Gradual development of non-adjacent dependency learning during early childhood

Mariella Paul[1,2], Claudia Männel[1,3], Anne van der Kant[4], Jutta L. Mueller[5,6], Barbara Höhle[4], Isabell Wartenburger[2,4], and Angela D. Friederici[1,2]

[1] Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences

[2] Berlin School of Mind and Brain, Humboldt Universität zu Berlin

[3] Department of Audiology and Phoniatrics, Charité – Universitätsmedizin Berlin, Germany

[4] Cognitive Sciences, Dept. Linguistics, University of Potsdam

[5] Institute of Cognitive Science, Osnabrück University

[6] Department of Linguistics, University of Vienna

Correspondence concerning this article should be addressed to Mariella Paul, Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstr. 1a, 04103 Leipzig, Germany, Tel: +49 341 9940-123, Email: paulm@cbs.mpg.de

The authors declare no conflict of interest. Research data are not shared due to privacy or ethical restrictions.

Acknowledgements: The Max Planck Society (MP, CM, ADF), the German Research Foundation (DFG) in the Research Unit FOR 2253 (Project number 258522519), project FR 519/20-1 (MP, CM, AK, BH, IW, ADF), and the Berlin School of Mind and Brain (MP) funded this project. Further, we would like to thank all participating families for their commitment, as well as Anika Schütze, Bettina Nowotny, Bianca Dietrich, Christina Rügen, Claudia Geißler, Eileen Stelter, Kristiane Klein, Leonie Faulhaber, and Sophia Richter for their help with data acquisition and their dedicated work with the participating children and families.

Research Highlights

- Transition between different developmental stages of non-adjacent dependency learning during early childhood evidenced by event-related brain potentials
- Children between 1 and 3 years of age showed learning of non-adjacent dependencies in a foreign language during passive listening
- Brain responses revealed associative non-adjacent dependency learning across the tested age range, triggered by passive listening
- Gradual decrease of the strength of associative non-adjacent dependency learning, during early childhood

Abstract

In order to become proficient native speakers, children have to learn the grammatical rules of their language. These grammatical rules can define morpho-syntactic relations between neighboring as well as distant elements of a sentence, so-called non-adjacent dependencies (NADs). Previous neurophysiological research suggests that NAD learning comprises different developmental stages during early childhood. Children up to 2 years of age show evidence of associative NAD learning under passive listening conditions, while children starting around the age of 3 to 4 years fail to show learning under passive listening, similarly to the pattern observed in adults. To test whether the transition between these developmental stages occurs in a gradual manner, we tested young children's NAD learning in a foreign language using event-related potentials (ERPs). We found ERP evidence of NAD learning across the age of 1 to 3 years. However, the amplitude of the ERP effect indexing NAD learning decreased linearly with increasing age. These findings indicate a gradual transition in children's ability to learn NADs associatively under passive listening during early childhood. Cognitively, this transition might be driven by children's increasing morpho-syntactic knowledge in their native language, hindering NAD learning in novel linguistic contexts during passive listening. Neuroanatomically, changes in brain structure might play a crucial role, especially the maturation of the prefrontal cortex, which promotes top-down learning, as opposed to bottom-up, associative learning. In sum, our study provides evidence that NAD learning under passive listening conditions undergoes a gradual transition between different developmental stages during early childhood.

Keywords Development, Language acquisition, Statistical learning, Artificial language learning, Non-adjacent dependencies, Event-related potentials

1 Introduction

In order to successfully communicate with their environment, infants must not only learn the words of their native language(s) but also their relations that define how these words combine into phrases and sentences. These relations can hold for neighboring elements in a sentence (e.g., *Mary is happy*) or non-neighboring elements (e.g., *The sister is singing*). These latter relations, so-called non-adjacent dependencies (NADs), require the learner to track dependent elements (*is* and *–ing*) across one or more intervening elements (*sing*-). NADs are a crucial aspect of natural languages and are present, for example, in subject-verb agreement and English tense marking. Nevertheless, NADs are relatively difficult to learn and behavioral evidence indicates that during both, natural language acquisition and artificial language learning, children learn NADs only in their second year of life, around 14-15 months of age, depending on the acquired language (Culbertson et al., 2016; Gómez & Maye, 2005; Höhle et al., 2006; Santelmann & Jusczyk, 1998).

The learning of NADs during early childhood has been shown to undergo different developmental stages, both, in terms of behavioral and neurophysiological learning measures (Culbertson et al., 2016; Mueller et al., 2019; van der Kant et al., 2020). For example, Culbertson and colleagues (2016) investigated French infants' learning of NADs in subject-verb agreement and observed two stages of different behavioral responses in a head-turn preference procedure. Across age, children displayed two cycles of a shift from a familiarity preference for encountered NADs to a novelty preference. The authors proposed that the processes related to the first stage, from 14 to 18 months of age, reflect initial surface-level representations based on *phonological* features of the NADs. In contrast, the processes related to the second stage, from 21 to 24 months of age, were interpreted as higher-level representations of the *morphological* features (Culbertson et al., 2016).

Further evidence of different developmental stages of NAD learning comes from studies using event-related potentials (ERPs). Friederici and colleagues (2011) tested German 4-month-old infants' NAD learning in a familiarization-test paradigm using a non-native language (Italian) containing NADs (e.g. *La sorella sta cantando*; *The sister is singing*). By means of ERPs, the authors showed that even 4-month-old infants succeeded at NAD learning. Learning was indexed by a late positive ERP effect, which was interpreted as associative learning of NADs based on phonological cues (Friederici et al., 2011). NAD learning in this miniature version of Italian was shown to display different developmental stages during early childhood, with an initial stage up to the age of 2 to 3 years, during which young children can learn NADs through passive listening, which likely triggers associative learning (Friederici et al., 2011; Mueller et al., 2019; van der Kant et al., 2020). This is followed by a stage starting at around 3 to 4 years, during which children and adults struggle to learn NADs under passive listening (without additional cues marking the NADs), but succeed in learning under active listening conditions, that is, when a task, administered during the whole experiment or during testing only, guides their attention towards the NADs (Friederici et al., 2013; Mueller et al., 2012; van der Kant et al., 2020; see also Pacton et al., 2015; Pacton & Perruchet, 2008). Notably, adults' neural signature of NAD learning under active listening conditions differed from infants' neural signature under passive listening: Adults showed an N400-like, negative ERP component and a P3, which were interpreted as indicating attention-driven lexicalization of the NAD processing (Mueller et al., 2009), in contrast to infants' associative learning based on phonological information processing, indexed by a late positive ERP component (Friederici et al., 2011, 2013). This difference was proposed to partly be driven by the maturation of the prefrontal cortex which supports top-down processes in contrast to the temporal cortex subserving bottom-up associative processes (Skeide & Friederici, 2016). Thus, ERP research on NAD learning further corroborates the observation of different developmental stages, as the neural signature of NAD learning seems to change with age, and is dependent on the presence or absence of a task.

