Word-learning from visual prevalence: evidence from first-person infant views

Elizabeth M. Clerkin, Chen Yu, & Linda B. Smith

Department of Psychological and Brain Sciences, Indiana University
Breaking into word learning
The problem of uncertainty & clutter
A mapping problem

- This conceptualization is at the heart of current debates about cross-situational word-referent learning (Medina, Snedeker, Trueswell, & Gleitman, 2011; Smith, Suanda, & Yu, 2014)
Today: The very early side of object name learning

8 to 10 month olds
Today: The problem is on the visual side

Long-time assumption that basic level categories are perceptually “given” (Rosch, et al., 1976; Gentner, 1982)

However, human vision and machine vision research shows that visual object recognition is not a trivial problem but one that requires massive visual experience with specific categories (Kourtzi & Connor, 2011; Pinto, Cox, & DiCarlo, 2008)

We know a lot about the words infants hear and how they are relevant to infant’s later word learning (see Hart & Risley, 1995)
The natural statistics of objects in infants’ views

We propose that the natural distribution of objects in infants’ day to day visual environments can help solve the problems of referential uncertainty and visual object recognition

Many natural distributions are extremely right-skewed

If a small set of objects are **pervasively** present, they could form an early select set of recognized objects for word learning
Infant perspective scenes
Head Camera (Looxcie)
Corpus collection

8 infants (3 male)
Ranging in age from 8.5 to 10.5 months, M = 9.2 months

8.5 total hours of footage (mean per subject = 1.1, SD = 0.54)
The total number of individual mealtime events in the sample was 147
  • Mean duration of 3.5 min (SD = 7.2)
The 917,207 (total) frames in the mealtime corpus were sampled at 1/5 Hz for coding
  • This yielded a total of 5,775 coded scenes
Corpus collection

- Images analyzed were captured during infant mealtimes
- Meal time = any eating/food related event
  - Occurs multiples times per day everyday
  - Likely involves a large variety of common objects
  - Occurs in many different contexts
Coding of the corpus

About 500 naïve adults (Amazon Mechanical Turk)
Task: name the five most obvious objects in the scene, using basic level nouns
Coders saw images in sequential sets of 20
Each scene was coded by 4 individuals
Object name categories

745 unique object names

First Nouns (133) – normatively learned by 16 months (Communicative Developmental Inventory - CDI)

Early Nouns (59) – normatively learned between 16 and 30 months (Toddler CDI)

Later Nouns (553) – other nouns named by coders that were not on either inventory

• AoA: 6.15 years (SD=1.56)
Object Occurrence by AoA Category

Proportion of Images

Object Name AoA Category

First

Early

Later

Mean

Median
Are parents naming these objects?

• Could visual pervasiveness itself – without high-frequency naming – support learning first object names?
• We think that it might:
  • By increasing the strength of visual object representations in cluttered scenes
  • And thus the likelihood of selecting, remembering, and thereby aggregating an object as a referent
Are parents naming these objects?

• The speech from these infant head camera videos was transcribed in entirety
  • Including child-directed speech, adult-directed speech, etc.
• “Naming moments” were extracted for each of the 745 objects that occurred in these scenes
Word Occurrence by AoA Category

Proportion of Segments Containing Speech

First  Early  Later

Object Name AoA Category

- Mean
- Median
Words and Objects in Infants' Mealtime Environments

Number of occurrences vs. Rank order by object frequency

Objects
Words
An associative model of word-referent learning

• 50 most frequent objects in each AoA category

• In each trial of the simulation, the system was presented with 5 object names drawn from 6 different naming frequencies and all the objects coders said were presented in that scene

• 6 different naming frequencies:
  • Entirely random - that is: First, Early, and Later Nouns were named equally often and with no regard to the objects in the scene
  • The objects present in a scene and only those objects had their names presented to the system
  • The other 4 conditions are intermediate between these two
An associative model of word-referent learning

• A co-occurrence matrix was built from these training trials
• On 1,000 simulations, the system was probed on the 50 most frequent nouns from the First, Early, and Later Nouns lists
• On the test trials, if the maximal cell in the co-occurrence matrix for each object name was its corresponding object, the trial was counted as correct
How do infants break into learning object names?

The pervasiveness of a few object categories present – day in and day out – in infants’ visual environments may solve referential ambiguity, and this visual learning may begin before word learning

The non-uniform frequency statistics of visual objects in every day scenes brings forward a signal of likely referents in a visually noisy and cluttered world
Acknowledgements

Our participating families
The members of the Cognitive Development Lab at IU, especially Lizzie Hart, Ariel La, Swapnaa Jayaraman, Caitlin Fausey, Dominique Saviano, and Mary Loftus

Our funding sources

The Cognitive Development Lab

[Logos of NSF and NIH]
<table>
<thead>
<tr>
<th>First Nouns</th>
<th>Early Nouns</th>
<th>Later Nouns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table</td>
<td>Tray</td>
<td>Shelf</td>
</tr>
<tr>
<td>Chair</td>
<td>Washing Machine</td>
<td>Bag</td>
</tr>
<tr>
<td>Shirt</td>
<td>Napkin</td>
<td>Container</td>
</tr>
<tr>
<td>Bowl</td>
<td>Knife</td>
<td>Cabinet</td>
</tr>
<tr>
<td>Bottle</td>
<td>Tissue</td>
<td>Lid</td>
</tr>
<tr>
<td>Spoon</td>
<td>Basket</td>
<td>Counter</td>
</tr>
<tr>
<td>Window</td>
<td>Sofa</td>
<td>Fireplace</td>
</tr>
<tr>
<td>Cup</td>
<td>Dryer</td>
<td>Bin</td>
</tr>
<tr>
<td>Plate</td>
<td>Bench</td>
<td>Tablecloth</td>
</tr>
<tr>
<td>Glasses</td>
<td>Can</td>
<td>Straw</td>
</tr>
<tr>
<td>Food</td>
<td>Yogurt</td>
<td>Painting</td>
</tr>
<tr>
<td>Door</td>
<td>Bucket</td>
<td>Handle</td>
</tr>
<tr>
<td>Telephone</td>
<td>Sauce</td>
<td>Seat</td>
</tr>
<tr>
<td>Couch</td>
<td>Belt</td>
<td>Wood</td>
</tr>
<tr>
<td>Picture</td>
<td>Walker</td>
<td>Outlet</td>
</tr>
</tbody>
</table>