have changed their answering heuristic as the study unfolded: in the first half of the experiment OVS sentences were less accurate, whereas in the second half accuracy was higher for the OVS sentences. This led to poor performance overall. In addition, it is possible that the task demands may have influenced performance. Previous studies involved an immediate comprehension task, while in the current study the participants first heard the construction in question, then 2 seconds later, this was followed by a short video, and then they were asked whether the sentence and video matched. It may be that the lag between the sentence presentation and comprehension probe was too great, and perhaps the working memory demands led the children to guess. More research needs to be done to further validate this potential explanation.

The fact that behavioral performance was at chance does not, however, negate the value of the pupillometry data which measured online processing while the participants listened to the sentences. Nevertheless, given the novelty of the methodology our hypotheses for pupil change were necessarily tentative. At the first NP we anticipated no differences between the conditions, but pupillometry revealed a smaller pupil change slope for the ambiguous OVS constructions compared to the other conditions, and no difference in peak size or peak latency. This difference is somewhat hard to interpret, it may be that the two SVO constructions, and the unambiguous OVS constructions evoked more processing resources because these three constructions are more straightforward, and therefore participants paid more attention and invested more resources during processing compared to the ambiguous OVS construction. If this is the case, than it is appears that children within this age group were sensitive to case marking cues, as both the unambiguous conditions were processed similarly. However, this of course cannot account for the ambiguous SVO construction, and more research needs to be conducted in an attempt to replicate and further understand this effect.

At the second NP, we anticipated larger pupil change, peak amplitude, and peak latency processing for the OVS constructions compared to SVO constructions given the difficulty associated with ambiguous OVS constructions and the research suggesting that 4-year-olds may be sensitive to case marking cues (e.g. Schipke et al., 2012). We found a smaller pupil change slope for the OVS ambiguous condition (compared to the unambiguous SVO and unambiguous OVS conditions), suggesting that participants are investing more processing at this point only to the conditions in which the case marking at the second NP is reconfirms the parse being built (unambiguous SVO/OVS) given that word order is apparent from the onset). Again, it may be that this age group invests more into those constructions which are straightforward from onset (unambiguous), which may indicate that this age group is sensitive to the case marking of the first NP. Additionally, there was a larger pupil change slope for ambiguous SVO constructions (compared to the both the OVS conditions), and no differences in terms of peak amplitude and peak latency. Again, these findings go against predictions and seem to suggest that that SVO conditions were more difficult to process than the OVS, indicating that the participants were investing more processing into the easier SVO conditions, than they were the more difficult and less common OVS conditions.

When looking at processing across the whole sentence, there was an effect of order approaching significance with the SVO conditions having greater pupil change slope than the OVS, which suggests that the participants were investing more resources into comprehending the typical SVO constructions. This is unlike adults who displayed a greater pupil diameter for OVS conditions compared to SVO.

These findings paint an interesting picture of comprehension and processing of word order and case ambiguity in German 4;5-year-old children, which is unlike that found in previous research. The pattern found here suggests this age group is investing more resources into the processing of SVO constructions at both NPs and across the whole construction. However, these results are not mirrored in the behavioral accuracy, with this group performing at chance for all conditions, but this may be the result of the task being too difficult. These are all unexpected findings, and it is difficult to explain within the framework of the coalition as prototype and cue validity. Some authors have argued that an increase in pupil size may be indicative of increased processing efficiency (Kuipers & Thierry, 2011; Aston-Jones & Cohen, 2005) and this may be why we see greater pupil size in the SVO construction for children; they may be investing more resources in efficiently processing those constructions they are most familiar with; however, brain activity and pupil change to these constructions would have to be measured directly to confirm this argument.

We also took the unique opportunity for exploratory analysis investigating the differences in pupil change between adults and children to exactly the same stimuli; this was the first investigation to do such a comparison. Despite the difficulty interpreting the child data, comparing adult and child pupil change seems to provide some clearer insight into the child data. At the first NP we saw no differences in the pupil change slope when comparing the adult and child data in the unambiguous conditions, but we did see differences when looking at the ambiguous conditions (with the adults having a larger pupil change slope). This suggests that adults and children were processing the unambiguous first NP similarly, however it appears that the child group is not investing as much of their resources into the ambiguous constructions (compared to the adults). Additionally, the child group had a larger peak amplitude (approaching significance) compared to the adult group. These data together suggest that the child group was investing more resources into the unambiguous conditions compared to the adults, perhaps this is because they invest their resources in those constructions that they are capable of processing most efficiently; those constructions that are straightforward from the onset.

The data from the comparison between the adult and child group at the second NP also showed that the two unambiguous constructions evoked a greater pupil slope change for the child group than for the adults and that the ambiguous OVS change of pupil slope was larger for adults than children. This is once again, consistent with the children investing more into the straightforwardly marked unambiguous constructions, where word order is apparent from the onset. The peak amplitude and latency analysis showed that the child group had a higher peak amplitude compared to the adult group, however it took longer for the adult group to reach their peak size. This could indicate that overall pupil change is quicker in children (see for example Piquado, Issacowitz, & Winfield, 2010 for an age comparison between younger and older adults) or there are more restraints on pupil change in children. No research has investigated the length of time it takes for the pupil of a child to reach maximum size, clearly it is important that this is investigated further to enable better interpretation of latency results when comparing children and adults.

Across the whole sentence we see a difference between the adults and children; the adult group displayed a larger pupil change slope than the child group in the two OVS conditions and the ambiguous SVO construction. The adult and child group display a similar pattern only in the unambiguous SVO construction, which it is well established in the literature that children of 4;5 years of age should be capable of comprehending. The other three conditions (two OVS conditions and the ambiguous SVO construction) seem to evoke a smaller pupil slope change in the child group, suggesting that they are not investing the same resources during the processing of these constructions.

The results of comparing the data from adults and children seem to suggest that children at the age of 4;5 years are making use of word order and are sensitive to case marking, which is in line with ERP research with this age group (Schipke et al., 2012). At the first NP, they invest processing resources in only those conditions that are clearly marked, however as the sentence unfolds they rely more heavily on word order during processing, which is seen in the increased pupil size at the second NP for the SVO constructions. However, more research with these types of stimuli is needed with

4.5 Conclusion

This study used pupillometry to measure the processing of case marking ambiguity and word order with a group of 4-year-old German children. Behaviorally, there was a chance level performance across all constructions, which was somewhat unexpected, but the relatively long latency between sentence presentation and comprehension task may have led to this poor performance. This revealed that the children were investing more in processing the SVO construction, which we argue was because this was the most familiar. These data seem to support the argument that pupil change is indicating processing efficiency during comprehension (Kuipers & Thierry, 2011; Aston-Jones & Cohen, 2005), and seem to support the ERP data from Schipke (2012) in which children the age of four are mainly relying on word order but are still aware of the case marking. Additionally, these data provide very interesting insights into language processing across age groups using pupillometry; more comparisons across age groups are needed, perhaps coupled with ERP, to expand on these findings.

4.6 References

Aston-Jones, G., & Cohen, J.D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance.*Annual Review Neuroscience, 28*, 401-450.

Bader, M., & Meng, M. (1999). Subject-Object ambiguities in German embedded clauses: An across-the-board comparison. *Journal of Psycholinguistic Research, 28*, 121-143.

Bahlmann, J., Rodriguez-Fornells, A., Rotte, M., & Mnte, T.F. (2007). An fMRI study of canonical and noncanonical word order in German. *Human Brain Mapping, 28(10)*, 940-949.

Bates, D.M., Maechler, M., & Bolker, B. (2012). *lme4: Linear mixed-effects models using S4 classes.* R package version 0.999999-0.

Bates, E., & MacWhinney, B. (1987). Competition, variation, and language learning. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 157193). Hillsdale, NJ: Lawrence Erlbaum.

Bates, E., & MacWhinney, B. (1989). Functionalism and the competition model. In B. MacWhinney & E. Bates (Eds.), *The cross-linguistic study of sentence processing* (pp. 373). New York: Cambridge University Press.

Dittmar, M., Abbot-Smith, K., Lieven, E., & Tomasello, M. (2008). German children's comprehension of word order and case marking in causative sentences. *Child Development, 79*, 11521167.

Friederici, A.D., Steinhauer, K., Mecklinger, A., & Meyer, M. (1998). Working memory constraints on syntactic ambiguity resolution as revealed by electrical brain responses. *Biological Psychology, 47*, 193-221.

Frisch, S., Schlesewsky, M., Saddy, D., & Alpermann, A. (2002). The P600 as an indicator of syntactic ambiguity. *Cognition, 85*, B83-92.

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

Hothorn, T., Bretz, F., Westfall, P., Heilberger, R.M, & Schuetzenmeister, A. (2013). *Simultaneous Inference in General Parametric Models.* R Foundation for Statistical Computing Vienna, Austria.

Just, M.A., & Carpenter, P.A. (1993). The intensity dimension of thought: Pupillometric indices of sentence processing. *Theoretical Issues in Ergenomics Science, 47(2)*, 310-339.