As reviewed above, in the ERP literature on NAD learning, the detection of NAD violations has been found to be indexed by different ERP responses, depending on the specifics of the experimental design and stimuli as well as the studied participant sample. In particular, in the studies using the Italian sentence materials reviewed above, the ERP polarity differed depending on participants' age and whether participants were tested under passive listening conditions or performed a task (Friederici et al., 2011, 2013; Mueller et al., 2009). Moreover, the polarity was shown to change with the discrimination difficulty of the chosen stimuli (Schaadt & Männel, 2019), infants' sex (Mueller et al., 2012), and whether school-age children were tested immediately after NAD learning or following a retention period involving sleep (Schaadt, Paul et al., 2020). The polarity of infants' ERP effects was also found to be associated with later language outcomes in studies testing lexical segmentation and phoneme discrimination (Kooijman et al., 2013; Schaadt et al., 2015). Therefore, we do not make specific predictions about the polarity of the measured ERP effects, but instead investigate whether it plays a role in the developmental stages of NAD learning and the transition between them.

Investigating the nature of this developmental transition can improve our understanding of children's acquisition of grammatical rules and language learning in general. A rich body of literature has demonstrated that language learning, and in particular learning grammatical rules, has its peak in infancy and degrades over development (e.g., Hartshorne et al., 2018; Johnson & Newport, 1989; Senghas et al., 2004; Singleton, 2005). It has been suggested that this time course is driven by a sensitive period of language learning influenced by brain plasticity (e.g. Knudsen, 2004; Kuhl, 2004, 2010; see also Skeide & Friederici, 2016). If NAD learning also undergoes a sensitive period, we would expect to see a gradual transition between the developmental stages of NAD learning described above.

In the present study, we set out to systematically investigate the transition between the different developmental stages of NAD learning under passive listening conditions using the same paradigm with Italian sentences as previous studies (Friederici et al., 2011, 2013; Mueller et al., 2009; van der Kant et al., 2020). Using a set of non-native sentences allows for combining the advantages of artificial grammars, by being well-controlled, and natural language, by being more naturalistic than artificial grammars. Based on previous findings with the same paradigm and artificial grammar paradigms, we propose that a transition between different developmental stages of NAD learning takes place between 2 and 4 years of age (Mueller et al., 2019; van der Kant et al., 2020). To investigate whether this transition occurs in a gradual manner, we exposed 1- to 3-year-old children to Italian sentences under passive listening conditions in a familiarization-test paradigm while recording ERPs. Based on a previous infant ERP study using the same paradigm (Friederici et al., 2011), we expected to see a late positive ERP effect in response to incorrect sentences (containing NAD violations) compared to correct sentences (containing familiarized NADs) during test phases. If the transition between the different stages of NAD learning indeed occurs in a gradual manner, we would expect to see a gradual, that is, a linear decrease in the amplitude of this ERP effect of NAD learning between the age of 1 and 3 years. As an exploratory analysis, we also included ERP polarity and investigated whether the ERP effect polarity was associated with children’s age, sex and behaviorally tested language development (comprehension and production), as it was in previous ERP studies (Friederici et al., 2011, 2013; Kooijman et al., 2013; Mueller et al., 2012; Schaadt et al., 2015).

2 Material and Methods

2.1 Participants

115 healthy children growing up in monolingual German families participated in this study. Of these, 40 were 1 year old (mean age: 12.80 months, SD: 0.54; 20 girls), 40 were 2 years old (mean age: 25.08 months, SD: 0.88; 16 girls) and 35 were 3 years old (mean age: 37.10, SD: 0.60; 18 girls). Children were orally informed about the experimental procedure and caregivers were informed both in written and oral form. Caregivers gave written informed consent for their children's participation in the study. Ethical approval was obtained from the Medical Faculty of the University of Leipzig. Forty-nine additional children had to be excluded from data analysis, either due to non-compliance during the experimental procedure (22 children) or because they contributed less than 10 artefact-free EEG trials per condition across all test phases and/or no trials in at least one test phase (27 children). For additional information on EEG trial numbers see Table 1.

2.2 Stimuli

2.2.1 EEG experiment

The stimuli for the EEG experiment were adapted from Mueller and colleagues (2009). Participants listened to Italian sentences (see Figure 1) spoken by a female native speaker of Italian. The sentences consisted of a determiner phrase (*La sorella* (*The sister*) or *Il fratello* (*The brother*)), followed by a verb phrase consisting of an auxiliary (*sta* (*is*) or a modal verb *puo* (*can*)), a verb stem (32 different verbs, e.g., *cant-* (*sing-*)), and a suffix (-*ando* (-*ing*) or -*are* (-∅)). In grammatical sentences, *sta* (*is*) is followed by a verb stem and -*ando* (-*ing*), while *puo* (*can*) is followed by a verb stem and -*are* (-∅). Ungrammatical sentences were created from grammatical sentences (Adobe Audition), such that the verb suffixes were exchanged, namely *sta* (*is*) was followed by a verb stem and -*are* (-∅), while *puo* (*can*) was followed by a verb stem and -*ando* (-*ing*). Sentences had a mean length of 2.43 s (SD: 0.12). Because we counterbalanced whether participants were familiarized with grammatical or ungrammatical Italian sentences, we will refer to the NADs presented in the familiarization phases as correct (i.e., regardless of grammaticality in Italian).

2.2.2 Tests of language development

For all three age groups, we behaviorally assessed language abilities via standardized tests. For 1-year-old children, we used the German version of the *Bayley Scales of Infant and Toddler Development Screening Tes*t (Bayley, 2015). For the language comprehension measure, we used the subscale receptive language of the Screening Test and for the language production measure, we used the subscale productive language. Both of these subscales assess children’s vocabulary in a playful interactive manner. For 2- and 3-year-old children, we used the *Sprachentwicklungstest für zweijährige Kinder* (SETK-2; Grimm et al., 2016) and *Sprachentwicklungstest für drei- bis fünfjährige Kinder* (SETK 3-5; Grimm, 2015), respectively. For 2-year-old children’s language scores, we used the averaged word and sentence comprehension and production subscales. For 3-year-old children’s language scores, we used the sentence comprehension and production measures. Because the Bayley Screening Test only offers raw scores, we used z-transformed raw scores of all tests to allow for comparisons across age groups.

2.3 Procedure

During the EEG experiment, participants were seated on their caregiver’s lap in a soundproof booth. The experiment consisted of four familiarization phases alternating with four test phases. Familiarization phases (3.3 minutes each) consisted of 64 correct sentences each (overall 256 sentences). Each familiarization phase was followed by a test phase (1.3 minutes each), consisting of 8 correct and 8 incorrect sentences each (overall 32 sentences per condition; see Figure 2). Note that the sentences that participants heard in the test phases were not repeated in any of the familiarization phases. Further, verb stems were divided into two sets, such that they were not presented in the familiarization phase immediately preceding a given test phase, but only occured in earlier or later familiarization phases. The inter-stimulus-intervals (ISI) were 580 ms in the familiarization phases and 1380 ms in the test phases. Between familiarization and test phases, there was a pause of 2780 ms. Overall, the EEG experiment took approximately 20 min, during which participants watched a silent children's movie (*Peppa Pig*, *Bummi*, or *Alles Trick 9*) in order to increase compliance. The EEG experiment and the standardized language tests were either administered during the same session or in two separate sessions (mean time between sessions: 13.35 days, SD: 14.78).

2.4 Data recording

EEG data were recorded from 24 Ag/AgCl electrodes (Fp1, F7, F3, Fz, F4, F8, FC3, FC4, T7, C3, Cz, C4, T8, CP5, CP6, P7, P3, Pz, P4, P8, O1, O2, M1, and M2) placed according to the International 10-20 System of Electrode Placement and secured in an elastic electrode cap (Easycap GmbH, Herrsching, Germany). Cz served as an online reference during recording. Electrooculograms (EOG) were recorded from 4 additional electrodes, placed supraorbitally (Fp2) and infraorbitally (V-) to the right eye to capture vertical eye movements, as well as laterally to the left (F9) and right eye (F10), for horizontal eye-movements. The signal was digitized with a sampling rate of 500 Hz.

