

Acquisition of Adjectival Agreement in German: Sensitivity to Grammar is Reflected in 3-Year-Olds' Pupil Dilation

Assunta Süß, Petra Hendriks, Tom Fritzsche, and Barbara Höhle

To acquire full competence in their first language, children have to detect the relevant grammatical contrasts and relations and be able to use them in comprehension as well as production. One of these grammatical relations is agreement, which is based on syntactic features, such as number and gender. There is an agreement controller and an agreement target, and the target changes its form according to one or more syntactic features of the controller.

The age at which children acquire agreement seems to be modulated by the specific agreement relation. For example, in Italian, agreement between determiner and noun was reported to be acquired earlier than subject-verb agreement (Moscati & Rizzi, 2014). Moscati and Rizzi (2014) have proposed that this modulation is caused by the syntactic complexity of the structure expressing the agreement relations, with movement being the defining factor for complexity. The more move operations are involved, the more complex the structure is and therefore the harder for children to acquire the specific agreement relation. Put differently, the more local the agreement relation is, the earlier children will be able to acquire it. The present study focuses on German-speaking children's development of sensitivity to gender agreement in the determiner phrase (DP), specifically to gender agreement between the attribute adjective and the noun. The central question is from which age on German-speaking children are sensitive to adjectival agreement and how this relates to their sensitivity to subject-verb agreement.

1. Acquisition of gender and gender agreement

To acquire agreement in their language, children have to figure out which elements in the syntactic structure function as the controller and which as the target, as well as which syntactic features are relevant for establishing agreement. Furthermore,

*Assunta Süß, University of Potsdam, University of Groningen, International Doctorate for Experimental Approaches to Language And Brain; assuess@uni-potsdam.de; Petra Hendriks, University of Groningen; Tom Fritzsche, University of Potsdam; Barbara Höhle, University of Potsdam. We thank all children and adults who came to the BabyLAB in Potsdam. This work was supported by the Erasmus Mundus Joint Doctoral Programme (EMJD) of the European Union, 2014—0685/001-001-EMJD (Framework Partnership Agreement 2012-2025).

the formal means (e.g., inflectional endings) that mark these relations have to be detected by the children. The current study investigates German-speaking children's sensitivity to gender agreement within DPs, more specifically, to agreement marking in the attribute adjective (such as "*ein grosses Haus*", 'a_Ø big_{NEUT} house_Ø'). The inflection of the attributive adjective is rather complex in German, as several categories (case, gender, number and definiteness) affect the form of the adjective. Therefore, the question arises whether adjectival agreement would still be acquired earlier than subject-verb-agreement (as would follow from the proposal by Moscati & Rizzi, 2014) or whether this high degree of morpho-syntactic complexity would cancel out the benefit of the locality of this relation.

Previous studies have shown that children become aware of gender features and gender agreement quite early in their language development. German-speaking children use determiner forms that agree with their nouns in gender correctly as soon as determiners are regularly produced, around the age of two years (Clahsen, 1988; Szagun, Stumper, Sondag & Franik, 2007). However, there is no research so far that studied sensitivity to gender agreement in their speech input in German-speaking children of this age. Studies on children acquiring other languages provide evidence that the representation of gender features develops early in language acquisition. This has been reported in a preferential-looking experiment, where Spanish-speaking children of three years looked towards a target picture faster after it had been labeled by a determiner-noun phrase when the distractor showed the referent of a noun of different gender compared to a noun with same gender (Lew-Williams & Fernald, 2007). Similar results have been obtained for French-speaking 2-year-olds (van Heugten & Shi, 2009). However, knowledge on gender marking is not limited to words already present in a child's vocabulary. Dutch-speaking children as young as 2;0 years were shown to be sensitive to the association of a nominal diminutive marker (attached to a nonce stem) and the noun gender (van Heugten & Johnson, 2010). They listened longer to stimuli in which these novel nouns were paired with the determiner that corresponded in gender, compared to stimuli paired with a gender incongruent determiner. This finding indicates that children of this age already have an abstract representation of the gender feature, instead of an item-based representation of specific determiner-noun combinations. Melançon and Shi (2015) showed that French-speaking 2;6 year-olds can quickly assign new nouns to gender categories when trained with a gender marked indefinite article and use this knowledge immediately in a word recognition task when the new word was presented with a different, untrained gender-marked determiner. All of these findings emphasize that gender features are acquired within the third year of life and that children are sensitive to the impact that gender has on the form of the agreeing elements (determiners in all these studies).