Kuipers, J.R., & Thierry, G. (2011). N400 amplitude reduction correlates with an increase in pupil size. *Frontiers in Human Neuroscience, 5*, 1-5.

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

R Core Team (2012). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria.

Rösler, F., Pechmann, T., Streb, J., Röder, B., & Hennighausen, E. (1998). Parsing of sentences in a language with varying word order: Word-by-word variations of processing demands are revealed by event-related brain potentials. *Journal of Memory and Language, 38*, 150-176.

Schipke, C. S. , Knoll, L.J., Friederici, A.D., & Oberecker, R. (2012). Preschool children's interperation of object-initial sentences: Neural correlates of their behavioural performance. *Developmental Science, 15*, 762-774.

Schlesewsky, M., Bornkessel, I., & Frisch, S (2003). The neurophysicological basis of word order in German. *Brain & Language, 86*, 116-128.

Schlesewsky, M., Fanselow, G., Kliegl, R. & Krems, J. (2000). Preferences for grammatical functions in the processing of locally ambiguous wh-questions in German. In Hemforth, B. & Konieczny, L. (eds.) *Cognitive Parsing in German.* Dortrecht: Kluwer.

Schmidtke, J. (2014). Second language experience modulates word retrieval effort in bilinguals: evidence from pupillometry. *Frontiers in Psychology, 5*, 137.

Zekveld, A.A., Kramer, S.E., Festen, J.M. (2010). Pupil response as an indication of effortful listening: The influence of sentence intelligibility. *Ear and Hearing, 31*, 480-490.

Chapter 5

The previous empirical chapters have all employed pupillometry to investigate the processing of different word orders. In this chapter, the direction is changed slightly to the investigation of word order using a different methodology, scalp-recorded event-related potentials (ERPs), to test competing theories concerning the difficulty that accompanies processing of object relative (OR) clause constructions. In doing so we are attempting to set the stage for further research with the same stimuli directly comparing the relationship between pupil change and brain activity. The relationship between pupil change and the underlying brain mechanisms is unclear and future research investigating the relationship between the two would be valuable and lead to pupillometry becoming a more valid measure.

Exploring Animacy in Object Relative Clauses using ERP

Abstract

Object relative (OR) constructions (*The lawyer that the banker irritated filed for a hefty lawsuit*) are more difficult to process than subject relative (SR) constructions (*The lawyer that irritated the banker filed for a hefty lawsuit*). Processing difficulties associated with ORs have been attributed to the increased working memory demands that accompany holding and assigning thematic roles to the moved object. However, an alternative view is that the less expected object relative construction incurs processing costs based on the indeterminacy of the sentence unfolding. Previous research has shown that the processing costs associated with object relative constructions can be alleviated if one of the noun phrases is inanimate. In the current study we extended this research and tested these two competing accounts by measuring event-related potentials (ERPs) while participants listened to object relative constructions in which the first noun phrase (NP) was either animate or inanimate. Analyses focused on the N400, an ERP component that indices expectation and the P600, linked to memory processes. At the second NP, the N400 was larger for animate object relative compared with inanimate object relative sentences, supporting the view that indeterminacy plays a role in the processing of object relative constructions. However, we found no differences in the P600 at the relative clause verb. Results are consistent with the view that the difficulties associated with object relative constructions are due to an expectancy violation, although it is difficult to rule out a further contribution from memory processes.

5.1 Introduction

Sentences like *The lawyer that irritated the banker filed for a hefty lawsuit* and *The lawyer that the banker irritated filed for a hefty lawsuit* have been the focus of considerable research because of their linguistic structure. The first is known as a subject relative (SR) and the second an object relative (OR). Both sentences include the relative clause (*that irritated the banker/that the banker irritated*); however, in subject relative constructions, the first noun phrase (*the lawyer*) is the subject of the verb (*irritated*) in the relative clause while in the object relative constructions the first noun phrase (*the lawyer*) is the object of the verb in the relative clause (*irritated*), and the subject of the main verb (*filed*). Despite the fact that both constructions have the same words, object relative constructions are more difficult to process than subject relative constructions, as evidenced by measures of comprehension accuracy, reading time, pupil dilation, and Event-Related Potentials (ERPs) (e.g. Ford, 1983; Just et al., 1996; King & Just 1991; Wanner & Mastros, 1978).

Subject relative and object relative constructions are considered filler gap dependencies in which an element (the filler) moves from its original (canonical) position leaving a (phonologically silent but syntactically relevant) trace of itself. In order to comprehend the sentence, the parser must associate the filler with the gap. In the examples below (SR (1) and OR (2)), the filler ($_i$) and the gap have been marked (t_i).

1. The lawyer$_i$ that t_i the irritated the banker filed for a hefty lawsuit.
2. The lawyer$_i$ that the banker irritated t_i filed for a hefty lawsuit

Semantic and pragmatic factors have been shown to influence the processing of object relative constructions; for example, by manipulating the animacy of the noun phrases (NPs) in object relative constructions, processing costs can be alleviated (e.g. Gennari & MacDonald, 2008; Mak, Vonk, & Schriefers, 2002,2006,2008; Traxler, Morris, & Seely, 2002; Traxler, Williams, Blozis, & Morris, 2005; Lowder & Gordon, 2012, 2014). For example, Traxler et al. (2002) investigated object relative and subject relative constructions, manipulating the animacy of the first and second noun phrase as shown in examples (3)-(6) using an eye tracking while reading paradigm.

3. Subject relative- animate NP1: The director that watched the movie

received a prize at the film festival.

4. Subject relative- inanimate NP1: The movie that the director watched received a prize at the film festival
5. Object relative- animate NP1: The director that the movie pleased received a prize at the film festival
6. Object relative- inanimate NP1: The movie that pleased the director received a prize at the film festival.

Eye tracking revealed that object relative processing difficulties were alleviated when the subject was inanimate, compared to when it was animate. Traxler et al. explained the difficulty associated with object relative compared to subject relative constructions, and the mediating effect of pragmatic effects such as animacy, in terms of two competing accounts; that we refer to as the expectation account and the memory account.

5.1.1 Expectation account

According to the expectation account, the difficulties accompanying object relative processing arise because the syntactic parser anticipates sentence continuations based on experience, and when these anticipations are not met, a processing cost is incurred (e.g. Forster, Guerrera, & Elliot, 2009; Gennari & MacDonald, 2008; Hale, 2001; Levy, 2008). Essentially, the more experience one has with a particular structure the easier that structure will be to interpret: As object relative constructions are relatively infrequent, they are harder to process.

Under the umbrella of expectation based accounts, Gennari and MacDonald (2008) argue that indeterminacy (the uncertainty of assigning grammatical roles to input) gives rise to the difficulties associated with object relative constructions. In a first experiment Gennari and MacDonald required participants to complete sentences at different points. This provided a measure of indeterminacy: the likelihood of different semantic interpretations of the construction based on sentence completion at different time points. The authors found more indeterminacy (a greater number of semantic interpretations provided) for the object relative clause when the subject was inanimate (e.g. *the director that the movie*) compared to when the object relative clause had an animate subject (e.g. *the movie that the director*). In their next two experiments they found that it was at the points of high indeterminacy that there were reading comprehension difficulties. The data from these studies support the argument that the indeterminacy of object relative constructions gives rise to processing difficulties compared to subject relative constructions.

5.1.2 Memory Theories

An alternative view places difficulties associated with object relative processing (and the mediating influence of animacy) on the burden that is incurred from holding two noun phrases in working memory during the processing of object relative constructions. This difficulty is proposed to arise in the encoding, storing, retrieving, and subsequently assigning the two NPs grammatical roles upon reaching the verb (e.g.Gibson, 1998; Gordon, Hendrick, & Johnson, 2001; Mak, Vonk, & Schriefers, 2002, 2006, 2008). In object relative constructions the moved element must be held in working memory longer before it can be retrieved and assigned grammatical properties (upon reaching the verb), and it is this integration and retrieval cost that makes object relative constructions more difficult. On this view, animacy effects arise because role assignment is easier in constructions where the sentential subject is inanimate because of the poor thematic fit of an inanimate NP as the object of the verb (Traxler et al., 2002).

Evidence for the memory account comes from a study by Gordon, Hendrick, and Johnson (2001), who found that the difficulty associated with object relative constructions was alleviated when the second NP was the pronoun you (e.g. *The banker that **you** praised climbed the mountain* compared to *The banker that **Bob** praised climbed the mountain*). This was interpreted as reducing the processing resources needed to hold a new discourse referent in working memory (Gibson, 1998). Additionally, they found that mismatching the type of NPs (i.e. proper name and/or job type) alleviated object relative processing costs (e.g. *It was **the barber** that **Bill** saw in the parking lot* was processed better than *It was **the barber** that **the lawyer** saw in the parking lot*).