The EEG data were analyzed using the Fieldtrip toolbox (Oostenveld et al., 2011) implemented in Matlab (MATLAB, 2017). Scripts for both, preprocessing and analysis can be found on the Open Science Framework (https://osf.io/43t9q/?view_only=d71462e36d184fde8f64a6be60239b39). The signal was re-referenced offline to the linked mastoids (the algebraic average of M1 and M2) and down-sampled to 250 Hz. We applied a kaiser-windowed finite-impulse response high-pass filter with half-amplitude cutoff (-6 dB) of 0.3 Hz and a transition width of 0.3 Hz. We also applied a kaiser-windowed finite-impulse response low-pass filter with a half-amplitude cutoff (-6dB) of 30 Hz and a transition width of 5 Hz. Data were segmented and time-locked to the onset of the suffix, with a 600 ms pre-stimulus period (to include the onset of the verb stem) and 1300 ms post-suffix-onset period. Artefact rejection was performed semi-automatically. Segments of the signal exceeding a z-value of 7 were highlighted automatically and screened manually to reject muscle and coarse-movement artefacts. To correct ocular artefacts, we used an independent component analysis (ICA) (“runica”, implemented in the FieldTrip toolbox; Oostenveld et al., 2011), decomposed the data from all channels into 26 ICA components, and rejected components corresponding to blinks and saccades. Afterward, we shortened the baseline period to 400 ms pre-suffix onset for plotting and averaging.

2.5 Statistical analysis

For the statistical analysis, we used linear models (LMs; see Frömer et al., 2018 for use of LMs and linear mixed models with EEG data), as implemented in the lme4 package (Bates et al., 2015) in R (R Core Team, 2017). LMs offer a reliable way to analyze all three age groups in one statistical model. The temporal region of interest (ROI) was defined as 600-1000 ms relative to the onset of the suffix (-are, -ando) and the spatial ROI included the electrode positions F3, Fz, F4, C3, Cz, C4, P3, Pz, P4, based on a previous ERP study with the same paradigm in infants (Friederici et al., 2011). Thus, we computed mean amplitudes across the defined temporal and spatial ROIs for each subject.

2.5.1 Predictors of ERP polarity

Fifty-seven percent of the participants (N=66) showed a positive-going ERP effect (correct vs. incorrect) in the temporal and spatial ROIs, while the other 43% (N=49) showed a negative-going ERP effect. We here used a logistic regression to investigate whether children's ERP polarity (positive vs. negative) could be significantly predicted by their age (1 year, 2 years, 3 years; using linear contrast coding; see Schad et al. (2020) for further information on contrast coding), sex (male, female; using sum contrast coding), or language comprehension or language production abilities (z-transformed raw scores of the language tests). This analysis was performed in R (R Core Team, 2017).

2.5.2 Age effects of NAD learning

Using the temporal and spatial ROIs defined above, we set up an LM in R (R Core Team, 2017) to test for the effect of age on children's ERP effect (correct vs. incorrect), as an indicator for NAD learning. As the dependent variable, we used the amplitude of the absolute (based on the results of the linear regression, see section 3.1) ERP difference wave (incorrect – correct; see Figure 4) averaged over the temporal and spatial ROIs, further averaged over trials in order to increase the signal-to-noise ratio for the LM, resulting in one ERP amplitude value per participant (see Figure 3). When considering the ERP amplitude values entered into the LM, there was still considerable variability (mean: 0.059; SD = 6.73; min = -31.06, max = 15.69). Therefore, we excluded outliers, defined as 2.5 times the median absolute cutoff (Leys, Ley, Klein, Bernard, & Licata, 2013). This procedure resulted in the exclusion of the datasets of 6 additional children. The results of the linear model were the same regardless of outlier exclusion. The results reported in the main text do not include outliers; for the same analyses including outliers, see supplementary materials S1. We entered *age* (1 year, 2 years, 3 years; using linear contrast coding) as a fixed effect (independent variable) into the model. We further added weights to the model, accounting for the number of trials of each average.

3 Results

3.1 Predictors of ERP polarity

The logistic regression analysis revealed that none of the tested variables (age, sex, language comprehension, and language production) significantly predicted whether children showed a positive or negative ERP effect polarity (all $p>0.05$; Table 2). Therefore, we used the absolute amplitude as the dependent measure in the LM.

3.2 Age effects of NAD learning

A likelihood-ratio test revealed that the LM including the fixed effect *age* explained significantly more variance than a restricted model with the factor *age* omitted ($F = 6.11$ $p = 0.015$). There was a significant main effect of *age* ($\beta = -1.20$, $p = 0.013$; Table 3), indicating that the absolute ERP amplitude significantly decreased with increasing age (Figure 3).

We followed up on this significant effect of age with separate LMs for each age group. These LMs are equivalent to one-sample t-tests, but include weights for the number of trials that constituted each average. These one-sample t-tests revealed that for all three age groups, the absolute amplitude of the difference wave was significantly different from 0 (1 year: $\beta = 5.23$, $p < 0.001$; 2 years: $\beta = 3.62$, $p < 0.001$; $\beta = 3.48$, $p < 0.001$).

4 Discussion

Previous studies have demonstrated that NAD learning undergoes different developmental stages during early childhood (Culbertson et al., 2016; Mueller et al., 2019; van der Kant et al., 2020). The aim of the present study was to investigate whether the transition between these stages occurs in a gradual manner. To this end, we exposed 1- to 3-year-old children to Italian sentences containing NADs and measured children's ERP responses to incorrect sentences (containing NAD violations) compared to correct sentences (containing familiarized

NADs). Independent of the tested age, children's ERP responses revealed that they were able to distinguish correct from incorrect sentences, and thus learned NADs, indexed by an ERP component between 600 and 1000 ms (relative to suffix onset) with a positive polarity in 57% of all children and a negative polarity in 43%. Previous studies found an association of children's ERP polarity and later language outcomes (Kooijman et al., 2013; Schaadt et al., 2015). In addition, ERP polarity in this paradigm has been related to participants' age and/or the presence or absence of a task (Friederici et al., 2011, 2013; Mueller et al., 2009). Yet, neither children's age, sex, nor behaviorally tested language abilities did predict a given child's ERP polarity in our study. Importantly, regardless of ERP polarity, the amplitude of the ERP effect of NAD learning decreased linearly with age, which we interpret as a decrease in strength of NAD learning. In previous studies, ERP amplitude has been shown to be indicative of the strength of learning (Boll-Avetisyan et al., 2018), as well as behavior, such as tone and phoneme discrimination abilities (Garcia-Sierra et al., 2011; Kujala et al., 2001). Our findings of an ERP component between 600 and 1000 ms in children aged 1 to 3 years extends and corroborates previous studies using the same Italian sentence materials containing NADs. These studies showed that 4-month-old infants learned NADs, indicated by a late positive ERP component (640-1040 ms relative to suffix onset). This late positivity was interpreted to reflect associative learning, that is, infants learned associations between surface-level phonological features of the dependent elements (Friederici et al., 2011). Similarly, it is conceivable that the ERP effect in our study (regardless of polarity) in the same time-window is also indicative of associative learning. Our findings would then imply that across 1 to 3 years of age, children are in principle capable of learning NADs associatively from passive listening, but that the ability gradually decreases with increasing age. In line with this proposal, German adults have been shown to struggle to learn NADs from Italian sentences through passive listening (Friederici et al., 2013), but to successfully learn the NADs under active conditions, that is, in the presence of a task (Mueller et al., 2009; see also Pacton et al., 2015; Pacton & Perruchet, 2008). Together, these findings imply that during development, there is an initial stage of NAD learning, during which young children are able to learn NADs associatively, and a later stage, during which older children and adults need additional cues (e.g. Gómez, 2002; Grama & Wijnen, 2018; Newport & Aslin, 2004; for a review, see Wilson et al., 2018) or a task (Mueller et al., 2012; Pacton et al., 2015; Pacton & Perruchet, 2008) to guide their attention to successfully learn NADs. Our current findings contribute to this notion and, moreover, provide evidence that the transition between these developmental stages of NAD learning occurs in the form of a gradual decrease of associative learning during early childhood.