However, not much is known about children's sensitivity to adjectival gender agreement. The only research we are aware of is a study with French-speaking 4- and 6-year-olds that tested the production and processing of gender agreement in determiners and adjectives (Roulet-Amiot & Jakubowicz, 2006). In an elicited produc-

tion task, children were asked to complete a sentence by describing a picture with a DP including an adjective. All stimuli drawings depicted an animal and an object, which served as the grammatical subject and object of the sentence, respectively. The experimenter started the sentence, such as 'Here the elephant is wearing ...', and the child was supposed to complete the sentence with e.g. '... a green shirt'. If s/he did not produce the expected adjective, the experimenter would ask directly, 'Which color is it?', and then repeat the elicitation task. Trials varied in gender of the noun (feminine vs. masculine) and position of the adjective (pre- vs. post-nominal). Both age groups produced more gender agreement errors with the adjectives than with the determiners. Considerable reduction in the number of errors across ages was only observed for the adjectives: while the error rate of the 4-year-olds was around 17.1%, it was only 4.6% for the 6-year-olds. In contrast, the error rate for the determiners was below 1% for both age groups. Productions where both the determiner and adjective carried an agreement violation were rare as well (around 1% for both age groups). Interestingly, an effect of the position of the adjective was found. In contrast to the authors' expectations, the children produced many more post-nominal adjectives (83%). However, the post-nominal adjectives showed an overall higher percentage of agreement errors than the pre-nominal adjectives. The authors consider this as an effect of the higher syntactic complexity in the post-nominal construction as the noun moves in front of the adjective, an explanation that is in line with Moscati and Rizzi's proposal (2014). In the comprehension part of the study, the same children were exposed to DPs including determiners, adjectives and nouns with congruent gender agreement or violations. These DPs were preceded by a sentential context. The participants had to judge whether the DP was a semantically appropriate continuation of this context or not via a press on a touch screen. Only the data from the 6-year-olds could be analyzed. Overall, they showed longer response latencies for non-agreeing compared to agreeing DPs. This effect was present independent of whether the determiner or the adjective caused the agreement violation. However, for the adjectives, this grammaticality effect was driven by the pre-nominal position. The latencies for the post-nominal adjectives were high in general, independent of agreement. With a 17.1% error rate in the production of adjectival agreement at the age of four years, this study shows that acquisition of this phenomenon is not complete at an early age, but that it still develops between the ages of four and six. Overall, these results provide evidence that a) determiner agreement is acquired earlier than adjectival agreement, as shown by the results from the production study, and b) that the syntactically more complex position, although preferred by the children for the production task, is associated with more agreement errors and a decreased sensitivity to agreement violations.

2. Complexity hierarchy of agreement relations

We suggest to integrate adjectival gender agreement into the agreement hierarchy put forth by Moscati and Rizzi (2014). According to this account, the lower

the number of movement operations that are involved in a given syntactic configuration, the earlier children should acquire the agreement relations within this configuration. More specifically, for Italian they proposed the following hierarchy: determiner-noun < subject-verb < clitic-past participle agreement. To test this hypothesis, sensitivity to these agreement configurations was investigated in Italian-speaking children. Three age groups (3;4, 4;6, and 5;4 years) were tested in a task in which the children were presented with pairs of sentences only differing in their agreement and they had to choose the correct one. All age groups achieved high levels of performance for determiner-noun agreement with no age effects. The youngest age group showed significantly lower performance for subject-verb agreement and even lower performance for clitic-past participle agreement. This pattern was also observable in the older children, even though not necessarily statistically significant between all conditions and with overall higher achievements. These results support the assumption that the complexity in which an agreement configuration occurs is in fact relevant for its acquisition.

Our study tests German-speaking children's sensitivity to gender agreement in adjectives, an agreement relation that is not present in the original agreement hierarchy proposed by Moscati and Rizzi (2014). We suggest that DP-internal adjective-noun agreement in German be ranked as less complex than subject-verb agreement. There are theories that state that adjectives are directly generated in the pre-nominal position of the noun in German (Giorgio & Longobardi, 1991) such that no movement operations would be necessary. In contrast, subject-verb agreement requires complex movement operations in German: The verb has to move from the verb phrase to the tense phrase and the subject then needs to move to an even higher position in the structure (e.g. Vikner, 1995). In the course of language acquisition, we thus expect adjective-noun agreement to be acquired earlier than subject-verb agreement.