5.1.3 Event-Related Potentials

In the study reported here, the processing of object relative constructions was investigated using scalp-recorded event-related potentials (ERPs). We were concerned with two sets of ERP components: the N400, and the P600. The N400 is a negative voltage deflection that occurs approximately 300-500 ms after word onset, and is believed to index semantic expectancy within a context: the less expected the word the larger the N400 effect (Kutas & Hillyard, 1980) (i.e. there is a greater negativity during the N400 time window at the last word in the sentence *He spread the warm bread with* ***socks*** compared to *He spread the warm bread with* ***butter***). The P600, sometimes referred to as a late positivity, is a late positive-going voltage deflection occurring approximately 600ms post-stimulus onset, believed to reflect the analysis/reanalysis of syntactic information in working memory (Osterhout et al., 1994); it is larger (more positive) when there is a violation of thematic role expectancy (i.e. there is a greater positivity during the P600 time window at the word ***to*** in the sentence *The broker persuaded* ***to*** *sell the stock was tall* compared to sentences like *The broker persuaded* ***me*** *to sell the stock was tall*).

5.1.4 ERP and Eye Tracking studies of Animacy in object relative constructions

Weckerly & Kutas (1999) investigated the reading of object relative sentences while animacy was manipulated as seen in (7) & (8) below.

7. Inanimate NP1: The poetry that the editor recognized dpressed the publisher of the struggling. . .
8. Animate NP1: The editor that the poetry depressed recognized the publisher of the struggling . . .

A larger negativity was found during the N400 time window at the first NP (*poetry/editor*) when this was inanimate. They argued that this N400 was driven by the unexpectedness of an inanimate first NP, and the difficulty of associating this unexpected element into the context of the sentence. The authors also found a N400 trend at the second NP (*editor/poetry*) with the inanimate NP (*poetry*) once again evoking more negativity. This suggests that encountering an inanimate second NP in the subject position of an object relative construction is less expected; this did not reach significance however.

At the verb in the relative clause (*recognized* in (7)/*depressed* in (8), a positivity during the P600 time window was evoked by the animate first NP constructions. Weckerly & Kutas argue that assigning a human as a subject and an inanimate noun as an object falls in line with typical thematic role assignment, but when this does not prove to be the case (as in 10) there are inherent difficulties that tax working memory and the P600 may reflect this difficulty. Additionally, they found an early anterior negativity resembling a left anterior negativity (LAN) component (between 200-500ms) and a late posterior positivity resembling the P600 component (between 400-700ms) at the main clause verb (depressed in (9)/recognized in (10)) in the inanimate first NP constructions. The LAN occurs approximately 300 ms after word onset and is believed to index the working memory involved in assigning thematic roles (Kluender & Kutas, 1993a; 1993b). Weckerly & Kutas suggested that the P600 at the main clause verb represented the continuation of the previous animacy difficulty as well as the difficulty associated with object relative constructions in general, and the LAN indexed the working memory demands associated with object relative constructions.

However, as Lowder and Gordon (2012; 2014) point out, the Weckerly & Kutas' (1999) stimuli had a potential confound: the words being compared across the conditions are not the same. When comparing the main clause verb across the conditions and the relative clause verb across the conditions, they were comparing two different words (*depressed/recognized*) making it difficult to tease apart whether the differences in the study were the result of the different verbs or from the animacy manipulation itself.

In a more recent ERP study, Metzler and Braun (2013) included object relative sentences in which the first NP was either animate (OR-animate) or inanimate (OR-inanimate), but the second NP was always animate. Importantly, this made OR-animate sentences reversible (Example 9) while OR-inanimate sentences were irreversible (Example 10).

9. OR-animate: The man/[*]who the woman is teaching/is discussing a hard problem/

10. OR-inanimate: The glass/that the man is washing/has a small chip in it/

During the second segment containing the second NP and relative clause verb, the authors found an effect of animacy during the N400 time window, with the OR-inanimate stimuli evoking a greater negativity than the OR-animate stimuli, suggesting that there is an expectancy violation upon encountering an inanimate first NP. At the onset of the third segment (which corresponds the end of the relative clause) they found an effect of animacy in terms of a late

[*]/ indicates the boundary between analysis segments

posterior central positivity, with the OR-animate stimuli evoking a greater positivity than the OR-inanimate stimuli. The authors compared this late positivity to the P600 component and argued that it reflected the increased syntactic demands associated with reversible OR-animate sentences. They ”suggest that the neural generators of the P600 signal in response to violations of syntactic structure and animacy constraints may also be involved in processing normal, nonanomalous sentences that are characterized by increased demands on syntactic processing.” (Meltzer & Braun, p. 12). Overall the data are consistent with memory based accounts but, again, the sentences for the two conditions were not particularly well matched. Very little of the constructions were held consistent, thus making it all the more difficult to tease apart what was driven by the reversibility manipulation and what was driven by the differences between the items themselves.

To combat these potential confounds, the current study used improved stimuli based on an eye-tracking study by Lowder and Gordon (2014) in which the only difference between the conditions was the animacy of the first NP. As can be seen below, both Example (11) and Example (12) are object relative constructions which only differ in the animacy of the first NP.

11. The senator that the journalist criticized accused the governor of embezzling millions of dollars.

12. The article that the journalist criticized accused the governor of embezzling millions of dollars

Lowder and Gordon found that participants spent longer looking at the second NP (*journalist*) and the embedded verb (*criticized*), in the OR-animate condition relative to the OR-inanimate condition. These results suggest that OR-animate constructions are more difficult than OR-inanimate constructions as early as the second NP (before the verb has been encountered). Gordon and Lowder argued that the differences at the relative clause verb reflect the difficulty associated with holding two animate NPs in working memory. The parser must attend more closely to the input when there are two animate NPs (that can easily be confused) to ensure it is accurately encoding and retrieving the correct thematic roles (who is doing what to whom) upon reaching the verb. However, Lowder and Gordon acknowledge that they cannot completely rule out the expectation account based on their study, and it is possible that, both expectancy and memory contribute to object relative difficulties and animacy effects (cf. Stuab, 2010).

5.1.5 Study Aims

In the current study we used Lowder and Gordon's improved stimuli in an ERP experiment aiming to test the expectation based and memory based accounts. We will look at the ERP components to the second NP and the relative clause verb (Gennari & MacDonald, 2008; Lowder & Gordon, 2014). The memory based accounts assume that the difficulty associated with object relative constructions is a result of the encoding, storing, and retrieving of the moved element at the relative clause verb (this is exacerbated when 2 animate NPs are encoded). It therefore predicts a late positivity/P600 at the relative clause verb, signifying syntactic working memory demands for encoding the two animate NPs for OR-animate constructions. In contrast, the expectation account predicts an N400 component at the second NP signifying the unexpectedness of the unfolding object relative construction for the OR-animate constructions.

5.2 Methods

5.2.1 Participants

Eighteen native speakers of English participated in this study, all were right handed, and had normal or corrected to normal vision. Each participant was compensated AU$22.50 for their participation.

5.2.2 Materials

The stimuli were recorded by a female native speaker of Australian English at a normal speaking rate (sampled at 44.1 kHz) at an acoustics laboratory at Macquarie University. There were a total of 72 critical stimuli in each condition, and 108 filler trials. The materials were based on the stimuli used by Lowder & Gordon (2014; see examples 13 and 14), but these were supplemented with new sentences.

13. The speaker that the treasurer consulted led the way to the company's new headquarters. (OR - animate)
14. The map that the treasure consulted led the way to the company's new headquarters. (OR - inanimate)

There were two critical conditions: the object relative animate condition (OR-animate), and the object relative inanimate condition (OR-inanimate). The two conditions were identical apart from the head noun phrase; in the OR-animate this NP was animate, and in the OR-inanimate condition the NP was inanimate. The first NP served as the object of the relative clause verb and the subject of the main clause verb. Fillers consisted of subject relative constructions, garden path sentences, active sentences, and passive sentences. Half the filler sentences began with an animate NP and half began with an inanimate NP. After each sentence the participants were presented a True/False question probing either the action in the relative clause or the action in the main clause; half of the questions were false and half were true. The recordings and corresponding questions were placed into three counterbalanced versions and rotated in a Latin square design, and the order of items was randomized per participant.

5.2.3 Apparatus

Stimulus presentation was programmed using Experiment Builder software. Participants sat a comfortable distance away from the display with a chin rest maintaining their head position. Pupil diameter was recorded with an EyeLink 1000 remote eye tracker, although data are not reported here. EEG was recorded at 1000 Hz with a Synamps II amplifier, using the 28 Ag-AgCL electrodes (FP1 and FP2 electrodes were removed due to the chin rest required for the eye tracking component) attached to an elastic cap (EasyCap) according to the International 10-20 system. EEG recordings were referenced online using the left mastoid, and were re-referenced offline to the average of the left and right mastoid. Impedances were kept to a minimum. Ocular movement was recorded by electrodes placed on the corner, above, and below the left eye.

