Moving beyond “nouns in the lab”: Using naturalistic data to understand why infants’ first words include uh-oh and hi

Kennedy Casey, Christine Potter, Mira Nencheva, Casey Lew-Williams, & Erica Wojcik
How does early word learning unfold in naturalistic contexts?

How do infants map labels onto objects?
Word learning: Theories

Key predictors of AoA:
- Concreteness
- Imageability
- Frequency

Goodman et al., 2008; McDonough et al., 2011; Swingley & Humphrey, 2018
Word learning: Theories

Theories of noun learning depend on **stable visual referents**

Cross-situational mechanism:

![Diagram showing different cups]

Akhtar & Montague, 1999; Smith & Yu, 2008; Vouloumanos & Werker, 2009
Word learning: Methods

parent-report surveys

MacArthur-Bates
CDI

✓ UNDERSTANDS
✓ SAYS / GESTURES

Fenson et al., 1994

eye-tracking studies

Fernald et al., 2008

Bergelson & Swingley, 2012; Frank et al., 2016; Tincoff & Jusczyk, 1999; 2012
Evidence for a “noun bias”? 

15 most commonly-produced English words at 16 months:

Frank et al., 2016
## Earliest-produced words in 15 languages

<table>
<thead>
<tr>
<th>Language</th>
<th>Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Croatian</td>
<td>mommy, daddy, grandma</td>
</tr>
<tr>
<td>Danish (American)</td>
<td>hi, woof-woof, thank-you, bye-bye, no, yes, bye-bye, daddy, vroom</td>
</tr>
<tr>
<td>French (French)</td>
<td>mommy, daddy, baby, yes</td>
</tr>
<tr>
<td>French (Quebecois)</td>
<td>yum, grandma, vroom, bye-bye, no, no, yes, bye-bye, daddy, vroom, food</td>
</tr>
<tr>
<td>Hebrew</td>
<td>mommy, yum, bye-bye, no, thank-you, bye-bye, grandma, vroom, bye-bye, no, bye-bye, Peekaboo, moo</td>
</tr>
<tr>
<td>Italian</td>
<td>mommy, yum, bye-bye, no, cat, meow, motorcycle, baby, banana, baa-baa</td>
</tr>
<tr>
<td>Kiswahili</td>
<td>mom, yes, bye-bye,互联互通, moo, bye-bye, kitty, water, ball, doll</td>
</tr>
<tr>
<td>Croatian</td>
<td>vroom, mommy</td>
</tr>
<tr>
<td>Norwegian</td>
<td>yum, hi, bye-bye, yes</td>
</tr>
<tr>
<td>Russian</td>
<td>meow, daddy, woof-woof, grandpa, aunt, mommy, grandma</td>
</tr>
<tr>
<td>Slovak</td>
<td>mommy, daddy, woof-woof, grandma, vroom, food</td>
</tr>
<tr>
<td>Spanish (Mexican)</td>
<td>yum, woof-woof, bread, no, bye-bye, baby, cereal ball</td>
</tr>
<tr>
<td>Swedish</td>
<td>mommy, yum, thank-you, woof-woof, hi, peekaboo, meow, moo</td>
</tr>
<tr>
<td>Turkish</td>
<td>mom, yum, bye-bye, no, water, ball, doll</td>
</tr>
</tbody>
</table>

Adapted from Frank et al., 2021
Everyday words

- Do not fit into established lexical categories
- Highly frequent and early-learned
- Grounded in common routines / social interactions

See exceptions: Bergelson & Swingley, 2013; Syrnyk & Meints, 2017
Stable referents → early learning

How do everyday words fit into learning theories?
Current investigation

**Study 1:** Behavioral experiment
- *Evidence of comprehension?*

**Study 2:** Corpus-based observational research
- *Real-world input statistics?*
Study 1: Behavioral experiment

Evidence of comprehension via eye-tracking?

Standard LWL design
N = 33 infants
Age range = 10-16m
Study 1: Behavioral experiment
Evidence of comprehension via eye-tracking?

Standard LWL design
N = 33 infants
Age range = 10-16m
Study 1: Behavioral experiment

Evidence of comprehension via eye-tracking?

Standard LWL design
N = 33 infants
Age range = 10-16m

No evidence of reliable comprehension
What does uh-oh look like?
Current investigation

**Study 1:** Behavioral experiment
- *Evidence of comprehension?* (No, based on looking time)

**Study 2:** Corpus-based observational research
- *Real-world input statistics?*
Study 2: Video corpus analysis

Real-world input associated with everyday words?