Brandt-Kobe and Höhle (2014) tested German-speaking children's sensitivity to subject-verb agreement and its violation using the visual world paradigm. Three- and 5-year-old children were exposed to sentences and a visual display that showed the referent of the sentence subject, the referent of the sentence object (the target picture), and an unrelated distractor. Sentences were simple transitive sentences (e.g., *Der Hund bringt den Ball*, 'the dog brings the ball') with either correct subject-verb agreement or an agreement violation with respect to number. The authors measured the duration children looked at the target picture as well as the latency of when they first looked towards the target picture. The hypothesis was that a violation in agreement would disrupt syntactic processing and cause a delay in identifying the target picture, i.e. the referent of the sentence object.

Results revealed no significant effect of grammaticality for the 3-year-olds but the older children had longer looking times and shorter latencies when the agreement was congruent. This indicates their sensitivity to agreement as well as their detection of agreement violations.

Considering these results of sensitivity to subject-verb agreement and the underlying complexity, we hypothesize that sensitivity to adjectival agreement will be acquired earlier as this agreement configuration does not involve movement and is therefore less complex.

3. Pupillometry

For this study, we used a technique that is suitable to capture children's sensitivity to deviant speech stimuli and at the same time should not affect the results by requiring the children to master a challenging task. A method fulfilling these needs is pupillometry, the measurement of the pupil dilation while listening to speech and looking at a visual display. Changes in the diameter of the pupils are not only a reflection of changes in illumination but are also associated with cognitive processes like working memory load (Kahneman & Beatty, 1966) or violations of expectation (Karatekin, 2007). With respect to sentence processing, several studies have shown higher pupil dilation when participants were exposed to linguistically complex stimuli (Schluroff, 1982; Just & Carpenter, 1993). Thus this method is assumed to indicate the cognitive effort that is needed to process syntactic structures.

Most importantly for this study, pupillometry has proven useful for investigating sensitivity to deviances from the correct linguistic form in children on a different aspect, namely mispronunciation detection. In these studies participants were simultaneously presented pictures of objects with their spoken labels, either pronounced correctly or slightly mispronounced. A larger pupil dilation was found for mispronounced compared to correctly pronounced words in adults and 2;6-year-old monolingual children (Fritzsche & Höhle, 2015). Pupil dilation even indicated sensitivity to featural distance between the correct and the mispronounced object names as the increase of the pupil was correlated to the degree of their phonetic distance (Tamási, McKean, Gafos, Fritzsche & Höhle, 2017). All of these findings indicate that pupillometry should be capable of measuring sensitivity to the well-formedness of linguistic stimuli. In this study, we extended the application of pupillometry to the investigation of syntactic agreement. We expected that differences in pupil size reflect sensitivity to agreement violations such that violations would result in larger pupil dilations than correct gender agreement between adjectives and nouns. In particular, we hypothesized that, if children are sensitive to adjectival agreement, pupil dilation would be larger in sentences with ungrammatical adjectival agreement than in sentences with grammatical adjectival agreement.

4. Method

4.1. Participants

Fifty-four children (29 girls) from two age groups took part in this experiment: 2;6-year-olds ($n=30$, ranging from 2;5.13 to 2;6.9) and 3;0-year-olds ($n=24$, 3;0.10 to 3;1.1). The data from another 13 children had to be excluded, one

for technical reasons and 12 for not completing the test session. All children were raised in a monolingual German environment in the region of Berlin and Potsdam. None were born prematurely or had any reported developmental language difficulties. Parents of the children filled out the FRAKIS questionnaire, the German version of the MacArthur-Bates CDI (Szagun, Stumper & Schramm, 2009); all of the children obtained FRAKIS scores that were in the typical range for their age. Parents were reimbursed for their travel costs and children received a small gift.

As controls, 41 German-speaking adults (28 women, mean age 25;2 years) were tested, all students of the University of Potsdam. Adults followed the exact same procedure and were rewarded either via course credits or financial reimbursement for their time.

4.2. Stimuli

4.2.1. Linguistic materials

In German, each noun is assigned to one of three gender categories (feminine, masculine and neuter). However, there are no formal gender makers on most nouns. Agreement markers for gender are found on the determiner and on attributive adjectives. As German morphology is predominantly fusional, there are no separate markers for gender, but gender marking is often collated into one form along with other grammatical categories like case and number. In the following, we will provide more detail on gender agreement on determiners and adjectives in German.