These findings raise the question of which processes underlie the developmental stages of NAD learning, including their transition. A behavioral study by Culbertson and colleagues (2016) proposed that an initial developmental stage of NAD learning (around 1 to 1.5 years of age) is characterized by associative learning of phonological features, while a later stage (around 2 years of age) is characterized by higher-level morphological learning. Our findings suggest that 2-year-old and even 3-year-old children are still able to learn NADs associatively from passive listening, but indicate that children show smaller effects for associative learning with increasing age. Considering Culbertson and colleagues' (2016) behavioral findings, it is possible that the ability to learn not only the surface-level phonological features, but also the higher-level morphological features of the NADs is slowly developing between 1 and 3 years, but this learning strategy was not triggered by our passive listening task. Similarly, it is possible that different measures, such as the head-turn preference procedure compared to ERPs and different NAD learning paradigms, tap into different learning processes. This difference in measures might also explain the differences between our study and a recent study using a similar paradigm with functional near-infrared spectroscopy (fNIRS), which found NAD learning from the same Italian sentences in 2-year-old, but not 3-year-old children (van der Kant et al., 2020). While fNIRS informs us about the brain areas underlying

NAD learning during early childhood, EEG may be more sensitive to detect children's decreased response to NAD violations at 3 years of age. Our results of a gradual decrease of associative NAD learning are therefore not necessarily at odds with previous studies reporting different developmental stages of NAD learning even before 3 years of age (Culbertson et al., 2016; van der Kant et al., 2020), but using electrophysiological measures and a passive listening design may have allowed to investigate children's associative learning abilities more closely.

In the following, we discuss two potential explanations for the observed gradual decrease of associative NAD learning during early childhood in the present study: (1) entrenchment of children's knowledge of their native language, and (2) maturational brain changes during early childhood. Regarding the former, children's early established (or entrenched) learning may influence expectations during later stages of learning (see Thiessen et al., 2016). These expectations facilitate subsequent learning of similar items, but hinder learning of new, dissimilar items. In line with this idea, the entrenchment has been shown to occur and hinder learning of new items in infants' use of lexical stress cues in word segmentation (Jusczyk et al., 1999) and in learning to read (Zevin & Seidenberg, 2002, 2004). Similarly, it is possible that through the course of early childhood, children's knowledge of the NADs in their native language becomes entrenched. Indeed, evidence from natural language studies show that native-language NAD learning slowly develops between 1 and 3 year. For example, French-learning infants can detect NAD violations in their native language starting around 14 months to 18 months, depending on the exact NADs tested (Culbertson et al., 2016; Nazzi et al., 2011; van Heugten & Shi, 2010). English-learning infants are able to detect NAD violations in '*is -ing*' constructions in their native language at 18 months, but not 15 months, and only when the NADs have no more than 3 intervening syllables (Santelmann & Jusczyk, 1998; see also Höhle et al., 2006 for evidence from German-learning infants). However, detecting violations does not necessarily mean that children learn the higher-level morphological rule and comprehend the meaning of the NADs, that is, successfully perform a sentence picture-matching task when the sentence contained NADs. These abilities seem to develop later, between 21 and 30 months (Culbertson et al., 2016; Legendre et al., 2010). It is possible that this increasing knowledge of children's native language NADs makes learning NADs in a foreign language (or an artificial language, such as our miniature version of Italian) more difficult with increasing age. Indeed, infants' NAD learning in an artificial language has been linked to processing NADs in their native language (Lany & Shoaib, 2020). This effect of entrenchment on learning novel NADs would explain why children's ability to learn NADs associatively decreased with age in our study. Taken together, children's knowledge of the NADs of their native language builds up over the first three years of life, possibly making learning of NADs in another language (such as our miniature version of Italian) more difficult for older children.

The second explanation of the gradual decrease in associative NAD learning refers to maturation of the developing brain. Associative NAD learning has been proposed to demand an interplay between posterior temporal brain areas and the premotor cortex (Friederici, 2012; Gervain et al., 2008, 2011; see Skeide & Friederici, 2016). These regions are involved in language comprehension and production more generally (Bruderer et al., 2015; Möttönen & Watkins, 2009; Rodd et al., 2015) and functionally connected through ventral and dorsal fiber pathway (Dubois et al., 2016; Perani et al., 2011). The ventral pathway is already well myelinated at birth (Perani et al., 2011) and available to infants for learning NADs at a very young age, likely providing the neurobiological basis of infants' ability to learn NADs associatively (see Friederici et al., 2011; Skeide & Friederici, 2016). In contrast, the development of higher-level learning of morpho-syntactic NADs is likely linked to the maturation of the prefrontal cortex and the arcuate fasciculus as the dorsal pathway, connecting the posterior temporal cortex and the pars opercularis (part of the inferior frontal gyrus), which has been shown to be specifically involved in syntactic processing in adults

(Rodd et al., 2015; Vigneau et al., 2006). The arcuate fasciculus, unlike the dorsal pathway which connects to the premotor cortex, shows continuous maturation until early adulthood, and has been linked to development of syntactic processing (Skeide et al., 2016). Specifically, in young children, syntactic information triggers activation in the left temporal cortex, but not the left inferior frontal cortex; at 3 years of age, at the latest, both regions are activated during syntactic processing (Dehaene-Lambertz et al., 2002, 2006; Skeide et al., 2016). In summary of these results, Skeide & Friederici (2016) proposed a transition from associative, bottom-up learning mainly based on the temporal cortex to higher-level, top-down learning around the age of 3 years involving the left inferior frontal cortex. Further studies need to evaluate whether higher-level morphological NAD learning shows a similar gradual *increase* during the developmental period, during which associative learning of phonological features *decreases*. The interplay between a decrease in bottom-up learning and an increase in top-down learning, driven by the continuous maturation of the arcuate fasciculus, most likely provides the neurobiological basis for the gradual decrease of associative learning of NADs during early childhood, as observed in the current study.

Overall, our results support the notion of a gradual transition between the two developmental stages of NAD learning. This gradual transition may point towards a sensitive period for associative NAD learning during early childhood. Under this view, younger children would have an advantage of learning NADs under passive listening conditions (Friederici et al., 2011; Mueller et al., 2019; van der Kant et al., 2020) compared to older children and adults (Friederici et al., 2013; Mueller et al., 2012; Mueller et al., 2019). This advantage would decrease gradually with age, as indicated by the linear decrease in our study. Older children and adults may still be able to learn NADs under passive listening in the presence of facilitating factors, such as additional cues (Frost & Monaghan, 2016; Gómez, 2002).