5.2.4 Design and Procedure

At the start of each trial a fixation cross appeared on the screen. After 2000 ms, the sentence recording was presented aurally through headphones; the fixation cross remained on the screen for the entirety of sentence and for an additional 2000 ms post recording offset. Participants were asked to focus on the fixation cross, attend to the aurally presented sentence, and to try to avoid blinking. After the fixation cross disappeared from the screen a True/False comprehension probe was presented, and participants were asked to judge whether the statement on the screen was true or false based on the sentence they previously heard. All ratings were input by pressing a corresponding T key for True or F key for False on a keyboard.

5.2.5 Data Processing and Analysis

EEG data were analysed using the Neuroscan System offline. Vertical electrooculogram (VEOG) artefacts were filtered and removed using a standard ocular artefact reduction algorithm using Neuroscan. Recordings were subjected to a bandpass filter with cut-off frequencies of 0.01-70 Hz. Data were then epoched between -150-700 ms, to encompass the N400 and P600 time windows. These epochs were triggered at the second NP and the relative clause verb.

Following Weckerly & Kutas (1999) electrodes were grouped in two regions of interest (ROIs) for analysis and were defined as anterior (F7, FT7, FC3, F3, F8, FT8, FC4, F4) and posterior (TP7, FT8, P7, CP3, CP4, P3, P4, P8), additionally they were further divided by left and right hemisphere. The ERP voltage was averaged for each ROI. For each of the components of interest, we calculated the average amplitude within the time windows used in the Weckerly and Kutas (1999) study. The N400 time window was between 300-500ms, and the P600 time window was between 500-700 ms. Time window data were then submitted to ANOVAs using R (R Core Team, 2012) and Tukey's contrasts from the *multcomp* package (Horthorn et al, 2013); we ran a 2x2x2 analysis with the factors type (OR-inanimate/OR-animate), ROI (anterior/posterior), and hemisphere (right/left). Behavioral data were submitted to a one-way ANOVA, with the factor of type (OR-inanimate/OR-animate).

5.3 Results

5.3.1 Second NP (relative clause subject)

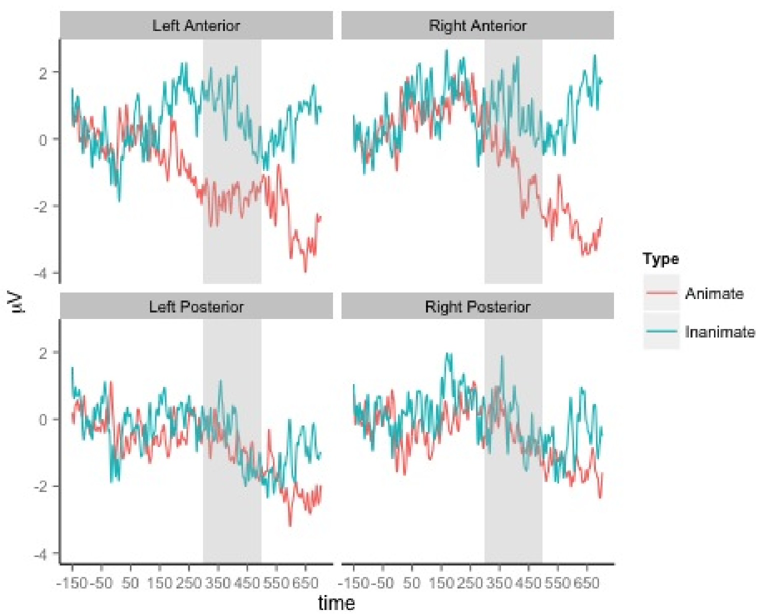

Figure 5.1: Averaged ERPs by ROI for the N400 (grey) time window at the 2nd NP

Figure 5.1 shows the averaged ERPs by ROI at the second NP; the N400 time window (300-500ms) being highlighted in grey. Figure 4 shows the topographical map during the N400 time window for both animate and inanimate conditions. We found a main effect of ROI (F(1,17)= 4.66, p<0.05) with the anterior evoking more negativity than the posterior. Contrary to expectations, there was no main effect of type (F(1,17)= 1.37, p=0.26). There was, however, an interaction between type and ROI (F(1,17)=4.61, p<0.04).

Tukey's contrasts revealed a significant effect of type at the anterior (OR-animate more negative than OR-inanimate) but not posterior ROI. In other words, the expected N400 effect was found but only in the anterior ROI. See Figure 5.2 for ERP topography for the OR-animate condition, and Figure 5.3 for the OR-inanimate condition.

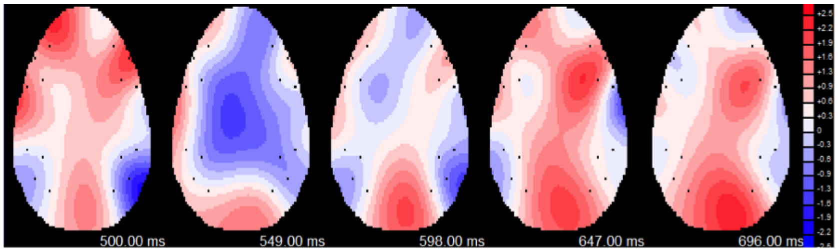

Figure 5.2: ERP topography during the N400 time window at the 2nd NP-Animate

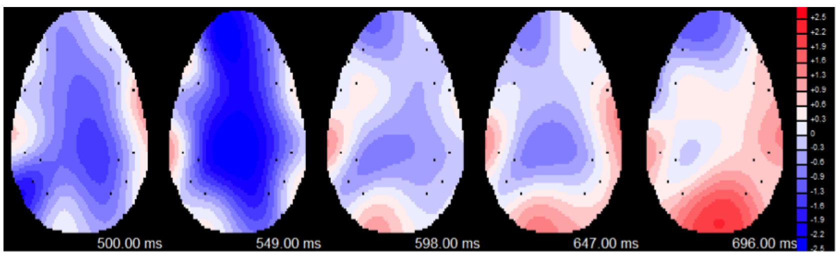

Figure 5.3: ERP topography during the N400 time window at the 2nd NP-Inanimate

5.3.2 Relative Clause Verb

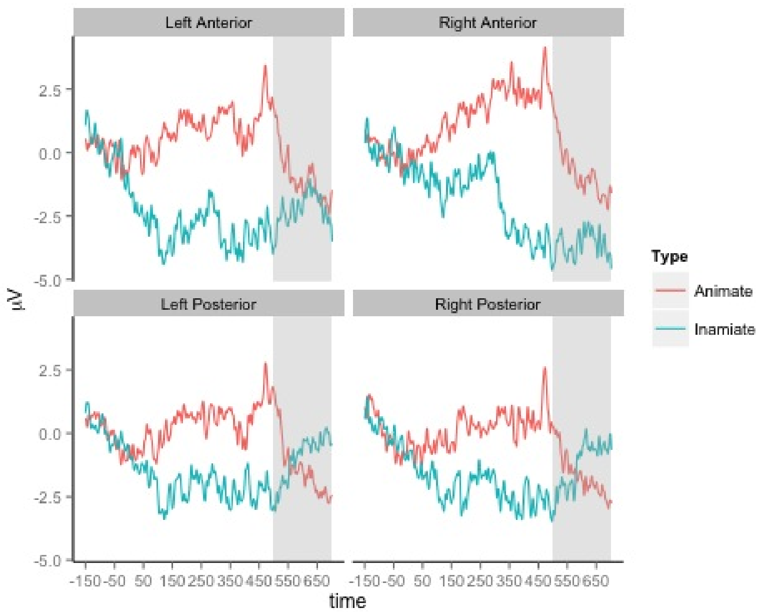

Figure 5.4: Averaged ERPs by ROI for the P600 (grey) time window at the relative clause verb

Figure 5.4 shows the averaged ERPs by ROI at the relative clause verb; the P600 time window (500-700 ms) highlighted in grey. During the P600 time window there were no main effects (all *ps*>0.27) and no interactions (all *ps*>0.35). See Figure 5.5 for ERP topography for the OR-animate condition, and Figure 5.6 for the OR-inanimate condition.