Leaving aside case and number (as we only consider nominative singular phrases in this study), the form of the attributive adjective depends on the interaction of gender and definiteness (Tab. 1). After a definite determiner, the adjective is not uniquely marked for gender but the same inflection is used across genders (*der_{MASC}/die_{FEM}/das_{NEUT} rote ...*, ‘the red ...’). After an indefinite determiner, the adjective is unambiguously marked for gender (*ein_{MASC/NEUT} roter_{MASC} Bus_(MASC)*, ‘a red bus’, *eine_{FEM} rote_{FEM} Tür_(FEM)*, ‘a red door’, *ein_{MASC/NEUT} rotes_{NEUT} Haus_(NEUT)*, ‘a red house’). The current study used this so called ‘strong inflection’ where the gender information is expressed exclusively on the adjective. We chose to include only masculine and neuter gender, as this allowed us to avoid ambiguity in gender marking.

Table 1: Gender agreement in German indefinite and definite DPs (nominative, singular). Used forms are marked in gray.

	DETERMINER			ADJECTIVE		
	<i>fem</i>	<i>masc</i>	<i>neuter</i>	<i>fem</i>	<i>masc</i>	<i>neuter</i>
<i>indefinite</i>	eine	ein		rot-e	rot-er	rot-es
<i>definite</i>	die	der	das	rot-e		

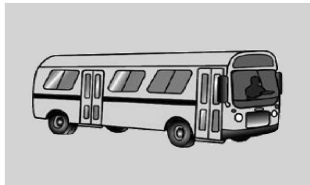
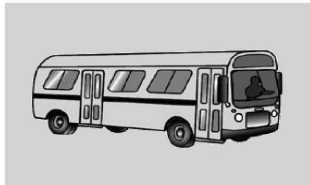


Twenty neuter and 20 masculine nouns were selected as stimuli. All were monosyllabic and contained no obvious semantic, morphological or phonological cues to gender. Five adjectives that frequently occur in children’s input at this age were chosen (*rot, blau, gelb, klein, groß*, ‘red, blue, yellow, small, big’).

The sentence started with ‘There is ...’ (*Da ist ...*) before the DP. The DPs consisted of the indefinite article *ein* ‘a’, which is the masculine and neuter form of this determiner in German, the attributive adjective, and the noun. Every test sentence was preceded by the carrier phrase ‘Look!’ (*‘Schau!’*) in order to draw the child’s attention to the picture.

The sentences were recorded in a soundproof booth (IAC) by a female German native speaker, using an audio-technica AT4022a microphone and Audacity (version 2.1) for digitizing the signal with the sampling rate of 44100 Hz.

The volume of all sentences was normalized to 70dB after the recording. Grammatical as well as ungrammatical sentence versions were cross-spliced from grammatical recordings using Praat (Boersma & Weenink, 2015). Mean sentence length was 3893 ms.

Table 2: Examples of sentence-picture combinations used in the experiment, top: ‘bus’ (*‘Bus’*, masculine), bottom: ‘house’ (*‘Haus’*, neuter). The original pictures were colored.

congruent	<i>Schau, da ist ein gelber Bus!</i>	
	Look _{IMP} , there is a yellow _{MASC} bus _{MASC} 'Look, there is a yellow bus!'	
incongruent	<i>*Schau, da ist ein gelbes Bus!</i>	
	Look _{IMP} , there is a yellow _{NEUT} bus _{MASC} 'Look, there is a yellow bus!'	
congruent	<i>Schau, da ist ein blaues Haus!</i>	
	Look _{IMP} , there is a blue _{NEUT} house _{NEUT} 'Look, there is a blue house!'	
incongruent	<i>*Schau, da ist ein blauer Haus!</i>	
	Look _{IMP} , there is a blue _{MASC} house _{NEUT} 'Look, there is a blue house!'	

4.3. Study design

The study contained one manipulation, namely the gender agreement of the adjective that could be congruent or incongruent. Every noun was used only once for each participant and was combined with a gender congruent or gender incongruent form of the adjective. Each participant was assigned to one of eight pseudo-randomized versions of trial order, balanced for gender and agreement congruency. In total, each participant was presented with 40 items, 20 of them neuter and 20 masculine, half of them gender-congruent and half of them violating the agreement relation.

