Lexical access during naturalistic listening in middle childhood and early adolescence

Briony Waite,[1] Tanya Levari,[1] Anthony Yacovone,[1,2] Jesse Snedeker[1] | [1]Harvard University, [2]Boston University

Background. To understand spoken language, listeners must quickly recognize words and retrieve their meanings. In adults, this process of lexical access is facilitated by predicting (or *pre*-activating) upcoming words.[1] ERP studies with adults show that lexical predictability is strongly correlated with N400 amplitude (a component indexing ease of lexical access).[2-3] Recent work has documented similar patterns in children (5–10),[4] but, we know little about whether and how lexical access and prediction during spoken language comprehension changes over middle childhood and early adolescence. These age groups are of particular interest because the school years are a time of rapid growth in reading ability, vocabulary, and discourse comprehension. To fill this gap, the current study investigates the effect of predictability on the N400 in naturalistic language comprehension in 9 and 14-year-old children. If linguistic prediction depends on language experience, we might expect larger effects of predictability on the N400 1) in older children and 2) in children with larger vocabularies.

Method. Data collection is ongoing. We test two age groups: middle childhood and young adolescents. The current sample includes 20 children (*M* = 9.9 yrs, *range* = 9.5–10.5 yrs) and 20 adolescents (*M* = 14.6 yrs, *range* = 14–15 yrs). Our final sample will have 30 per age group. Participants listened to a 30-minute children's story while continuous EEG was recorded. We analyzed N400 amplitude to each of the 2,614 content words (electrodes Cz and Pz, 350–550ms). We also measured participants' vocabulary using the PPVT[6] (14s: *M* = 118, *range* = 98–139; 9s: *M* = 118, *range* = 101–146, standard scores). We measured predictability (cloze), semantic association (word2vec), and lexical frequency (SUBTLEX$_{WF}$) values for all content words, as well as control variables including concreteness, acoustic length, word position in sentence and story, and baseline effects of immediately surrounding words.

Results. We used linear mixed-effects models to assess which lexical properties predict the N400 response across and within age groups. In keeping with our design (two age groups lacking continuity), we treated age categorically. Cloze significantly predicts N400 amplitude in both age groups ($p < .05$), with no reliable effects of Frequency and Semantic Association. There was also a significant three-way interaction between Cloze, Age Group (9s and 14s), and PPVT standard score ($b = 0.04$, *SE* = 0.01, $p < .05$), such that 9-year-olds with higher PPVT scores show a stronger effect of Cloze than 9-year-olds with lower PPVT scores but there is no difference in the Cloze effect by PPVT score for 14-year-olds.

Discussion and conclusions. Our preliminary results show changes in lexical access during naturalistic listening from ages 9–14. Both age groups are sensitive to top-down cues (predictability). However, vocabulary only modulates the use of these cues in younger children. This is consistent with a theory which posits that vocabulary acquisition helps children construct top-down representations during discourse processing, facilitating prediction. These data contrast prior EEG research with native adult speakers that finds no effect of individual language ability on N400 amplitude.[7] We see at least two ways to reconcile these findings: First, language ability may play a role in linguistic prediction while predictive processes are still developing, such that language ability modulates reliance on top-down factors in younger children, but this effect slowly disappears across middle childhood. Second, rather than age, the difficulty level of the discourse may determine whether language ability affects prediction. Our story has a third-grade reading level, making it age-appropriate for our 9-year-olds but likely too easy for our 14-year-olds. Thus, language ability might modulate reliance on top-down factors during comprehension of discourses at emergent levels of understanding. Future work should explore the effects of discourse difficulty in younger children and adults. Future work will also explore the role of literacy in predictive processing, given that vocabulary and literacy are highly correlated and children in our age range are developing readers.[8]

References. [1]Altmann & Kamide, 1999; [2]Kuperberg & Jaeger, 2016; [3]Federmeier & Kutas, 1999; [4]Levari & Snedeker, 2024; [5]Huettig & Pickering, 2019; [6]Dunn & Dunn, 2007; [7]Kim et al., 2018; [8]Duff et al., 2015.