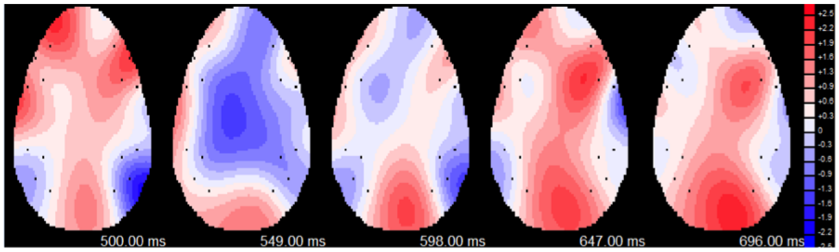

Figure 5.5: ERP topography during the P600 time window at the relative clause verb-Animate

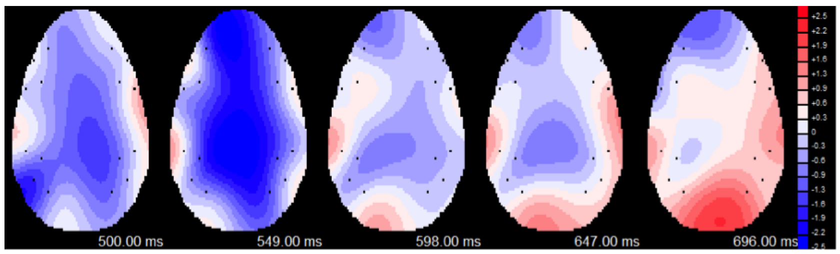

Figure 5.6: ERP topography during the P600 time window at the relative clause verb-Inanimate

5.4 Discussion

In this study, we investigated the processing of object relative constructions with the animacy of the first NP manipulated. Previous research was confounded by stimuli that did not match across conditions, making it difficult to tease apart whether animacy or stimulus differences were driving the observed effects. Participants were asked to answer a True/False question probing the interpretation of the sentence they had heard. Comprehension accuracy was above 85% for both stimuli types, but the OR-inanimate conditions were answered significantly more accurately than the OR-animate conditions. This is in line with previous research that has shown OR-inanimate comprehension is more accurate than OR-animate (e.g Lowder & Gordon, 2014) and, more generally, that object relative constructions with an inanimate and animate NP are easier to comprehend than object relative constructions with two animate NPs (e.g. Metzler & Braun, 2013; Weckerly & Kutas, 1999).

ERPs to two critical words were investigated in order to test competing theories of object relative processing. Expectation based accounts assume that the difficulty that arises from object relative constructions has to do with the expectation of how the sentence will unfold based on what has come before. While another noun phrase is an unexpected continuation from a first noun phrase, the argument is that animacy of this second NP also influences the expectation. Thus, when, in a sentence like *The speaker that the* *treasurer consulted led the way to the company's new headquarters*, the parser encounters the animate first NP (*The speaker that...*), it is less expected that this would be followed by another animate NP (e.g. *the treasurer...*) and this unexpected continuation is the locus of processing difficulties in object relative constructions. When the first NP is inanimate (as in the OR-inanimate constructions, e.g., *The map that the treasurer consulted led the way to the company's new headquarters*), an animate second NP is not as unexpected. Consequently, the OR-inanimate conditions are easier to process than OR-animate conditions. Thus, in terms of the ERP data, expectation based accounts predicted an N400 when comparing the conditions at the second NP (signifying the unexpected second animate NP).

Consistent with this prediction, we found that the OR-animate constructions evoked a more negative response at anterior electrodes during the N400 time window, compared to the OR-inanimate constructions. These data suggest that there is an unexpectancy cost associated with OR-animate constructions, and upon encountering a second animate NP the parser experiences a greater violation of expectation. This is in line with previous ERP research that argues the difficulties associated processing OR-animate constructions arise from an expectancy violation (e.g. Gennari and MacDonald, 2008). Additionally, using an eye-tracking paradigm with the same type of stimuli as the current study, Lowder and Gordon (2014) found longer looking at the second NP for the OR-animate relative to the OR-inanimate constructions. The current data, coupled with the previous research, suggests that there is an expectancy violation at the 2nd NP that causes processing difficulties with object relative constructions that have two animate NPs.

In contrast to the expectation account, memory based accounts assume that object relative processing difficulty is due to the working memory demands of encoding, storing, and retriving the two NPs and subsequently assinging them correct roles. When the parser encounters two animate NPs (as in the OR-animate condtions) these NPs tax working memory, given they are more easily confused (than one animate and one inanimate NP), and therefore more resources are needed to keep track of who is doing what to whom. In terms of the ERPs, memory based acounts predicted a P600 at the relative clause verb, signifying the working memory demands associated with the encoding and storage of two animate NPs.

Unexpectedly, we found no differences between the conditions within the

P600 window, which suggests that there are not measurable working memory differences when assigning thematic roles at the verb between OR-animate and OR-inanimate constructions. This finding was quite unexpected, because previous research across different modalities has found difficulty at the relative clause verb in object relative constructions (e.g. King & Kutas, 1995; Metzler & Braun, 2013; Lowder & Gordon, 2014; 'Müller, King, & Kutas, 1997; Weckerly & Kutas, 1991). In the current study it seems as though difficulty arises at the second NP, but this does not affect the assignment of thematic roles at the relative clause verb. It is possible that the difficulties that were experienced at the second NP have been resolved, making thematic role assignment straightforward upon reaching the verb.

As noted above, it may be that previous research has been confounded with differences across conditions, making it difficult to interpret the cause of the animacy difference in the object relative constructions (differences in the items themselves, or differences due to experimental manipulation). In the current study we used improved stimuli, with only the animacy of the first NP manipulated, thus avoiding the confounds affecting previous studies. It is also possible that the lack of findings at the relative clause verb are a result of the preceeding animacy effect from the 2nd NP carrying over to the verb and confounding potential differences. The unexpectedness at the 2nd NP in the OR-inanimate conditions (as seen in the form of an N400) may carry over to the subsequent word, thus affecting interpretation of the brain activity.

The data here seem to support the expectancy account of processing difficulties associated with object relative constructions. However, we cannot entirely rule out memory based accounts (Lowder and Gordon, 2014, also report this problem), and it is possible that there is some combination of both memory and expectancy processes contributing (e.g. Staub, 2010). There was a clear N400 effect at the second NP for the OR-animate constructions, suggesting that there is an unexpectancy upon reaching an animate 2nd NP after an animate 1st NP. The lack of differences makes interpretation at the critical relative clause verb more difficult, but the preceding animacy effect at the 2nd NP may have confounded the interpretation of this segment.

5.5 Conclusion

In this study we tested competing theoretical explanations of the processing difficulties associated with object relative constructions. Comprehension accuracy was higher for OR-inanimate constructions than for OR-animate constructions, suggesting that object relative constructions with two animate NPs are more difficult to comprehend than those with one animate and one inanimate NP. In terms of theoretical implications the ERP data from the 2nd NP suggest that difficulties arise due to (un)expectancy (N400). No differences were found at the relative clause verb, and it may be that the difficulties (and potential role revision) at the second NP makes the role assignment at the verb straightforward (thus causing no differences between OR-animate and OR-inanimate conditions) or that there were carry over effects from the 2nd NP impeding the interpretation of results during the relative clause verb; however more research needs to investigate these improved object relative constructions using ERP. Overall, the data suggest that difficulties arising from object relative constructions are due to expectancy violations.

5.6 References

Bates, D.M., Maechler, M., & Bolker, B. (2012). *lme4: Linear mixed-effects models using S4 classes*. R package version 0.999999-0.

Ford, M. (1983). A method for obtaining measures of local parsing complexity throughout sentences. *Journal of Verbal Learning and Verbal Behavior 22*, 203218.

Forster, K. I., Guerrera, C. & Elliot, L. (2009). The maze task: Measuring forced incremental sentence processing time. *Behavioral Research Methods, 41*, 163-171.

Gennari, S.P., & MacDonald, M.C. (2008). Semantic indeterminacy in object relative clauses. *Journal of Memory and Language, 58*, 161-187.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition, 68*, 1-76.

Gordon, P.C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory and Cognition, 27*, 1411-1423.

Hale, J. (2001). A probabilistic early parser as a psycholinguistic model. In Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics (pp. 159-166).

Just, M.A., & Carpenter, P.A. (1993). The intensity dimension of thought: Pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology, 47*, 310-339.

Just, M. A., Carpenter, P. A., Keller, T. A., Eddy, W. F., & Thulborn, K. R. (1996). Brain activation modulated by sentence comprehension. *Science, 274*, 114-116.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language, 30*, 580-602.

King, J. W. & Kutas, M. (1995). Who did what and when? Using wordand clause level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience, 7*, 376-395.

Kluender, R. and Kutas, M. (1993a). Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience, 5*, 196-214.

Kluender, R. and Kutas, M. (1993b). Subjacency as a processing phenomenon. *Language and Cognitive Processes, 8*, 573-633.

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

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106*, 1126-1127.

Lowder, M.W., & Gordon, P.C. (2014). Effects of animacy and nounphrase relatedness on the processing of complex sentences. *Memory and Cognition, 42*, 794-805.

Lowder, M.W., & Gordon, P.C. (2012). The pistol that injured the cowboy: Difficulty with inanimate subject-verb integration is reduced by structural separation. *Journal of Memory and Language, 66*, 819-832.

Mak, W.M., Vonk, W., & Schriefers, H. (2006). Animacy in processing relative clauses: The hikers that rocks crushclause processing. *Journal of Memory and Language, 54*, 466-490.

Mak, W.M., Vonk, W., & Schriefers, H. (2008). Discourse structure and relative clause processing. *Memory and Cognition, 36*, 170-181.

Mak, W.M., Vonk, W., & Schriefers, H. (2002). The influence of animacy on relative clause processing. *Journal of Memory and Language, 47*, 50-68.

Meltzer, J.A., & Braun, A.R. (2011). An EEG-MEG dissociation between online syntactic comprehension and posthoc reanalysis. *Frontiers in Human Neuroscience, 5*, 1-15.]