4.4. Procedure

The study was administered as a single-picture looking task. The picture was presented in the center of a 1280x1024 TFT monitor. Each picture was presented for exactly 8000ms; 1000ms after its appearance, the corresponding sentence started to play. After the end of the sentence, the picture remained on the screen for another 4000ms (slightly varying, depending on the length of the sentence).

The study was implemented on a Tobii 1750 eye-tracking system in a dual computer set-up (Tobii, Sweden). Pictures were presented on a 17-inch monitor, and sentences were played through speakers hidden behind the screen. ClearView software was used for stimulus presentation and data collection.

During the experiment, children sat on their parent's lap and the eye tracker recorded their gaze position and pupil data. Parents were listening to masking music over headphones. Children were instructed by the experimenter to look at the pictures in front of them ('Hey, do you want to look at some pictures?'). Before each item, a picture with a small colored sun in the middle was displayed, accompanied by a short ringing sound, to draw the child's attention to the monitor. Each testing session lasted approximately seven minutes. During the session, the experimenter observed the testing in an adjacent room via a video camera that was placed next to the eye tracker.

4.5. Results

Only trials with more than 25% valid data points in the time during the presentation of the sentence (ms 1000-5000) were analyzed. Data loss was due to blinking or looking away. Missing data was discarded and we did not apply any interpolation procedure. Pupil sizes from the left and the right eyes were averaged to obtain a single value.

In order to time-lock the pupillary response to the critical information and to account for spontaneous variation in pupil diameter, the pupil size measure was baseline-corrected for each individual trial using a 200ms period prior to the adjective offset. The mean pupil size from this baseline period was subtracted from all data points of a given trial. The time window of analysis was defined *a priori* as starting at the offset of the adjective, where the gender is marked, and lasted for 3000ms.

We opted for a Growth Curve Analysis of the data to account for the dynamics of pupil size changes in this 3000ms window. By including higher order polynomial terms for time, nonlinear changes can be modeled (Mirman, Dixon & Magnuson, 2008). The analysis was implemented in R (version 3.4.2; R Core Team, 2017) using the lme4 package (version 1.1-14, Bates, Maechler, Bolker & Walker, 2015).

We specified the following fixed effects: agreement congruency and time as within-participant predictors, and in addition for the children's data age as a between-participant predictor (2;6-year-olds and 3;0-year-olds). The contrasts of the factors were coded by a +.5/- .5 dummy coding. The continuous predictor of time was included with three potencies to allow for a linear, a quadratic, and a cubic trend. The random component structure included intercepts for participants and items as well as adjustments of the agreement congruency effect for individual participants (i.e. random slope)¹.

Agreement congruency is the main factor of interest as this would reveal differences in the processing of gender agreement. However, interactions with time are expected here due to the dynamics of the pupillary response.

First, the adult data will be discussed (Fig. 1). The simple effect of agreement congruency was not significant (estimate: .015, $p=.311$); there is no overall difference in pupil size between congruent and incongruent sentences. Pupil size over time changed in a linear fashion (estimate: 9.68, $p<.001$), as well as in a quadratic and cubic order of time (estimate: -11.05, $p<.001$, estimate: -4.11, $p<.001$). The quadratic term reflects the symmetric shape of the curve with a maximum dilation around 2000 ms. The cubic term captures the asymmetry of the curve in the sense that the latency of the peak is not in the middle of the analysis window and the pupil size change does not go back to zero at the end of the window. While these effects of time describe the general pupil dynamics, the relevant question is whether these changes over time differ across the congruency conditions. Pupil dilation change over time is larger for ungrammatical sentences in the linear (estimate: 7.59, $p<.001$), the quadratic (estimate: 3.34, $p<.001$), as well as in the cubic term (estimate: -1.52, $p=.004$). These effects show that ungrammatical compared to grammatical sentences lead to an increasingly larger pupil (linear), a higher peak (quadratic), and a stronger

1 The full model specification was: $\text{pupil_change} \sim \text{agreement} * \text{age} * \text{poly}(\text{time}, \text{degree} = 3) + (1 \mid \text{item}) + (1 + \text{agreement} \mid \text{participant})$ for the children and without the term “age” for the adults.

asymmetry in the curve (cubic), which is reflected by a later peak in the ungrammatical condition and also a more sustained dilation at the end of the analysis window. From these results, we conclude that this method is suitable to measure sensitivity to agreement violations.