5 Conclusion

Our findings show a gradual decrease of associative NAD learning under passive listening during early childhood. Children at 1 to 3 years of age showed neurophysiological evidence of associative NAD learning under passive listening conditions, but the amplitude of this ERP effect linearly decreased with age. We propose that this linear decrease may be driven by entrenchment of children's knowledge of their native language NADs, which hinders NAD learning in a foreign language. In addition, brain maturation during early childhood likely contributes to children's increasing ability to utilize higher-level, morphological features of the input through top-down learning, and to their decreasing ability to learn NADs associatively under passive listening conditions. Our study provides evidence that the transition between different developmental stages of NAD learning occurs in a gradual manner, pointing toward a sensitive period for NAD learning during early childhood.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/doi:10.18637/jss.v067.i01.

Bayley, N. (2015). *Bayley Scales of Infant and Toddler Development—Third Edition (1. Auflage)*.

Boll-Avetisyan, N., Nixon, J. S., Lentz, T. O., Liu, L., van Ommen, S., Çöltekin, Ç., & van Rij, J. (2018). Neural Response Development During Distributional Learning. *Interspeech 2018*, 1432–1436. https://doi.org/10.21437/Interspeech.2018-2072

Bruderer, A. G., Danielson, D. K., Kandhadai, P., & Werker, J. F. (2015). Sensorimotor influences on speech perception in infancy. *Proceedings of the National Academy of Sciences*, *112*(44), 13531–13536. https://doi.org/10.1073/pnas.1508631112

Culbertson, J., Koulaguina, E., Gonzalez-Gomez, N., Legendre, G., & Nazzi, T. (2016). Developing knowledge of nonadjacent dependencies. *Developmental Psychology*, *52*(12), 2174–2183. https://doi.org/10.1037/dev0000246

Dehaene-Lambertz, G., Dehaene, S., & Hertz-Pannier, L. (2002). Functional Neuroimaging of Speech Perception in Infants. *Science*, *298*(5600), 2013–2015. https://doi.org/10.1126/science.1077066

Dehaene-Lambertz, G., Hertz-Pannier, L., Dubois, J., Mériaux, S., Roche, A., Sigman, M., & Dehaene, S. (2006). Functional organization of perisylvian activation during presentation of sentences in preverbal infants. *Proceedings of the National Academy of Sciences*, *103*(38), 14240–14245. https://doi.org/10.1073/pnas.0606302103

Dubois, J., Poupon, C., Thirion, B., Simonnet, H., Kulikova, S., Leroy, F., Hertz-Pannier, L., & Dehaene-Lambertz, G. (2016). Exploring the Early Organization and Maturation of Linguistic Pathways in the Human Infant Brain. *Cerebral Cortex*, *26*(5), 2283–2298. https://doi.org/10.1093/cercor/bhv082

Friederici, A. D. (2012). The cortical language circuit: From auditory perception to sentence comprehension. *Trends in Cognitive Sciences*, *16*(5), 262–268. https://doi.org/10.1016/j.tics.2012.04.001

Friederici, A. D., Mueller, J. L., & Oberecker, R. (2011). Precursors to Natural Grammar Learning: Preliminary Evidence from 4-Month-Old Infants. *PLoS ONE*, *6*(3), e17920. https://doi.org/10.1371/journal.pone.0017920

Friederici, A. D., Mueller, J. L., Sehm, B., & Ragert, P. (2013). Language Learning without Control: The Role of the PFC. *Journal of Cognitive Neuroscience*, *25*(5), 814–821. https://doi.org/10.1162/jocn_a_00350

Frömer, R., Maier, M., & Abdel Rahman, R. (2018). Group-Level EEG-Processing Pipeline for Flexible Single Trial-Based Analyses Including Linear Mixed Models. *Frontiers in Neuroscience*, *12*, 48. https://doi.org/10.3389/fnins.2018.00048

Frost, R. L. A., & Monaghan, P. (2016). Simultaneous segmentation and generalisation of non-adjacent dependencies from continuous speech. *Cognition*, *147*, 70–74. https://doi.org/10.1016/j.cognition.2015.11.010

Garcia-Sierra, A., Rivera-Gaxiola, M., Percaccio, C. R., Conboy, B. T., Romo, H., Klarman, L., Ortiz, S., & Kuhl, P. K. (2011). Bilingual language learning: An ERP study relating early brain responses to speech, language input, and later word production. *Journal of Phonetics*, *39*(4), 546–557. https://doi.org/10.1016/j.wocn.2011.07.002

Gervain, J., Berent, I., & Werker, J. F. (2011). Binding at Birth: The Newborn Brain Detects Identity Relations and Sequential Position in Speech. *Journal of Cognitive Neuroscience*, *24*(3), 564–574. https://doi.org/10.1162/jocn_a_00157

Gervain, J., Macagno, F., Cogoi, S., Peña, M., & Mehler, J. (2008). The neonate brain detects speech structure. *Proceedings of the National Academy of Sciences*, *105*(37), 14222–14227. https://doi.org/10.1073/pnas.0806530105

Gómez, R. L. (2002). Variability and detection of invariant structure. Psychological Science, 13 (5), 6.

Gómez, R., & Maye, J. (2005). The Developmental Trajectory of Nonadjacent Dependency Learning. *Infancy*, *7*(2), 183–206. https://doi.org/10.1207/s15327078in0702_4

Grama, I., & Wijnen, F. (2018). Learning and generalizing non-adjacent dependencies in 18-month-olds: A mechanism for language acquisition? *PLOS ONE*, *13*(10), e0204481. https://doi.org/10.1371/journal.pone.0204481

Grimm, H. (2015). Sprachentwicklungstest für drei- bis fünfjährige Kinder (3;0-5;11 Jahre) Diagnose von Sprachverarbeitungsfähigkeiten und auditiven Gedächtnisleistungen (3., überarbeitete und neu normierte Auflage). *Göttingen: Hogrefe*.

Grimm, H., Aktas, M., & Frevert, S. (2016). SETK-2. Sprachentwicklungstest für zweijährige Kinder (2;0-2;11 Jahre). Diagnose rezeptiver und produktiver

Sprachverarbeitungsfähigkeiten (2., überarbeitete und neu normierte Auflage). *Göttingen: Hogrefe*.

