What’s the Relevant Data for Toddlers’ Word Learning?

Sumarga H. Suanda, Linda B. Smith, & Chen Yu

INTRODUCTION

Toddlers’ learning environment is a data hodgepodge. Parts of the data are “high quality” (i.e., referentially transparent), other parts are “low quality” (i.e., referentially ambiguous). Knowing what parts of the data are relevant for learning is central to understanding the mechanisms that underlie early word learning.

Research Goal

In this research, we investigated the nature of the data that contribute to toddlers’ learning and examined the implications of this data for the mechanisms that underlie early word learning.

METHOD OVERVIEW

1. We observed two-year-olds, equipped with head cameras, and their parents playing with novel objects.

2. We then tested toddlers’ knowledge of word-object pairings.
   - Toddlers were tested twice for each word.
   - Words were classified as “learned” if they chose correctly on both trials.

3. We analyzed the quality of naming events.
   - We computed the number of pixels in toddlers’ views taken up by each object.
   - Naming event quality was quantified as the proportion of all object pixels that belonged to the named object.
   - As illustrated below, naming events varied widely in their quality.

4. We tested whether an associative model could account for learning and whether all the data, or just a subset, were needed to do so.

RESULTS

1. Naming event quality from toddlers’ perspectives
   - We computed the number of pixels in toddlers’ views taken up by each object.
   - Naming event quality was quantified as the proportion of all object pixels that belonged to the named object.
   - As illustrated below, naming events varied widely in their quality.

2. Predicting learning from naming event data
   - For each toddler, we built co-occurrence matrices that either integrated naming event quality or did not.
   - From the populated matrices of each toddler, we computed the associative strengths between words and their referents (p(oi|wj)).

3. Are toddlers’ using just the “good” data?
   - The highest quality naming events were neither necessary nor sufficient to account for toddlers’ learning patterns.
   - When a substantial amount of naming events were incorporated, we were able to account for toddlers’ learning patterns, even when those events were lower quality.

4. For predicting learning, does it matter which data are sampled as long as enough data are sampled?
   - We tested whether we could account for learning by simply incorporating enough data, irrespective of quality (i.e., by randomly choosing which data were included).

ACKNOWLEDGEMENTS

We thank many members of the Computational Cognition and Learning Lab, and the Cognitive Development Lab for assistance with this work.

This research was supported in part by the National Science Foundation (BCS 092428, SMA 1203420) and the National Institutes of Health (ROI-HD07665, R01-EY017843, K99-HD082558).

Illustrations of child with oranges and of market were retrieved in May 2016.

For more information, please contact: Sumarga (Imra) Suanda (sumanda@indiana.edu).