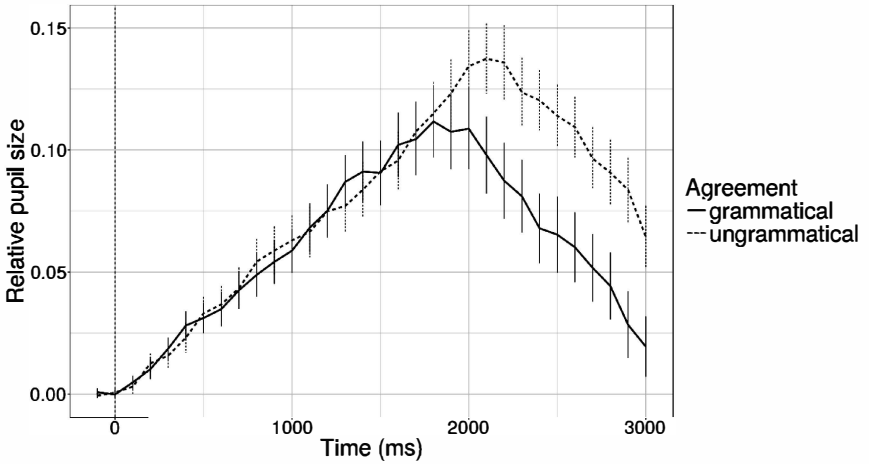


Figure 1: Mean pupil size of adult participants, baselined per trial. Time point zero marks the offset of the agreement marking adjective. Solid line indicates sentences with congruent agreement, dashed line indicates sentences with incongruent agreement markers. Error bars indicate one standard error of the mean.

In the children’s data, the effect of age was not significant nor did it interact with agreement (estimate: $-.013$, $p=.601$, estimate: $.005$, $p=.803$). This is why the data for both age groups are pooled in Fig. 3. Pupil size for ungrammatical sentences is numerically larger than for grammatical sentences (as in adults), however, as a simple effect this is not significant (estimate: $.009$, $p=.550$).

The influence of time on the pupil size dynamics, independent of sentence condition, is comparable to the adults (linear estimate: -54.37 , $p<.001$; quadratic estimate: -4.45 , $p<.001$; cubic estimate: 34.57 , $p<.001$).

The interaction of agreement with time was only significant in the quadratic term (estimate: -1.96 , $p=.007$) while the linear and cubic interaction terms did not reach significance (linear estimate: $-.20$, $p=.781$, cubic estimate: $.33$, $p=.650$). This means that the peak dilation was larger for ungrammatical than for grammatical sentences and the shape of the curves is very similar.

To summarize, pupil size changes in response to linguistic stimuli in children and adults are quite comparable, with a peak at around 2000 ms after the critical agreement information. In adults, both the amplitude and the latency

of this pupillary response are affected by the congruency condition, with larger and later peaks for ungrammatical compared to grammatical sentences. Moreover, the amplitude difference is long lasting: until the end of the analysis window. In contrast, the effect in children is very subtle and no latency difference or longer-lasting effect is found. There is only a small difference in the peak of the pupillary response curve (expressed by the quadratic term of time interacting with agreement), which goes in the same direction as in adults: larger dilations for ungrammatical compared to grammatical sentences. This can be taken as an indicator that children in both age groups are at least beginning to be sensitive to gender agreement violations. The age difference of six months between both groups of children did not have an effect.

