Decoding Early Vocabulary Acquisition: Naturalistic evidence from EEG encoding models across the First Five Years

Katharina Menn[1,2], Claudia Männel[2,3], Florian Scharf[4] & Lars Meyer[1,5]

[1]RG Language Cycles, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, [2]Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, [3]Department of Audiology and Phoniatrics, Charité – Universitätsmedizin Berlin, [4]University of Kassel , Germany, [5]Clinic for Phoniatrics and Pedaudiology, University Hospital Münster

Infant language acquisition is marked by rapid learning, with comprehension of some basic words emerging within the first year of life. Measuring the earliest stages of lexical acquisition experimentally remains challenging due to infants' limited attention and overt response capacities. In this study, we assess the acquisition of lexical knowledge from electrophysiological (EEG) recordings of infants listening to naturalistic children's stories. Our quasi-longitudinal dataset comprised a final sample of 51 children (31 female) between 3 months to 5 years of age, assessed twice within a 3-month time window. Children heard translation-equivalent stories in their native language (German) and an unfamiliar language (French). We quantified word processing from continuous EEG through temporal response functions of EEG deconvolution models. Analyses revealed a significant word onset response between 100–300 ms in the native language, which increased with infant age (t=2.34, p=.021). This response was absent in the unfamiliar language. Together, this indicates that infants' brains registered the onsets of native words in fluent speech. To assess whether this effect indeed reflects the neural representation of word forms in infants' emerging mental lexica, we tested for a modulation of onset responses by age of acquisition. We calculated the probability of a child producing a word based on vocabulary trajectories from Wordbank. Strikingly, our mixed-effects models demonstrate a decrease in the word onset response for more familiar words (t=-2.17, p=.029). We interpret this as evidence that familiar words are processed more efficiently. More generally, it suggests that infants' emerging lexical representations are mirrored in their EEG responses to natural speech. In sum, we provide first evidence for the usefulness of EEG deconvolution to trace the emergence of the mental lexicon during naturalistic stimulation. Using language models and syntactic annotations, we currently aim for applying our EEG methodology for revealing the emergence of semantic and syntactic processing.