Meltzer, J.A., & Braun, A.R. (2013). P600-like positivity and left anterior negativity responses are elicited by semantic reversibility in nonanomalous sentences. *Journal of Neurolinguistics, 26*, 129-148.

Meltzer, J.A., McArdle, J.J., Schafer, R.J., & Braun, A.R. (2010). Neural aspects of sentence comprehension: syntactic complexity, reversibility, and reanalysis. *Cerebral Cortex, 20*, 1853-1864.

Müller, H.M., King, J.W., & Kutas, M. (1997). Eventrelated potentials elicited by spoken relative clauses. *Cognitive Brain Research, 5*, 193-203.

Osterhout, L., & Holcomb, P. J., & Swinney, D. A. (1994).Brain potentials elicited by garden path sentences: Evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20*, 786-803.

R Core Team (2012). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria.

Ruchkin, D.S., Johnson, R., Canoune, H., & Ritter, W. (1990). Short-term memory storage and retention: An event-related potential study. *Electroencephalography and Clinical Neuroscience, 76*, 419-439.

Staub, A. (2010). Eye movements and processing difficulty in object relative clauses. *Cognition, 116*, 71-86.

Sabourin, L., & Stowe, L. (2004). Memory effects in syntactic ERP tasks. Brain and Cognition, 55, 392-395

Steinhauer, K., & Drury, J.E. (2012). On the early left-anterior negativity (ELAN) in syntax studies. *Brain and Language, 120*, 135-162.

Traxler, M.J., Morris, R.K., & Seely, R.E. (2002). Processing subject and object relative clauses: Evidence from eye-movements. *Journal of Memory and Language, 47*, 69-90.

Traxler, M.J., Williams, R.S., Blozis, S.A., & Morris, R.K. (2005). Working memory, animacy, and verb class in the process of relative clauses. *Journal of Memory and Language, 53*, 204 224.

Wanner, E.,& Maratsos, M. (1978). An ATN approach in comprehension. In: M. Halle, J. Bresnan, & G. Miller (Eds.), *Linguistic Theory and Psychological Reality* (pp. 119-161). Cambridge, MA: MIT Press.

Weckerly, J. & Kutas, M (1999). An electrophysiological analysis of animacy effects in the processing of object relative sentences. *Psychophysiology, 36*, 559-570.

Chapter 6

Conclusions

In this thesis sentence processing was investigated using a psychophysiological measure known as pupillometry as well as Event-Related Potentials (ERP). The scope of the thesis was broad, investigating the processing of several different movement constructions with native speakers of English and second language learners of English, as well as word order and case marking in German speaking adults and children. Pupillometry and ERP allowed us to test competing linguistic theories and use novel methodologies to investigate the processing of word order. In doing so we also aimed to establish pupillometry as an effective way to investigate the processing of word order thus broadening the methodological spectrum. In the first chapter, I provided a thorough review of pupillometry, and the seminal studies that adhered to Kahneman's (1973) criteria which established pupillometry as an effective psychophysiological measure. Research into the syntactic constructions of interest was also reviewed, providing a solid foundation with which to base the empirical studies in this thesis. Below, I revisit the findings of each empirical study and state the implications. The chapter will conclude with potential future directions.

6.1 L1/L2 intermediate gap processing & the SSH

Chapter 2 compared the processing of English intermediate gap constructions, by both native speakers of English and proficient L2 speakers of English (that were native speakers of German). Intermediate gap constructions are considered filler gap dependencies; they contain an intermediate gap which is believed to reactivate the filler at a clausal boundary thus facilitating the later integration of filler and gap. Clahsen and Felser (2006) put forward the Shallow Structure Hypothesis (SSH), which argues that L2 speakers rely on lexical, pragmatic, and real world knowledge during comprehension in their non-native language, causing them to build shallow syntactic representations, unlike native speakers. Intermediate gap structures have been used to investigate the SSH, and while some research has shown evidence for the SSH (e.g. Marinis, Roberts, Felser, & Clahsen, 2005) other research has not (e.g. Dekydtspotter, Schwartz, & Sprouse, 2006; Pliatsikas & Marinis, 2013).

In the empirical study in chapter 2, with native speakers of English and proficient speakers of English (native German speakers), I investigated the processing of the same constructions used by Marinis et al. However, I used auditory rather than written presentation, with a novel methodology, pupillometry. I compared conditions with an intermediate gap (t'_i, e.g., The manager who_i the consultant claimed t'_i that the new proposal had pleased t_i will hire five workers tomorrow) or without an intermediate gap (e.g. The manager who_i the consultant's claim about the new proposal had pleased t_i will hire five workers tomorrow). Pupillometry at the site of the intermediate gap (t'_i) revealed no differences between conditions or between the two language groups, suggesting that both groups processed the intermediate gap similarly. At the site of the final gap (t_i), both groups showed facilitation (in terms of a greater pupil slope) at the gap site in the condition that had an intermediate gap compared to the condition that did not. While the native group showed greater pupil change at this segment compared to the L2 group, the pattern of results from both groups showed remarkable similarity, suggesting that the groups were processing these constructions in the same way, contrary to the SSH.

Understanding the difficulties that arise during second language processing is imperative to understand second language acquisition, and in turn L2 instruction. This study provided evidence that L2 speakers of English are capable of rich and complex syntactic processing in their non-native language, given the evidence for facilitated processing at the final gap site like that of native speakers. Thus, Chapter 2 provides evidence against the SSH using a novel methodology, and adds to the growing body of research indicating that L2 speakers may be able to comprehend their non-native language similarly to native speakers. Further research with complementary methodologies will lead to a more complete understanding of the way in which second languages are processed

6.2 Pupillometry and word order and case marking processing

In Chapters 3 and 4, I investigated German word order and case marking in adults and children using pupillometry. In German, nominative and accusative case are marked on the article in the simple transitive, and while the preferred word order is SVO, this case marking renders word order to be a less important in order to correctly assign thematic roles. However, in German, it is only in the masculine gender that nominative and accusative forms are unambiguously marked; the feminine and neuter forms, on the other hand, are ambiguous as they are the same in both the nominative and accusative form. In Chapters 3 and 4 word order (SVO/OVS) and ambiguity (ambiguous NP1/unambiguous NP1) were manipulated to investigate the processing of these constructions (see Table 6.1).

Table 6.1: Example sentences with case marking and word order manipulation

	Unambiguous NP1	Ambiguous NP1
SVO	**Der Uhu** kitzelt gleich das Meerschwein	Das Meerschwein kitzelt gleich **den Uhu**
	The owl $_{nom}$ tickles the guinea pig$_{acc}$	The guinea pig$_{nom}$ tickles the owl$_{acc}$
	The owl tickles the guinea pig	The owl tickles the guinea pig
OVS	**Den Uhu** kitzelt gleich das Meerschwein	Das Meerschwein kitzelt gleich **der Uhu**
	The owl$_{acc}$ tickles the guinea pig$_{nom}$	The guinea pig$_{acc}$ tickles the owl$_{nom}$
	The guinea pig tickles the owl	The guinea pig tickles the owl

Adult German speakers show high accuracy in comprehension for all of the different conditions (e.g. Bader & Meng, 1999; Schlesewsky, Fanselow, Kliegl & Krems, 2000; Schriefers, Friederici, & Kühn, 1995), but previous research had found that it was not until around the age of 4 that German speaking children are able to interpret subject-first constructions reliably (Dittmar et al., 2008; Grünloh et al. 2011; Bates & MacWhinney 1987, 1989; Schipke et al., 2012), and it is not until approximately 7 years of age that object-first constructions are interpreted with high accuracy (e.g. Dittmar et al., 2008).

In the empirical study in Chapter 3, I investigated comprehension by adults of German constructions with word order and ambiguity manipulated using pupillometry with the aim of broadening the validity of pupillometry. Additionally, I aimed to test competing theories that involve the processing of the first NP in an unambiguously marked NP object-first construction. Some authors have suggested that the difficulty that occurs with an unambiguous object-first construction is the result of working memory demands (e.g. Matzke et al., 2002; Rösler et al., 1998), while others have attributed the difficulty to the violation of the expectation of canonicity (e.g. Schlesewsky et al., 2003). I chose pupillometry to test these two competing theories as it is a measure of cognitive processing costs, and may correlate with brain activity. Additionally, testing these well-researched constructions using a novel methodology provided insight into what the pupil reveals about processing, and strengthened the foundation of pupillometry.

I found no differences in cognitive processing costs as measured by pupillometry at the first NP across all construction types and concluded that that the difficulties that arise when understanding object-first constructions were not the result of working memory demands. As Matzke et al. (2002) stated, items (NPs) being stored in working memory should be apparent in neural responses (or in this case pupillary responses) at the first NP when comparing object-first constructions given that the parser must hold the accusatively marked object in working memory as the sentence unfolds. Hence, I concluded that the difficulties associated with object-first constructions previously found at the first NP are most probably the result of a violation of expectancy.