Figure 1. Amplitude of the N400 for highest cloze (black), mid cloze (red), and lowest cloze (blue) words, split by age and vocabulary.
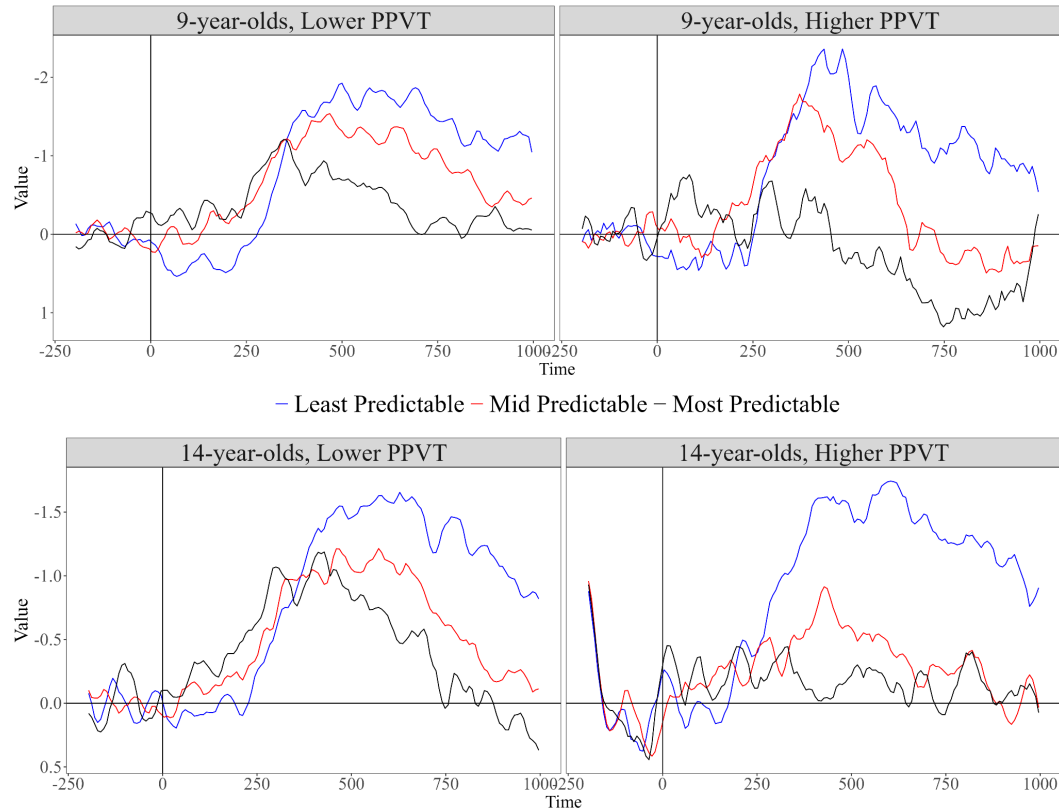
Table 1. Final model with both age groups included. Positive betas indicate smaller N400s. Control variables (e.g. features of prior word) are only included when significant.

Factors	Beta	Std. Error	t value	p value
Intercept	-2.89	1.75	-1.65	0.11
Age Group	-1.10	2.34	-0.47	0.64
PPVT Standard Score	0.02	0.01	1.22	0.23
Cloze	1.53	0.97	1.58	0.12
Log Frequency Prior Word	-0.21	0.10	-2.13	0.03 *
Cloze Prior Word	0.23	0.10	2.34	0.02 *
Interaction Terms:				
Age Group:PPVT	0.01	0.02	0.33	0.75
Age Group:Cloze	-3.86	1.31	-2.95	0.00 **
PPVT:Cloze	-0.01	0.01	-1.20	0.23
Age Group:PPVT:Cloze	0.04	0.01	3.20	0.00 **