InWordDB

A comprehensive database of infant word segmentation

Christina Bergmann & Alejandrina Cristia

LSCP, CNRS, ENS-DEC, EHESS, Paris, France

Background: Infant word segmentation

Typical segmentation study[1]:

- Expose infant to words
- Test with passages (or vice versa)
  Test contains familiar (known) and novel (unknown) words
- Dependent measure:
  Listening time (LT) to both types of test trials

Significant difference between familiar and novel:
↬Words were segmented!
NB: Direction varies! (Longer LT either to novel or familiar)

Research Question

Can we identify factors that influence the outcome of a segmentation study?

What role does **age** play?
↬ Switch from preference for **familiar** test stimuli to preference for **novel** test stimuli[2,3]
(prediction based on infant information processing)

Impact of **language background**?
Emerging evidence (French, Spanish)[4,5,6]

Database Information

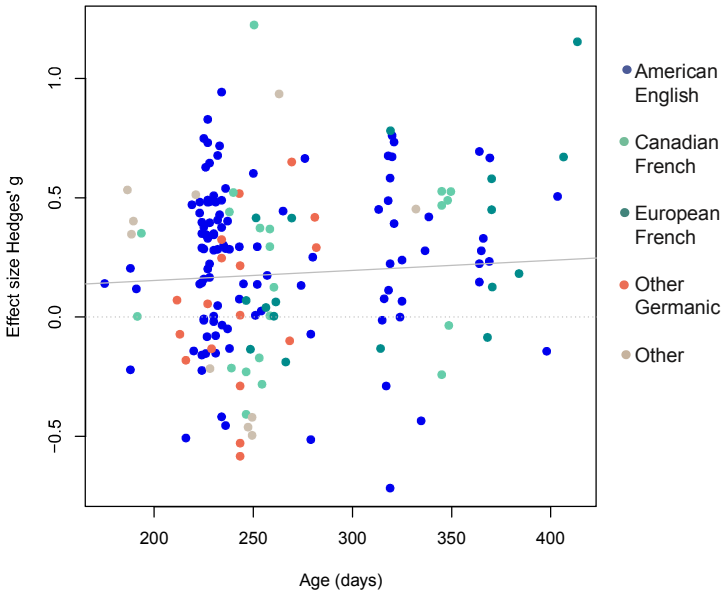
Study ID	Participants	Stimuli	Procedure	Results
Author(s)	Age in days	Language	N test trials	LT Familiar / Novel
Title	Number	Words	Fam criterion	SD for both LTs
Year	% Girls	Paragraphs	Method	% Familiarity pref.
DOI or link	% Drop-out	Edge align		**Effect size**

51 studies encoded (mostly published articles) including 175 unique experimental conditions including data from 3951 infants

Age: 5¾ - 14 months (mean: 8½ months)

Results

Mean weighted ES positive: Hedge's g = 0.2 (SE = 0.003)
Significantly above 0 with $p < 0.001$



Open Science

Access: ***sites.google.com/site/InWordDB***

- Database publicly available (with analysis scripts, plots, explanations, ...)
- Continuously updated

Join us!

↬ Add to it and **submit data** (published or not)
- Null results are also of interest!

↬ **Coordinate extensions** of the database:
- Older children
- Artificial speech input (Statistical learning)
- Word-object mapping
- ...

Effect Sizes for repeated measures

"Ingredients" to compute Effect Size (**ES**) for rm
1. Infant behavior per condition
2. Standard deviation of 1.
3. Correlation between conditions
4. Number of infants tested

↬ These depend on reporting standards in a field!
(NB: correlation usually not reported)

ES for repeated measures: **Formula** (Hedge's g)

$$\left(1 - \frac{3}{4N_{Participants} - 5}\right) \times \frac{LT_{Familiar} - LT_{Novel}}{\sqrt{SD^2_{Familiar} + SD^2_{Novel}/2}}$$

Correct for small sample sizes — Effect Size

ES is weighted by the correlations (3.)
↬ How "systematic" was infant behavior during test?

References

1. Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology, 29*(1), 1-23.
2. Houston-Price, C., & Nakai, S. (2004). Distinguishing novelty and familiarity effects in infant preference procedures. *Infant and Child Development, 13*(4), 341-348.
3. Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. *Advances in Infancy Research, 5*, 69-95.
4. Polka, L., & Sundara, M. (2012). Word Segmentation in Monolingual Infants Acquiring Canadian English and Canadian French: Native Language, Cross-Dialect, and Cross-Language Comparisons. *Infancy, 17*(2), 198-232.
5. Nazzi, T., Mersad, K., Sundara, M., Iakimova, G., & Polka, L. (2014). Early word segmentation in infants acquiring Parisian French: task-dependent and dialect-specific aspects. *Journal of Child Language, 41*(3), 600-633.
6. Bosch, L., Figueras, M., Teixidó, M., & Ramon-Casas, M. (2013). Rapid gains in segmenting fluent speech when words match the rhythmic unit: evidence from infants acquiring syllable-timed languages. *Frontiers in Psychology, 4*:106.

Contact: InWorDB@gmail.com