Hartshorne, J. K., Tenenbaum, J. B., & Pinker, S. (2018). A critical period for second language acquisition: Evidence from 2/3 million English speakers. *Cognition*, *177*, 263–277. https://doi.org/10.1016/j.cognition.2018.04.007

Hohle, B., Schmitz, M., Santelmann, L. M., & Weissenborn, J. (2006). The Recognition of Discontinuous Verbal Dependencies by German 19-Month-Olds: Evidence for Lexical and Structural Influences on Children's Early Processing Capacities. *Language Learning and Development*, *2*(4), 277–300. https://doi.org/10.1207/s15473341lld0204_3

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, *21*(1), 60–99. https://doi.org/10.1016/0010-0285(89)90003-0

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The Beginnings of Word Segmentation in English-Learning Infants. *Cognitive Psychology*, *39*(3–4), 159–207. https://doi.org/10.1006/cogp.1999.0716

Knudsen, E. I. (2004). Sensitive Periods in the Development of the Brain and Behavior. *Journal of Cognitive Neuroscience*, *16*(8), 1412–1425. https://doi.org/10.1162/0898929042304796

Kooijman, V., Junge, C., Johnson, E. K., Hagoort, P., & Cutler, A. (2013). Predictive Brain Signals of Linguistic Development. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00025

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, *5*(11), 831–843. https://doi.org/10.1038/nrn1533

Kuhl, P. K. (2010). Brain Mechanisms in Early Language Acquisition. *Neuron*, *67*(5), 713–727. https://doi.org/10.1016/j.neuron.2010.08.038

Kujala, T., Kallio, J., Tervaniemi, M., & Näätänen, R. (2001). The mismatch negativity as an index of temporal processing in audition. *Clinical Neurophysiology*, *112*(9), 1712–1719. https://doi.org/10.1016/S1388-2457(01)00625-3

Lany, J., & Shoaib, A. (2020). Individual differences in non-adjacent statistical dependency learning in infants. *Journal of Child Language*, *47*(2), 483–507. https://doi.org/10.1017/S0305000919000230

Legendre, G., Barrière, I., Goyet, L., & Nazzi, T. (2010). Comprehension of Infrequent Subject–Verb Agreement Forms: Evidence From French-Learning Children. *Child Development*, *81*(6), 1859–1875. https://doi.org/10.1111/j.1467-8624.2010.01515.x

Möttönen, R., & Watkins, K. E. (2009). Motor Representations of Articulators Contribute to Categorical Perception of Speech Sounds. *Journal of Neuroscience*, *29*(31), 9819–9825. https://doi.org/10.1523/JNEUROSCI.6018-08.2009

Mueller, J. L., Friederici, A. D., & Mannel, C. (2012). Auditory perception at the root of language learning. *Proceedings of the National Academy of Sciences*, *109*(39), 15953–15958. https://doi.org/10.1073/pnas.1204319109

Mueller, J. L., Friederici, A. D., & Männel, C. (2019). Developmental changes in automatic rule□learning mechanisms across early childhood. *Developmental Science*, *22*(1), e12700. https://doi.org/10.1111/desc.12700

Mueller, J. L., Oberecker, R., & Friederici, A. D. (2009). Syntactic learning by mere exposure—An ERP study in adult learners. *BMC Neuroscience*, *10*(1), 89. https://doi.org/10.1186/1471-2202-10-89

Nazzi, T., Barrière, I., Goyet, L., Kresh, S., & Legendre, G. (2011). Tracking irregular morphophonological dependencies in natural language: Evidence from the acquisition of subject-verb agreement in French. *Cognition*, *120*(1), 119–135. https://doi.org/10.1016/j.cognition.2011.03.004

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, *48*(2), 127–162. https://doi.org/10.1016/S0010-0285(03)00128-2

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). *FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data* [Research Article]. Computational Intelligence and Neuroscience; Hindawi. https://doi.org/10.1155/2011/156869

Pacton, S., & Perruchet, P. (2008). An attention-based associative account of adjacent and nonadjacent dependency learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(1), 80–96. https://doi.org/10.1037/0278-7393.34.1.80

Pacton, S., Sobaco, A., & Perruchet, P. (2015). Is an attention-based associative account of adjacent and nonadjacent dependency learning valid? *Acta Psychologica*, *157*, 195–199. https://doi.org/10.1016/j.actpsy.2015.03.002

Perani, D., Saccuman, M. C., Scifo, P., Anwander, A., Spada, D., Baldoli, C., Poloniato, A., Lohmann, G., & Friederici, A. D. (2011). Neural language networks at birth. *Proceedings of the National Academy of Sciences*, *108*(38), 16056–16061. https://doi.org/10.1073/pnas.1102991108

R Core Team. (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rodd, J. M., Vitello, S., Woollams, A. M., & Adank, P. (2015). Localising semantic and syntactic processing in spoken and written language comprehension: An Activation Likelihood Estimation meta-analysis. *Brain and Language*, *141*, 89–102. https://doi.org/10.1016/j.bandl.2014.11.012

Santelmann, L. (1998). Sensitivity to discontinuous dependencies in language learners: Evidence for limitations in processing space. *Cognition*, *69*(2), 105–134. https://doi.org/10.1016/S0010-0277(98)00060-2

Schaadt, G., & Männel, C. (2019). Phonemes, words, and phrases: Tracking phonological processing in pre-schoolers developing dyslexia. *Clinical Neurophysiology*, *130*(8), 1329–1341. https://doi.org/10.1016/j.clinph.2019.05.018

Schaadt, G., Männel, C., van der Meer, E., Pannekamp, A., Oberecker, R., & Friederici, A. D. (2015). Present and past: Can writing abilities in school children be associated with their auditory discrimination capacities in infancy? *Research in Developmental Disabilities*, *47*, 318–333. https://doi.org/10.1016/j.ridd.2015.10.002

Schaadt, G., Paul, M., Muralikrishnan, R., Männel, C., & Friederici, A. D. (2020). Seven-year-olds recall non-adjacent dependencies after overnight retention. *Neurobiology of Learning and Memory*, *171*, 107225. https://doi.org/10.1016/j.nlm.2020.107225

Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, *110*, 104038. https://doi.org/10.1016/j.jml.2019.104038

Senghas, A., Kita, S., & Özyürek, A. (2004). Children Creating Core Properties of Language: Evidence from an Emerging Sign Language in Nicaragua. *Science*, *305*(5691), 1779–1782. https://doi.org/10.1126/science.1100199

Singleton, D. (2005). The Critical Period Hypothesis: A coat of many colours. *International Review of Applied Linguistics in Language Teaching*, *43*(4), 269–285. https://doi.org/10.1515/iral.2005.43.4.269

Skeide, M. A., Brauer, J., & Friederici, A. D. (2016). Brain Functional and Structural Predictors of Language Performance. *Cerebral Cortex*, *26*(5), 2127–2139. https://doi.org/10.1093/cercor/bhv042

Skeide, M. A., & Friederici, A. D. (2016). The ontogeny of the cortical language network. *Nature Reviews Neuroscience*, *17*(5), 323–332. https://doi.org/10.1038/nrn.2016.23

Thiessen, E. D., Girard, S., & Erickson, L. C. (2016). Statistical learning and the critical period: How a continuous learning mechanism can give rise to discontinuous learning:

Statistical learning and the critical period. *Wiley Interdisciplinary Reviews: Cognitive Science*, *7*(4), 276–288. https://doi.org/10.1002/wcs.1394

van der Kant, A., Männel, C., Paul, M., Friederici, A. D., Höhle, B., & Wartenburger, I. (2020). Linguistic and non-linguistic non-adjacent dependency learning in early development. *Developmental Cognitive Neuroscience*, 100819. https://doi.org/10.1016/j.dcn.2020.100819

van Heugten, M., & Shi, R. (2010). Infants' sensitivity to non-adjacent dependencies across phonological phrase boundaries. *The Journal of the Acoustical Society of America*, *128*(5), EL223–EL228. https://doi.org/10.1121/1.3486197

Vigneau, M., Beaucousin, V., Hervé, P. Y., Duffau, H., Crivello, F., Houdé, O., Mazoyer, B., & Tzourio-Mazoyer, N. (2006). Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. *NeuroImage*, *30*(4), 1414–1432. https://doi.org/10.1016/j.neuroimage.2005.11.002