Providence corpus
- 5 infants
- Age range = 11-24 months
- 114 at-home sessions (~1 hour each)
- 11,920 total tokens ($M = 993, SD = 827$)

Coding scheme
- Exact visual referent
- Situation surrounding production
- Match to experimental stimuli

Demuth et al., 2006
Study 2: Video corpus analysis

**Top-down:** Ecological validity of experimental stimuli?  
*Match vs. Non-Match*

**Bottom-up:** Characteristics of infants’ real-world input?  
*Visual vs. Situational*
Assessing the ecological validity of experimental stimuli

Visual Match

Situational Match

Non-Match

Study 1 target
(uh-oh)
Rare visual but common situational matches to stimuli
Study 2: Video corpus analysis

**Top-down:** Ecological validity of experimental stimuli?  
Match vs. Non-Match

**Bottom-up:** Characteristics of the real-world input?  
Visual vs. Situational
Visual stability?

Co-occurrence with consistent visual referents?
Everyday words are variable at the **visual level**

**Prototypical visual referent?**

![Graph showing the proportion of tokens against rank-ordered visual referents. The x-axis represents rank-ordered visual referents, and the y-axis represents the proportion of tokens. The graph shows a bell-shaped curve, indicating variability in visual referents.]

**Hypothetical visual input**

- target child falling
- microphone falling off
- crayon breaking
- cup falling
- child hiding
- phone ringing
- bubble popping
Everyday words are variable at the **visual level**

**Prototypical visual referent?**

![Graph showing the proportion of tokens for rank-ordered visual referents.](image)

- Target child falling
- Microphone falling off
- Crayon breaking
- Cup falling
- Child hiding
- Phone ringing
- Bubble popping

**uh-oh**
Everyday words are variable at the visual level

Prototypical visual referent?

- Co-occurred with hundreds of unique visual referents:
  \[ M = 343 \text{ unique referents} \]
  \[ \text{range} = 34 - 1,414 \]

- Appeared with unique visual referent for 1 in 3 tokens:
  \[ M = 34.5\% \text{ unique referents} \]
  \[ \text{range} = 19.0 - 45.6\% \]
Everyday words are variable at the visual level
Everyday words are variable at the **visual level**

**everyday words**

9%

- uh-oh
- ?

**concrete nouns**

85-92%

- cup

Bergelson & Swingley, 2013; Custode & Tamis-LeMonda, 2020
Everyday words are variable at the visual level

Referents vary within and across children
Study 2: Video corpus analysis

**Top-down:** Ecological validity of experimental stimuli?  
Match vs. Non-Match

**Bottom-up:** Characteristics of the real-world input?  
Visual vs. Situational
Situational stability?

Consistency in broader context surrounding production?
Everyday words are more stable at the situational level.
Discussion

- **Study 1**: Standard lab-based measures failed to show evidence of everyday word comprehension.

- **Study 2**: Naturalistic investigation found that everyday words do not co-occur with consistent visual referents but more reliably appear in stable situational contexts.

- Current theories/methods over-prioritize visual information.

- Visual cues matter, but what else?
Using naturalistic data to refine theories and methods

- **Past ecological work:**
  - Multimodal cues (e.g., Abu-Zhaya et al., 2017)
  - Contextual/spatial cues (e.g., Roy et al., 2015)

- **New questions:**
  - Frequency of occurrence in isolation? (e.g., Brent & Siskind, 2001; Lew-Williams et al., 2011)
  - Consistency of prosodic information? (e.g., Nencheva et al., 2021)
  - Frequency of occurrence at event boundaries? (e.g., Sonne et al., 2017)
  - Contingency on infant behavior? (e.g., Tamis-LeMonda et al., 2014)
  - Link to social reward? (e.g., Gros-Louis et al., 2014)
Using naturalistic data to refine theories and methods

- **Past ecological work:**
  - Multimodal cues (e.g., Abu-Zhaya et al., 2017)
  - Contextual/spatial cues (e.g., Roy et al., 2015)

- **New questions:**
  - Frequency of occurrence in isolation? (e.g., Brent & Siskind, 2001; Lew-Williams et al., 2011)
  - Consistency of prosodic information? (e.g., Nencheva et al., 2021)
  - Frequency of occurrence at event boundaries? (e.g., Sonne et al., 2017)
  - Contingency on infant behavior? (e.g., Tamis-LeMonda et al., 2014)
  - Link to social reward? (e.g., Gros-Louis et al., 2014)
Everyday words occur frequently in isolation

**everyday words** vs. **concrete nouns**

*frequency- and AoA-matched*
Isolation may be helpful for several reasons

- Clearly segmentable word boundaries (e.g., Lew-Williams et al., 2011)
- More consistent prosodic contours
Using naturalistic data to refine theories and methods

Lack of visual stability

Some situational stability

- **A way forward**: Naturalistic video corpora, including headcam data
  (e.g., Bergelson et al., 2019; Sullivan et al., 2021)
Acknowledgments

Funding
NICHD (F32 HD093139, R01HD095912)

Dr. Christine E. Potter
University of Texas at El Paso

Mira Nencheva
Princeton University

Dr. Casey Lew-Williams
Princeton University

Dr. Erica H. Wojcik
Skidmore College

Participating families

Ellie Breitfeld
Matthew Weatherhead