

Using Eye Movements to Study Infants' Visual Attention in Real-world Word Learning Contexts

Jenny Yang, Yayun Zhang, Chen Yu

Children have incredible language learning abilities. By 30 months, children already acquire over 500 productive vocabularies, many of which are object names. How do they become such skilled language learners? To learn object names, infants must link what they hear with what they see. However, daily scenes are often cluttered with multiple objects, making it challenging for infants to form the correct word-object associations. How do they solve this problem of referential uncertainty in which one word can refer to multiple different objects? Previous laboratory experiments studying word learning utilized simplified stimuli such as pictures of equally sized objects displayed on a solid background, which is not representative of the naturalistic settings infants are usually exposed to. Given the complex visual information that infants see, we want to understand how they visually select objects when object names are heard. In laboratory experiments, looking behavior has been primarily used as the outcome measure of infant word learning: Infants “know” a name when they look at the right referent upon hearing that name. Therefore, in the current study, we hope to understand the changes in looking behavior as a function of experience in 8 to 24-month-old infants as well as the learning mechanism through which they build name-object mappings.

The current study measured how infants' real-time looking behaviors change over the course of word learning. To do this, we first selected representative naming-moment frames from children's videos collected using head-mounted cameras. These videos capture infants' first-person view and the objects they are attending to during naturalistic toy play interactions. Using these egocentric videos, two types of naming moments were defined: 1) the labeled object is the only dominant object in view (high informative) and 2) all objects in view are equally salient (low informative). For the stimuli in our experiment, we recreated high and low informative scenes from the egocentric videos with novel toys. Using a screen-based eye-tracker, we recorded infants' gazes while presenting them with a series of high and low informative scenes mimicking infants' realistic word learning environment. In each series, one target object was labeled to resemble naturalistic naming moments. We analyzed infants' moment-by-moment gaze to examine 1) what objects they prefer to attend to (Object Looks), 2) whether or not they looked at the target object (Target Looks), and 3) if labeling affected their looking behavior (Labeling Effect).

Results illustrated that 1) infants consistently attended to a subset of all of the objects in view. This selective attention may help infants during word learning by reducing the referential uncertainty in the environment. 2) Infants also preferred attending to the target more in high informative scenes than in low informative scenes, demonstrating that more visually salient objects appear to draw infants' attention. 3) We discovered no difference in infant looking behavior before and after the label, meaning that infants do not change their existing looking patterns due to labeling. This suggests that caregivers can better help infants find the correct word-object mappings by naming objects infants are already attending to rather than redirecting their attention to another object.

Through this experiment, we recognized the importance of using more realistic stimuli to study early language learning. Using our methodology, we were able to study word learning in infants using naturalistic stimuli without losing the control of laboratory experiments. By collecting infants' real-time gazes through eye-tracking, we were able to understand how infants address the referential uncertainty problem and uncover mechanisms through which they learn to map words to the correct objects in real time. This is a critical first step in discovering how infants form word-object associations as the information they choose to attend to is what enters their system for cognitive processing.