Wilson, B., Spierings, M., Ravignani, A., Mueller, J. L., Mintz, T. H., Wijnen, F., van der Kant, A., Smith, K., & Rey, A. (2018). Non-adjacent Dependency Learning in Humans and Other Animals. *Topics in Cognitive Science*. https://doi.org/10.1111/tops.12381

Zevin, J. D., & Seidenberg, M. S. (2002). Age of Acquisition Effects in Word Reading and Other Tasks. *Journal of Memory and Language*, *47*(1), 1–29. https://doi.org/10.1006/jmla.2001.2834

Zevin, J. D., & Seidenberg, M. S. (2004). Age-of-acquisition effects in reading aloud: Tests of cumulative frequency and frequency trajectory. *Memory & Cognition*, *32*(1), 31–38. https://doi.org/10.3758/BF03195818

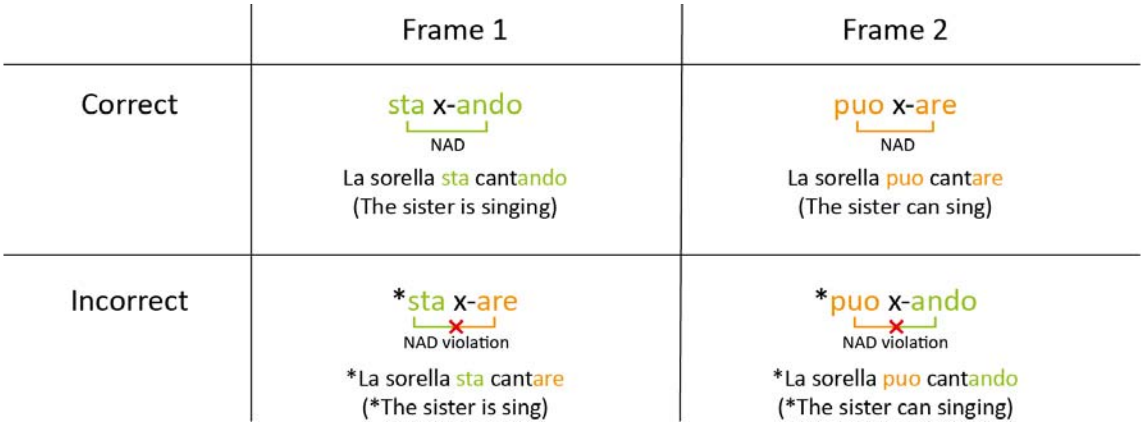
Figure Captions

Figure 1. Visualization of the stimuli for the EEG experiment with examples, adapted from Friederici and colleagues (2011). Ungrammatical sentences and frames are marked with an asterisk. Unicolored brackets visualize non-adjacent dependencies (NADs). Bicolored brackets and red crosses indicate NAD violations.

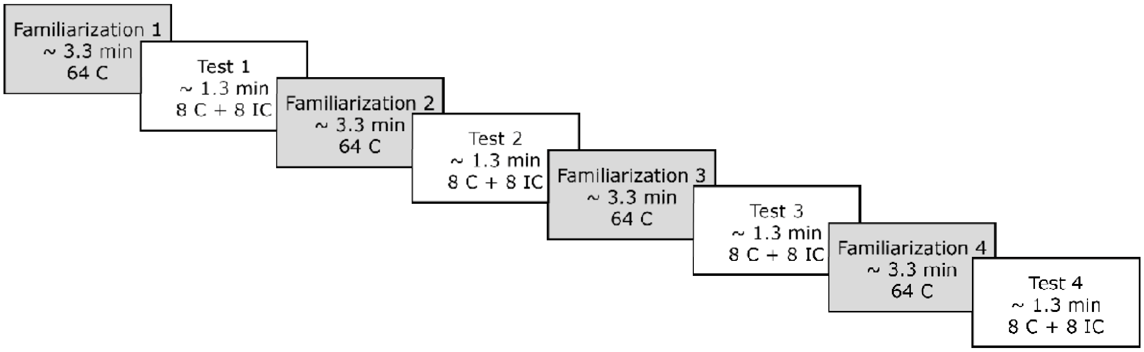
Figure 2. Experimental procedure: alternating familiarization and test phases. Participants listened to four familiarization phases (64 correct trials each) alternated with four test phases (eight correct and eight incorrect trials each). C: correct, IC: incorrect. Figure adapted from Friederici, Mueller, and Oberecker (2011).

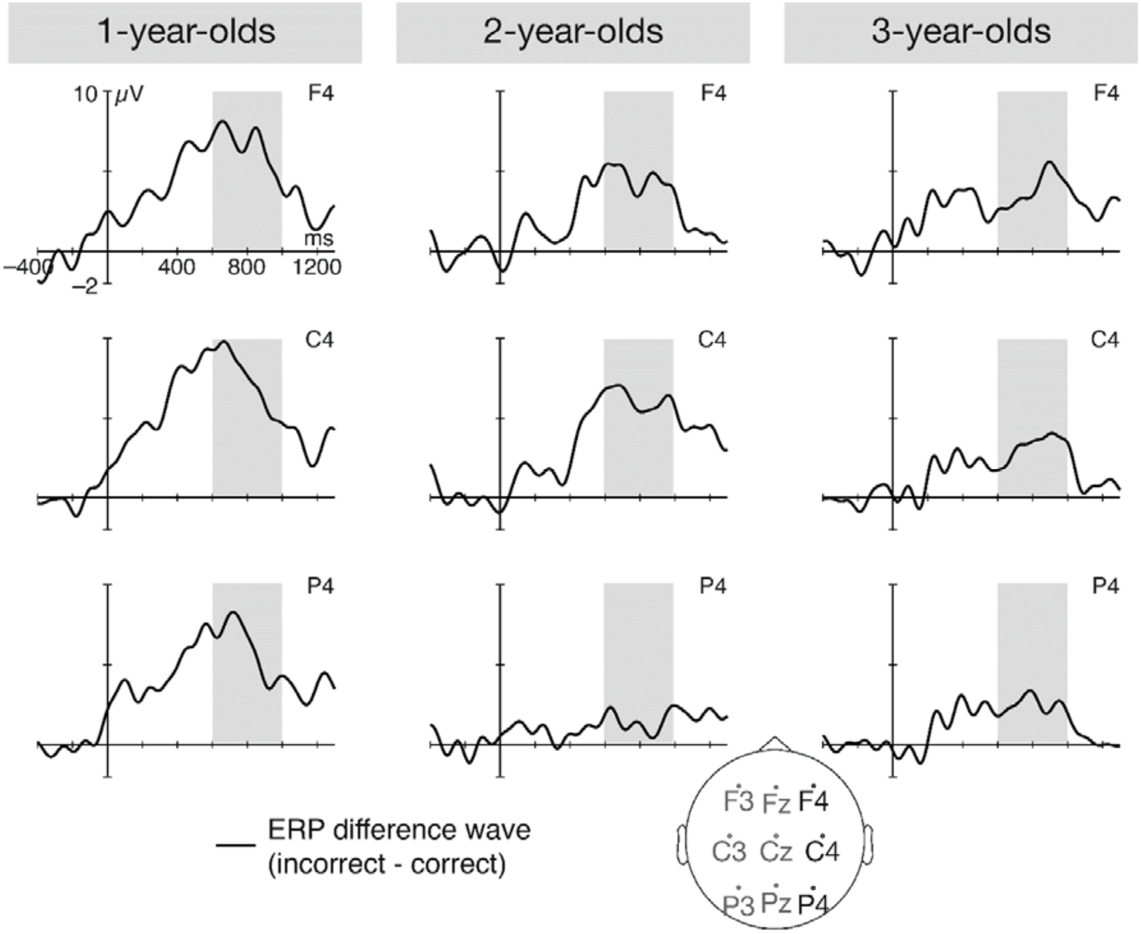
Figure 3. A. Amplitude of the ERP difference wave (incorrect - correct) across all age groups. B. Absolute amplitude of the ERP difference wave (incorrect - correct) plotted by age. Absolute ERP amplitude is significantly predicted by age.

