Effects of Home Environment & Semantic Structure on Early Lexical Development

Shannon Dailey, Andrei Amatuni and Elika Bergelson | Department of Psychology & Neuroscience, Duke University

BACKGROUND

While infants' vocabularies grow remarkably fast (Bergelson & Swingley, 2012; Dale & Fenson, 1996; Fernald et al, 1998), it is unclear what *causes* improved comprehension with age.

We explore whether advances in early word comprehension seem better explained by:

- *changes in the input data* (i.e. parents talking more or differently to older infants), or by
- *changes in the learner* (i.e. cognitive growth and linguistic exposure).

Here we contrast these explanations combining in-lab and at-home measures in a yearlong longitudinal sample.

We use ***early-learned, concrete nouns*** (e.g. ball, shoe, car) as a test case.

Research Questions:

1) How does ***word comprehension*** change with age?

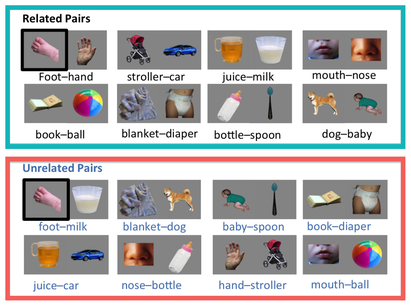
2) Does ***at-home exposure*** to the tested words change month-to-month?

METHODS

- Analysis from 44 infants, 6-18 months (SEEDLingS Corpus)

In-lab word comprehension (eye-tracking)

- At 6, 12, and 18 months longitudinally
- Probed infants' word comprehension using eye-tracking used to quantify infants' fixations to named object on a visual display
- Tested on 16 **semantically related** and **semantically unrelated** word pairs:



Home linguistic environment

- Monthly audio and video recordings (n=12 per child) at home for each infant

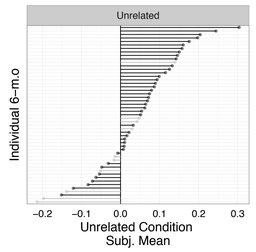
Child-directed object words are annotated, along with 3 properties of each:

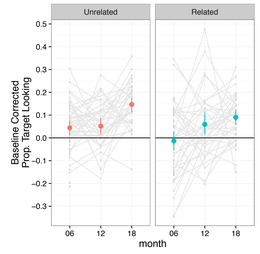
- type of utterance (e.g. command, question, singing)
- object presence (is the object present and attended to?)
- speaker

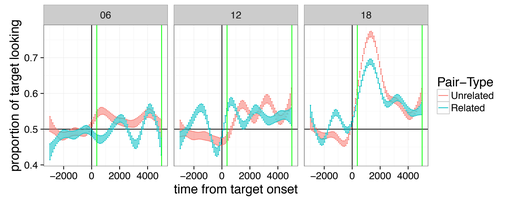
Measure	Time
Daylong home audio recording	Monthly, 6-17 mos
Hour-long home video recording	Monthly, 6-17 mos
In-lab word comprehension assessment	6, 12, 18 mos

RESULTS

In-lab word comprehension

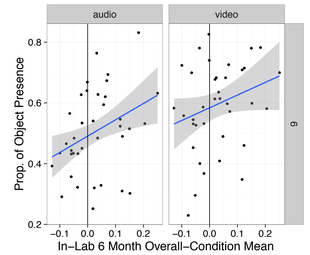






- Infants at all three ages were able to discriminate **unrelated** nouns, but only 12- and 18-month olds were successful in the **related** condition.
- Performance across timepoints improved but was not correlated.

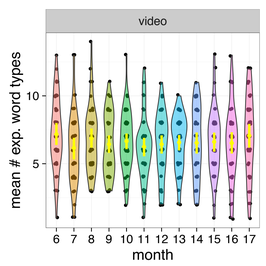
Object-label co-occurrence & comprehension

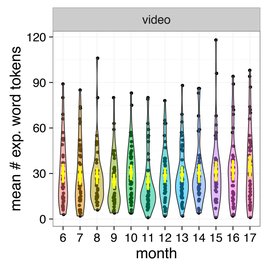


Home environment correlates with in-lab performance at 6mos.