I also analyzed pupil change slope, peak amplitude, and peak latency at the second NP and across the whole sentence. At the second NP the unambiguous SVO constructions evoked the smallest pupil change slope suggesting that the other three constructions (ambiguous SVO, unambiguous OVS, ambiguous OVS) that were not straightforwardly and unambiguously marked from the onset, caused processing difficulties at this point. The increased pupil size for the ambiguous OVS compared to the unambiguous SVO construction suggested that the parser was reanalyzing upon reaching the second NP, from the original and preferred SVO order to the OVS order. No differences between the two OVS constructions were found, suggesting, once again, that difficulties at the second NP were not arising due to working memory constraints (which would be greater for the ambiguous OVS than the unambiguous OVS). Interestingly, a difference was found at the second

NP between the two SVO constructions, suggesting that there were some processing costs for the ambiguously marked construction, even when the disambiguating second NP confirmed the SVO order.

Across the whole sentence I found the OVS constructions and ambiguous constructions evoked larger pupil change than the SVO constructions and the ambiguous constructions, suggesting that OVS constructions required more processing resources, as did ambiguous constructions. Additionally, the peak amplitude for the ambiguous conditions was higher than for the unambiguous conditions.

The data from this study also fall in line with previous research; object-first constructions require more processing resources than subject-first constructions. It also suggests that holding a noun phrase in working memory before thematic roles are assigned taxes working memory compared to those sentences where word order is apparent from the onset.

The data from the current study allows us to shed further light on the results from the ERP research investigating these types of constructions (e.g. Matzke et al., 2002; Rösler, et al., 1998; Schlesewsky, et al., 2003), and suggests that the ERP modulations are caused by canonicity violations, and not working memory demands. Chapter 3 not only gives us a greater understanding of the processing of these constructions but this study also built a solid foundation to conduct the same research with children (Chapter 4).

6.3 The processing of case marking and word order in German speaking children

In Chapter 4 the same stimuli and methodology from the previous study were used (see Table 5.1), but I tested comprehension and processing with a group of 4;5-year-old German speaking children. Previous research had found that German speaking children around the age of 4 years could understand subject-first constructions, but were at chance for object-first constructions (e.g. Dittmar et al., 2008; Grnloh et al. 2011; Bates & MacWhinney 1987, 1989; Schipke et al., 2012). ERP research revealed that there were difficulties for this age group with integrating the second NP into unambiguous OVS constructions, reflecting the costs associated with assigning thematic roles to this object-first construction. In Chapter 4 I used pupillometry in an attempt to find further insight into these differences.

In a comprehension task, the 4;5-year-olds performed at chance across all conditions, which was in contrast to previous research which predicted that these children should, at least, have been able to respond correctly to the unambiguous SVO sentences. We suggested that perhaps the relatively long gap between the end of the sentence and the comprehension response may have added to the difficulties with performance or played a role in the answering heuristic. Interestingly, comprehension performance in the second half of the study was higher for the OVS constructions than in the first half, and was higher for OVS than SVO constructions. Clearly, more research needs to be conducted to further explore the reasons for this surprisingly poor performance.

At the first NP we found that the ambiguous OVS construction evoked the smallest pupil change slope (compared to the three other conditions), suggesting that at this age the participants are investing only in those constructions that are either in the preferred SVO order or are clearly marked at the onset (unambiguous conditions). This therefore suggests that this age group was sensitive to both case marking (to determine that a condition was unambiguous) and word order (a preference for SVO interpretation). At the second NP we found a smaller pupil change for ambiguous OVS compared to the two unambiguous conditions. This indicates that these children invested more resources at the 2nd NP in those unambiguous conditions where the case marking was reconfirming the word order that had been apparent at the outset of the sentence (NP1). Across the whole sentence there was an effect of word order approaching significance with the SVO conditions evoking a larger pupil change slope than the OVS. The data from the whole sentence, again, suggests that the participants were investing in those conditions that were in the typical and straightforward SVO conditions.

The current data is in line with Schipke (2012) who measured ERP with auditory presentation, and found that children the age of 4 were able to detect case marking but were still mainly relying on word order during comprehension. Grünloh, Lieven, and Tomasello (2011) found that 4-year-old children make use of intonation during processing of OVS sentences, and it is possible that the auditory presentation for the study had some impact on the somewhat counterintuitive results; more research exploring auditory presentation and pupillometry with children is needed.

I also directly compared the adult and child data. At the first NP, ambiguous conditions evoked a larger pupil change slope in the adults compared to the children, and the peak amplitude was smaller for the child group than the adult group (approaching significance). This was in line with the pattern found when the children's data was analyzed separately in which they were found to be investing more resources in the constructions where word order was apparent from the onset

At the second NP there was a greater pupil change slope for the child group in the unambiguous conditions while the adult group showed a larger pupil change slope for the ambiguous OVS condition. In other words, the children, again, invested more in those constructions that reconfirmed the word order, while adults invested more in those constructions that were confirmed at this point to be in the less typical OVS constructions. The child group also showed a larger peak amplitude and a smaller peak latency compared to the adult group, showing the differences in investment of resources and the time it took to reach the peak of investment between the two groups. Across the whole sentence, the adult group had a greater pupil size for all the conditions except the unambiguous SVO condition, and there were no differences in terms of peak size or latency. This suggests that the only condition that the two groups processed similarly was the unambiguously marked and preferred SVO order.

The child data is clearly quite different from the adult data; children seem to be investing more of their resources into the simpler SVO constructions, however this does not seem to be mirrored in behavioral accuracy. Together with the individual data it seems that children at the age of 4.;5 invest in those constructions that fall in the SVO order. This seems to be in line with proposals that pupil change may be indicative of efficient processing (Kuipers & Thierry, 2011; Aston-Jones & Cohen, 2005). However given that most previous research points to pupil change as an indicator for processing cost (e.g. Kahneman & Beatty, 1966; Hess & Polt, 1964), clearly there is need for more research to discriminate between these accounts.

By comparing adult and child pupil change to the same stimuli I utilized a unique opportunity to investigate differences between the two groups; to my knowledge this was the first study to do that. As discussed, it gave greater insight into the child data, though more research is needed to investigate pupil response differences between adults and children to the same stimuli further. In terms of peak latency and peak size we found some differences between the two groups: it may be that pupil change is delayed in adults or it may indicate differences in processing between the two groups, either way more research is needed investigating age differences in pupillometry to validate these findings.

6.4 Animacy and object relative clause processing

In Chapter 5, I used ERPs to test the processing of object relative clauses while manipulating the animacy of the first NP. Given the somewhat opaque link between pupil change and brain activity, it is important that research aims to relate the two, in order to better understand what the pupil is telling us about underlying brain activity. In this chapter I set the foundation of an ERP study investigating object relative clauses with the aim of future research directly relating the two methodologies.

Object relative clauses are consistently found to be more difficult to comprehend than subject relative clauses (e.g. Wanner & Mastros, 1978), and it has also been found that using an inanimate NP in an object relative construction can alleviate some of this difficulty (e.g. Gennari & MacDonald, 2008; Mak, Vonk, & Schriefers, 2002,2006,2008; Traxler, Morris, & Seely, 2002; Traxler, Williams, Blozis, & Morris, 2005; Lowder & Gordon, 2012, 2014). Two main theories have been put forward to explain the processing difficulty that accompanies object relative clauses: the first is expectation based and the second is working memory based. The expectation based accounts assume that the difficulties that accompany object relative constructions are the result of the indeterminacy of the sentence unfolding, while the working memory based accounts assume the difficulties are the result of holding and assigning roles to the two NPs which in turn taxes working memory.

Previous studies investigating object and subject relative clauses have been confounded by the fact that the stimuli were not well matched across conditions, making it difficult to tease apart whether differences were caused by the manipulation or by the different words themselves. Gordon and Lowder (2014) used eye tracking to investigate the effects of animacy by manipulating only the animacy of the first NP, while keeping the rest of the constructions consistent across stimuli; thus avoiding the confound of the previous research (e.g., *The speaker (OR-animate)/map (OR-inanimate) that the treasurer consulted led the way to the company's new headquarters*). Hence, in Chapter 5, I was the first to use ERPs to investigate brain responses to object relative (OR) constructions using stimuli which were appropriately matched (based on Gordon and Lowder's stimuli from the eye-tracking study). I was then able to test the two theories that have been proposed to account for the difficulties associated with object relative constructions.

Behaviorally, the results supported previous research (e.g Lowder & Gordon, 2014): object relatives were answered with high accuracy, but the OR-inanimate constructions were answered more accurately than the OR-animate constructions. To test the competing theories I looked at ERPs to the second NP and the relative clause verb and found an anterior negativity for the OR-animate condition compared to the OR-inanimate condition during the N400 time window. As this component is believed to index unexpectedness (Kutas & Hillyard, 1980), this finding supported the view that difficulties associated with object relative constructions are expectation based: upon reaching the second animate NP in an OR-animate construction expectations are violated.