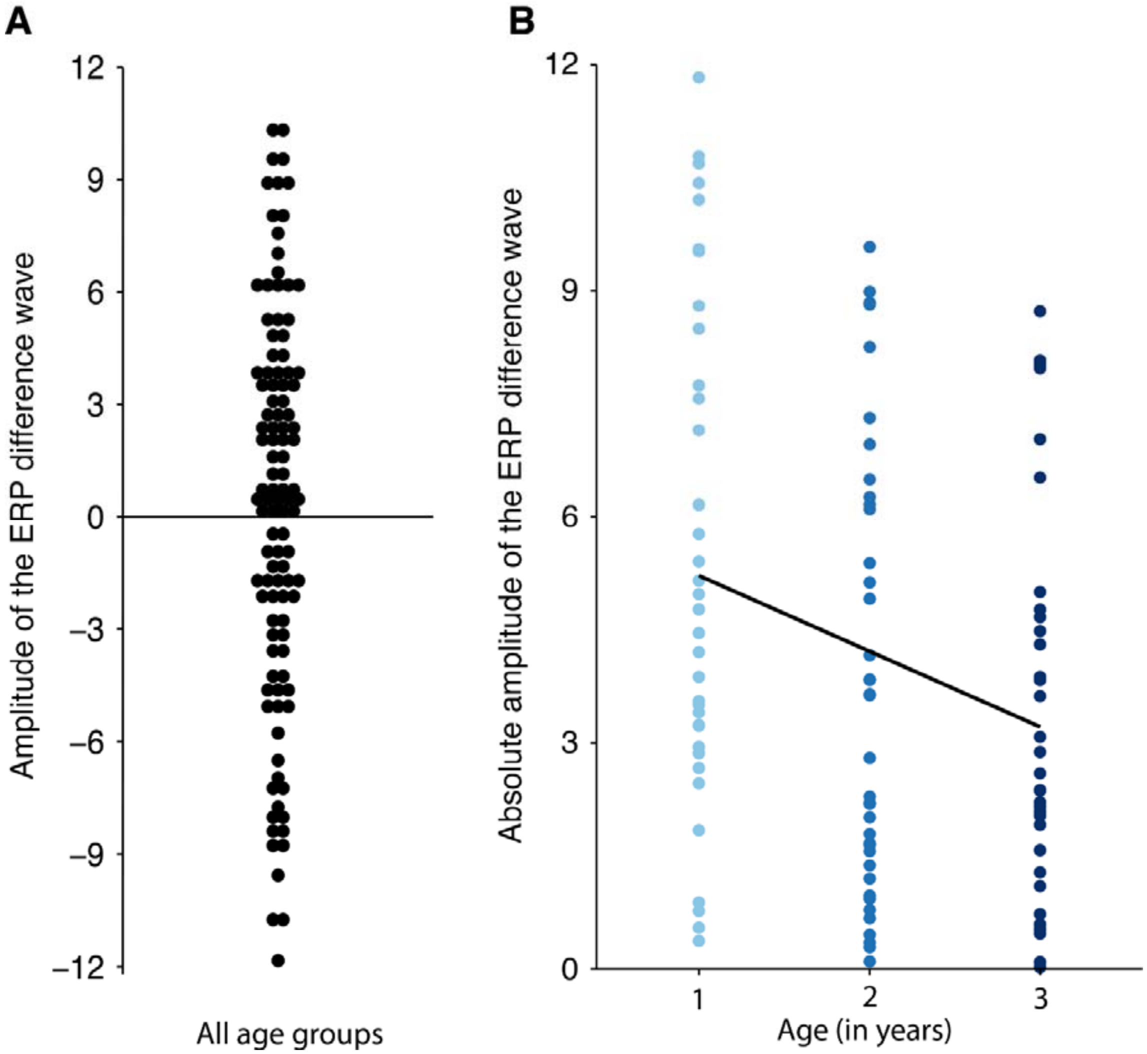
Figure 4. Grand-average ERP difference waves (incorrect - correct) for the three age groups. To account for the absolute amplitudes used in the analyses, the ERP difference wave for the children with negative polarity is flipped for visualization purposes.



	Frame 1	Frame 2
Correct	sta x-ando NAD La sorella sta cantando (The sister is singing)	puo x-are NAD La sorella puo cantare (The sister can sing)
Incorrect	*sta x-are NAD violation *La sorella sta cantare (*The sister is sing)	*puo x-ando NAD violation *La sorella puo cantando (*The sister can singing)




1-year-olds
2-year-olds
3-year-olds
10
μV
F4
–400
–2
400
800
1200
ms
C4
P4
ERP difference wave
(incorrect - correct)
F3
Fz
F4
C3
Cz
C4
P3
Pz
P4


A
12
9
6
3
0
−3
−6
−9
−12
Amplitude of the ERP difference wave
All age groups
B
12
9
6
3
0
Absolute amplitude of the ERP difference wave
1
2
3
Age (in years)

Tables

Table 1. Overview of participants and trials. "Additional participants" refers to the ones additionally tested, but excluded based on either non-compliance or insufficient number of artefact-free EEG trials. Correct and incorrect trials refer to the average number of included trials in the correct and incorrect condition during test phases. The difference in trial numbers between the correct and incorrect conditions was not significant for any age group (all p>0.1).

	1 year	2 years	3 years
Participants (N)	40	40	35
Mean age (months)	12.80 (SD: 0.54)	25.08 (SD: 0.88)	37.10 (SD: 0.60)
Sex	20 girls	16 girls	18 girls
Additional participants (N)	28	15	6
Correct Trials (N)	21.90 (SD: 4.24)	24.25 (SD: 4.47)	26.77 (SD: 4.15)
Incorrect Trials (N)	22.53 (SD: 3.85)	23.93 (SD: 4.70)	27.26 (SD: 3.59)

Table 2. Summary of the logistic regression to predict children's ERP effect polarity

Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	0.71	0.19 – 2.71	0.619
age	1.03	0.52 – 2.05	0.922
sex	1.44	0.63 – 3.38	0.390
lang. comp.	1.16	0.71 – 1.92	0.565
lang. prod.	1.09	0.68 – 1.77	0.713
Observations	99		
R^2 Tjur	0.014		

Note: lang. comp. – language comprehension. lang. prod. – language production

Table 3. Summary of the LM predicting children's ERP amplitudes. Statistically significant p-values are highlighted in bold.

Predictors	*Estimates*	*CI*	*p*
(Intercept)	4.10	3.55 – 4.65	**<0.001**
age	-1.20	-2.17 – -0.24	**0.015**
Observations	109		
R^2 / R^2 adjusted	0.054 / 0.045		

Note: inc – incorrect condition, cor – correct condition