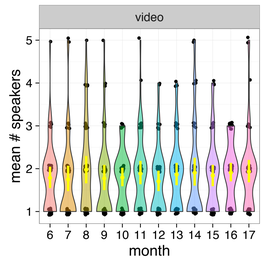
Home linguistic environment

The tested words are **very common** and **very consistent** in the infants' input.

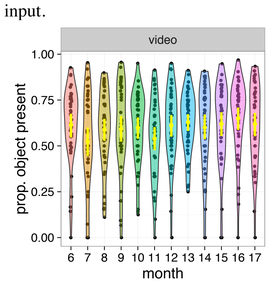




Input quantity (types and tokens) stable month to month.



# of speakers stable month to month.



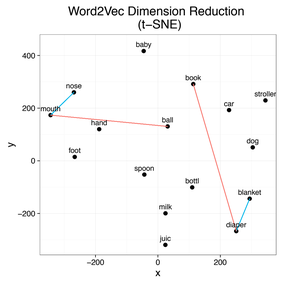
Object-label co-occurrence stable month to month.

Measuring semantic similarity

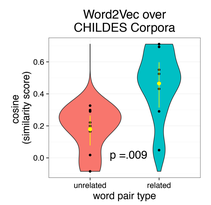
To measure if the word pairs we categorized as **related** or **unrelated** are quantifiably so, we calculated the cosine similarity between words using semantic vector space models.

Models were trained on the CHILDES corpus of early environmental linguistic input rather than adult language corpora, e.g. Wikipedia pre-trained vectors.

We verify that the related pairs are more semantically similar than the unrelated pairs in input to children.



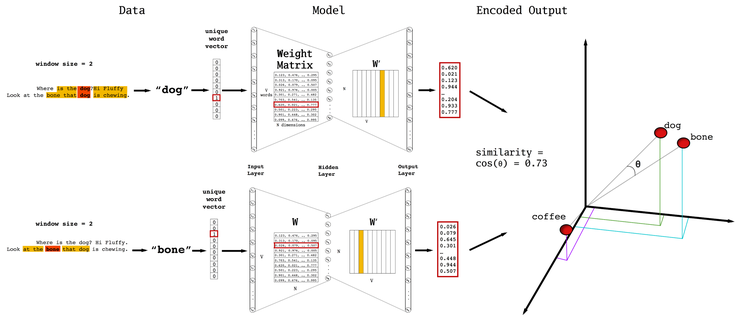
$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A}\cdot\mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}},$$



Word2vec encodes words as vectors (shallow neural network model).

Networks predict the most likely **context** words which could surround an input word. The vector representations are taken from the weight matrices between layers.

We can treat these vectorized words as points in an N dimensional space and compute semantic distance.



DISCUSSION

1) How does word comprehension change with age?

- Older infants understand the tested words better than younger infants.
- Understanding words in semantically-related visual contexts is more challenging.

2) Does at-home exposure to the tested words change month-to-month?

- First pass, no. Input for these words is incredibly stable month-to-month.

Conclusions

- No support for a *changing data account*.
- Results are compatible with two flavors of a *changing learner account*:
  - "*More data*" *account*: more input = greater learning (accumulator function)
  - "*Better learner*" *account*: linguistic, cognitive, and social gains allow older infants to take better advantage of the same kind of data.
- Models of semantic space trained on child-directed speech corpora converge with real data.

Ongoing & future directions

- Unique vs. general instances of words/categories
- Links to early production
- Links to non-linguistic development (e.g. pointing)

References

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (3111-3119). Chicago.

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences, 109*(9), 3253-3258.

Dale, P. S., & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments, & Computers, 28*(1), 125-127.

Fernald, A., Pinto, J. P., Swingley, D., Weinbergy, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science, 9*(3), 228-231.