At the relative clause verb, no differences between the two conditions were found in the P600 window, contrary to previous research. Weckerly and Kutas, 1999) argued that the larger P600 for the OR-animate conditions reflected difficulties recalling and assigning the two NPs, given that both animate NPs are equally as likely to take the subject or object role (e.g. Metzler & Braun, 2013; Weckerly & Kutas, 1999). However, our data suggest that their findings may instead have reflected the confounds in their stimuli (e.g. King and Kutas; Meltzer & Braun; Weckerly & Kutas). Nonetheless, it remains possible that both unexpectedness and working memory cause difficulties processing object relative constructions (Staub, 2010).

6.5 Overall Conclusions

Establishing who is doing what to whom is an essential part of language comprehension, and one of the central goals of psycholinguistics is to understand the mechanisms that allow linguistic comprehension. Word order is one potential cue to establish who is doing what whom, and in this thesis word order was investigated to garner more information about the processing of different word orders in different languages (and with different participant groups).

The research I have carried out and presented in this thesis has provided insight into the processing of different types of word orders in different languages, and with different participant groups. I used pupillometry and ERPs to explore linguistic theories, and adjudicate between these linguistic theories using novel methodology and improved stimuli. In this thesis I tested competing linguistics theories that have been debated based on findings from, for example, offline paradigms, self paced reading paradigms, and ERP. While these methods provided valuable insights into the movement structures in question, there still remained some debate, and pupillometry allowed us to use a novel methodology to shed some light into these theories. We found evidence against the SSH (Chapter 2), in support of expectation based accounts of word order difficulties in German (Chapter 3), and ERPs provided evidence for expectation based accounts of the difficulties associated with object relative constructions (Chapter 5) (thus setting the stage for future research to investigate the relationship between pupil change and brain activity).

Not only did pupillometry allow us to test competing theories in adult language processing, but it also provided us with interesting insights into language processing of children (Chapter 4), and is the first study to investigate the differences in pupil response between adults and children. This comparison allowed us to garner new insight into the child data, and investigate the differences that arise in terms of pupil slope, peak amplitude, and peak latency. Additionally, the results from this thesis broaden the scope of pupillometry research and provide more validation to its usefulness (Chapters 2-4). Pupil diameter change is easily measured with an eye-tracking machine, and as this thesis shows, it provides an effective means to measure processing costs associated with many linguistic constructions.

6.6 Future Directions

While pupil diameter change is increasingly used as a psychophysiological method to measure processing costs there has been little empirical research investigating what pupil diameter change represents. To say that pupil diameter change is an effective method of measuring ”processing cost”, ”processing load”, ”cognitive effort”, etc. gives no real insight into what underlying brain activity pupil change may be linked to. In the future more research needs to investigate both brain activity and pupil change. This will in turn allow us to test theories of language (and other fields) more specifically, and will provide greater and more focused insight into the questions at hand. If specific ERPs (e.g. N400, P600, etc.) are found to correlate with pupil change it will allow us to draw more conclusions from studies that have, or will use pupillometry. In this thesis, I demonstrated the sensitivity of pupillometry to a variety of linguistic constructions, across many participant groups. In addition, I discussed pupil change slope in the context of brain activity, and conducted an ERP study that could easily also be undertaken with pupillometry. Thus, this thesis has not only contributed to the body of knowledge in this field but also set the stage for future research.

6.7 References

Bader, M., & Meng, M. (1999). Subject-Object ambiguities in German embedded clauses: An across-the-board comparison. *Journal of Psycholinguistic Research, 28*, 121-143.

Bates, E., & MacWhinney, B. (1987). Competition, variation, and language learning. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 157193). Hillsdale, NJ: Lawrence Erlbaum.

Bates, E., & MacWhinney, B. (1989). Functionalism and the competition model. In B. MacWhinney & E. Bates (Eds.), *The cross-linguistic study of sentence processing* (pp. 373). New York: Cambridge University Press.

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics, 27*, 3-42.

Dekydtspotter, L., Schwartz, B.D., & Sprouse, R.A. (2006). The comparative fallacy in L2 processing research. In M.G. O'Brien, C. Shea, & Archibald, J. (Eds.) , *Proceedings of the 8th Generative Approaches to Second Language Acquisition Conference* (pp. 33-40). Summerville, MA: Cascadilla Proceedings Project.

Dittmar, M., Abbot-Smith, K., Lieven, E., & Tomasello, M. (2008). German children's comprehension of word order and case marking in causative sentences. *Child Development, 79*, 11521167.

Gennari, S.P., & MacDonald, M.C. (2008). Semantic indeterminacy in object relative clauses. *Journal of Memory and Language, 58*, 161-187.

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

Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science, 143*, 1190-1192.

Kahneman, D. (1973). *Attention and Effort.* Englewood Cliffs, N.J.: Prentice-Hall.

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science, 154*, 1583-1585.

King, J. W. & Kutas, M. (1995). Who did what and when? Using wordand clause level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience, 7*, 376-395.

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

Lowder, M.W., & Gordon, P.C. (2014). Effects of animacy and nounphrase relatedness on the processing of complex sentences. *Memory and Cognition, 42*, 794-805.

Lowder, M.W., & Gordon, P.C. (2012). The pistol that injured the cowboy: Difficulty with inanimate subject-verb integration is reduced by structural separation. *Journal of Memory and Language, 66*, 819-832.

Mak, W.M., Vonk, W., & Schriefers, H. (2006). Animacy in processing relative clauses: The hikers that rocks crushclause processing. *Journal of Memory and Language, 54*, 466-490.

Mak, W.M., Vonk, W., & Schriefers, H. (2008). Discourse structure and relative clause processing. *Memory and Cognition, 36*, 170-181.

Mak, W.M., Vonk, W., & Schriefers, H. (2002). The influence of animacy on relative clause processing. *Journal of Memory and Language, 47*, 50-68.

Marinis, T., Roberts, L., Felser, C., & Clahsen, H. (2005). Gaps in second language sentence processing. *Studies in Second Language Acquisition, 27*, 53-78.

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

Meltzer, J.A., & Braun, A.R. (2013). P600-like positivity and left anterior negativity responses are elicited by semantic reversibility in nonanomalous sentences. *Journal of Neurolinguistics, 26*, 129-148.

Osterhout, L., & Holcomb, P. J., & Swinney, D. A. (1994).Brain potentials elicited by garden path sentences: Evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20*, 786-803.

Pliatsikas, C., & Marinis, T. (2013). Processing empty categories in second language: When naturalistic exposure fills the (intermediate) gap. *Bilingualism: Language and Cognition, 16*, 167-182.

Rösler, F., Pechmann, T., Streb, J., Röder, B., & Hennighausen, E. (1998). Parsing of sentences in a language with varying word order: Word-by-word variations of processing demands are revealed by event-related brain potentials. *Journal of Memory and Language, 38*, 150-176.

Schipke, C. S. , Knoll, L.J., Friederici, A.D., & Oberecker, R. (2012). Preschool children's interperation of object-initial sentences: Neural correlates of their behavioural performance. *Developmental Science, 15*, 762-774.

Schlesewsky, M., Bornkessel, I., & Frisch, S (2003). The neurophysicological basis of word order in German. *Brain & Language, 86*, 116-128.

Schlesewsky, M., Fanselow, G., Kliegl, R. & Krems, J. (2000). Preferences for grammatical functions in the processing of locally ambiguous wh-questions in German. In Hemforth, B. & Konieczny, L. (eds.), *Cognitive Parsing in German.* Dortrecht: Kluwer.

Schriefers, H., Friederici, A.D., & Kühn K. (1995). The processing of locally ambiguous relative clauses in German. *Journal of Memory and Language, 34*, 227-246.

Traxler, M.J., Morris, R.K., & Seely, R.E. (2002). Processing subject and object relative clauses: Evidence from eye-movements. *Journal of Memory and Language, 47*, 69-90.

Traxler, M.J., Williams, R.S., Blozis, S.A., & Morris, R.K. (2005). Working memory, animacy, and verb class in the process of relative clauses. *Journal of Memory and Language, 53*, 204 224.

Wanner, E.,& Maratsos, M. (1978). An ATN approach in comprehension. In: M. Halle, J. Bresnan, & G. Miller (Eds.), *Linguistic Theory and Psychological Reality*(pp. 119-161). Cambridge, MA: MIT Press.

Weckerly, J. & Kutas, M (1999). An electrophysiological analysis of animacy effects in the processing of object relative sentences. *Psychophysiology, 36*, 559-570.

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, dass ich die vorgelegte Dissertation selbstständig und ohne unzulässige fremde Hilfe angefertigt und verfasst habe, dass alle Hilfsmittel und sonstigen Hilfen angegeben und dass alle Stellen, die ich wörtlich oder dem Sinne nach aus anderen Veröffentlichungen entnommen habe, kenntlich gemacht worden sind. Ich versichere weiterhin, dass die Dissertation in der vorgelegten oder einer ähnlichen Fassung bei keiner anderen Fakultät eingereicht worden ist.

Berlin, 1. September 2015