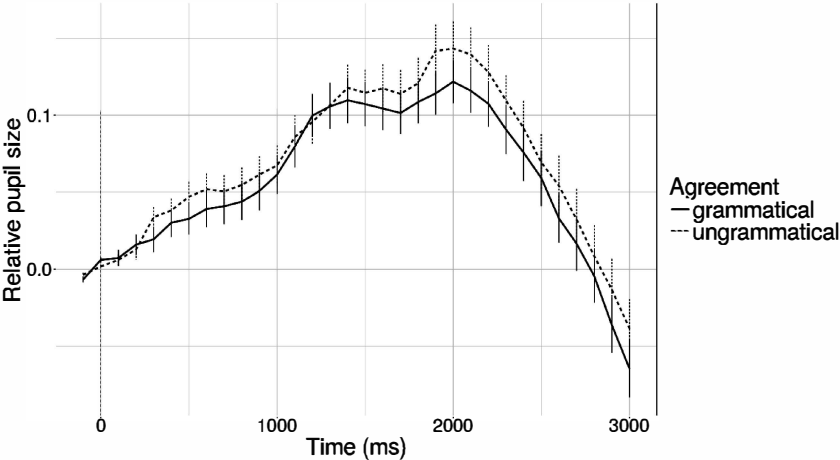


Figure 2: Mean pupil size of both groups of children combined, baselined per trial. Time point zero marks the offset of the agreement marking adjective. Solid line indicates sentences with congruent agreement, dashed line indicates sentences with incongruent agreement markers. Error bars indicate one standard error of the mean.

5. Discussion

The main aim of this study was to investigate if German-speaking children at the age of 2;6 and 3 years are sensitive to gender agreement marked on attributive adjectives. Multiple studies in different languages have shown that children are sensitive to gender features of nouns and establish the agreement relation between noun and determiner quite early, but there are only very few studies on the acquisition of attributive gender agreement. According to the complexity hierarchy by Moscati and Rizzi (2014), complexity of the syntactic

configuration that bears the agreement predicts the order in which children acquire different agreement relations. Within this framework, agreement of an attributive adjective is assumed to be less complex in German than subject-verb agreement, and thus German children should be sensitive to adjective agreement earlier than to subject-verb agreement.

This study used the method of pupillometry to test children's sensitivity to gender marking and collected data from German-speaking adults and children. Previous studies on several linguistic phenomena have shown that larger pupil dilations are associated with greater processing costs. Based on this, we predicted that sentences with gender-incongruent agreement markers would result in larger pupil dilations upon detecting the mismatch. Verifying the method first with adults, the results showed a significant difference in pupil dilation for sentences including a DP with congruent versus incongruent gender marking on the adjective. This shows that the method of pupillometry is suited to measure sensitivity to gender agreement violations.

The next step was to determine if children at the age of 2;6 and 3 years already show signs of sensitivity to gender agreement. If they do, their pupillary responses should show a pattern similar to that of the adult control group. While the results show an overall similarity to the results of the adults, the difference between the congruent and incongruent sentences was smaller and more focal in children than in adults. Nevertheless, an effect of gender agreement violations on pupil size was obtained, showing that children at this age are sensitive to this agreement relation. There was no difference in performance between the two age groups, which could be taken as indication that no crucial developmental change in the ability to process gender agreement occurs between 2;6 and 3 years.

Considering the complexity hierarchy of agreement and its implication for agreement acquisition, our prediction was that adjectival agreement would be acquired before subject-verb agreement. Comparing our data to the study by Brandt-Kobe and Höhle (2014) that did not find evidence that German 3-year-olds are sensitive to violations of subject-verb agreement supports this prediction. In this regard, our data fit the complexity hierarchy as proposed by Moscati and Rizzi (2014). However, there is still a possibility that sensitivity to subject-verb agreement in German-speaking children might be detected at an earlier age with a different experimental method. In the study on German subject-verb agreement, the dependent variables were latency of target fixation and the proportion of looks, based on the assumption that the agreement violation would disrupt the processing of the sentence. We cannot exclude that pupil dilation reflects sensitivity to grammatical violations more accurately than gaze pattern. Therefore, further research is needed that compares children's sensitivity to these two agreement relations using the same dependent measure.

One other crucial aspect that may contribute to differences in the developmental trajectory between subject-verb agreement and adjectival gender agreement is the difference between the grammatical features of number and gender.

While number is typically a semantic property of the noun referent, this is true only for a very restricted subset of cases for gender. For the majority of nouns, grammatical gender does not reflect a semantic property but rather an abstract lexical-syntactic property. Because of this, the complexity hierarchy based on locality and movement alone might not suffice to account for different acquisition trajectories. For a direct comparison of the featural differences between number and gender, minimizing locality/movement differences, subject-verb agreement could be tested against predicative adjectives. However, German does not mark gender agreement on predicative adjectives, thus this would have to be studied in a language with predicative adjectival agreement, like French. In such a plan, a comparison of results with predicative and attributive adjectives would also be desirable. Another source of difficulty for the acquisition of gender could lie in the number of categories. While there are two number categories (singular and plural) there are three for gender in German (feminine, masculine, and neuter) that might make the feature gender more difficult to acquire. Studies investigating the sensitivity to gender agreement have so far been only conducted in languages with only two gender categories (Dutch, French, and Spanish).

To conclude, our findings present the first evidence that children acquiring German²;6 years and older show beginning signs of sensitivity to gender agreement marking on attributive adjectives. This is concluded based on the different temporal dynamics of pupil dilations when children were exposed to sentences with correct agreement versus sentences with agreement violations. This early sensitivity of young children is comparable to adults' sensitivity regarding the direction of the effect, but is not yet adult-like in size and extent.

References

- Bates, Douglas, Maechler, Martin, Bolker, Ben, Walker, Steve (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- Boersma, Paul & Weenink, David (2015). Praat: doing phonetics by computer [Computer program]. Version 6.0.30, retrieved 22 July 2017 from <http://www.praat.org/>.
- Brandt-Kobe, Oda C., & Hoehle, Barbara. (2014). The detection of subject-verb agreement violations by German-speaking children: An eye-tracking study. *Lingua*, 144, 7–20.
- Clahsen, Harald (1988). *Normale und gestörte Kindersprache*. Amsterdam: John Benjamins.
- Giorgio, Alessandra & Longobardi, Giuseppe (1991). *The syntax of noun phrases: configuration, parameters and empty categories*. Cambridge: Cambridge University Press.
- van Heugten, Marieke & Johnson, Elisabeth K. (2010). Linking infants' distributional learning abilities to natural language acquisition. *Journal of Memory and Language*, 63(2), 197–209.
- Van Heugten, Marieke & Shi, Rushen. (2009). French-learning toddlers use gender information on determiners during word recognition. *Developmental Science*, 12(3), 419–425.

- Just, Marcel A. & Carpenter, Patricia A. (1993). The intensity dimension of thought: pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, 47(2), 310–339.
- Kahneman, Daniel & Beatty, Jackson (1966). Pupil diameter and load on memory. *Science*, 154(3756), 1583–1585.
- Karatekin, Canan (2007). Eye tracking studies of normative and atypical development. *Developmental Review*, 27(3), 283–348.
- Lew-Williams, Casey & Fernald, Anne (2007). Young Children Learning Spanish Make Rapid Use of Grammatical Gender in Spoke Word Recognition. *Psychological Science*, 18(3), 193–198.
- Melançon, Andréane, & Shi, Rushen (2013). Grammatical Knowledge and its Immediate Processing during Online Comprehension by Infants. In Sarah Baiz, Nora Goldman & Rachel Hawkes (Eds.), *Proceedings of the 37th annual Boston University Conference on Language Development* (2). Somerville, MA: Cascadilla Press. 251–264.
- Melançon, Andréane & Shi, Rushen (2015). Representation of abstract grammatical feature agreement in young children. *J. Child Lang.*, 42, 1379–1393.
- Mirman, Daniel, Dixon, James A. & Magnuson, James S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, 59(4), 475–494.
- Moscato, Vincenzo & Rizzi, Luigi. (2014). Agreement configurations in language development: A movement-based complexity metric. *Lingua*, 140, 67–82.
- R Core Team. (2017). R: A language and environment for statistical computing.
- Roulet-Amiot, Leslie, & Jakubowicz, Célia (2006). Production and perception of gender agreement in French SLI. *International Journal of Speech-Language Pathology* 8(4), 335–346.
- Schluroff, Michael (1982). Pupil responses to grammatical complexity of sentences. *Brain and Language*, 17(1), 133–145.
- Szagun, Gisela, Stumper, Barbara, Sondag, Nina & Franik, Melanie. (2007). The acquisition of gender marking by young German-speaking children: Evidence for learning guided by phonological regularities. *Journal of Child Language*, 34(3), 445–471.
- Szagun, Gisela, Stumper, Barbara & Schramm, Antonio (2009). *Fragebogen zur frühkindlichen Sprachentwicklung (FRAKIS) und FRAKIS-K (Kurzform)*. Frankfurt/M: Pearson.
- Tamási, Katalin, McKean, Cristina, Gafos, Adamantios, Fritzsche, Tom & Höhle, Barbara (2017). Pupillometry registers toddlers' sensitivity to degrees of mispronunciation. *Journal of experimental child psychology*, 153, 140–148.
- Vikner, Sten (1995). *Verb movement and expletive subjects in the Germanic languages*. New York: Oxford University Press.