

# Factors affecting lexical development by English speaking children

Junko Maekawa and Holly L. Storkel, University of Kansas

## Introduction

- ▶ Young children learn new words rapidly and effortlessly.
  - ▷ Fast mapping
- ▶ Phonological variables affecting lexical development:
  - ▷ Phonotactic probability – likelihood of occurrence of a phoneme sequence
    - High: phoneme sequences that are common in a language (e.g., “coat”)
    - Low: phoneme sequences that are rare in a language (e.g., “watch”)
  - ▷ Neighborhood density – the number of words (i.e., neighbors) that differ by one phoneme from a target word
    - Dense: A word has many neighbors (e.g., “sit” – 36 neighbors)
    - Sparse: A word has few neighbors (e.g., “these” – 9 neighbors)
  - ▷ Word frequency: number of occurrence of a word in a language
  - ▷ Word length: number of phonemes in the transcription
  - ▷ High probability words from dense neighborhoods are acquired more rapidly than low probability from sparse neighborhoods (Storkel, 2001, 2004; Storkel & Maekawa, in press).
  - ▷ Shorter words that occur frequently in a language are acquired at earlier ages than longer words that occur infrequently (Storkel, 2004).
- ▶ Questions
  - ▷ Do phonotactic probability, density, word frequency, and/or word length predict the age when words are first produced spontaneously by young children?
  - ▷ Are there similarities and differences across children in the patterns of significant predictors?

## Method

- ▶ Conversational samples from CHILDES
  - ▷ April: 22-35 months; 6 data points; 340 different root nouns
  - ▷ Peter: 21-37 months; 20 data points; 767 different root nouns
  - ▷ Allison: 16-34 months; 6 data points; 174 different root nouns
- ▶ Variables
  - ▷ Predictor variables
    - Phonotactic probability (Segment average, Biphone average)
    - Density
    - Frequency (Kucera and Francis, 1967)
    - Length
  - ▷ Outcome variable
    - Age of first production
      - Earliest age when the child spontaneously produced the word in the sample

## Analyses

- ▶ Primary analysis: backward regression to determine significant predictors
- ▶ Follow-up analyses: regression to determine predictors for each type of word for children with multiple significant predictors (i.e., Peter, Allison)

## Results

- ▶ April:
  - ▷ Length:  $t = 4.22, p < 0.001, r^2 < 0.07$ 
    - Short words were produced at earlier ages than long words
- ▶ Peter
  - ▷ Density:  $t = 2.11, p = 0.04, r^2 < 0.01$ 
    - Words from sparse neighborhoods were produced at earlier ages than words from dense neighborhoods, contrary to the previous studies.
  - ▷ Frequency:  $t = -3.06, p = 0.002, r^2 < 0.03$ 
    - High frequency words were produced at earlier ages than low frequency words
  - ▷ Length:  $t = 2.43, p = 0.02, r^2 < 0.02$ 
    - Short words were produced at earlier ages than long words

Table 1. Results of the follow-up regression analyses by word types for Peter

		$\beta$ estimate	Standard Error	$r^2$	$t$	$p$
Sparse words	*Frequency	-0.13	0.06	0.03	-2.33	0.02*
	Length	0.10	0.05	0.02	1.82	0.07
Dense words	*Frequency	-0.15	0.08	0.04	-1.96	0.05*
	*Length	0.20	0.08	0.05	2.44	0.02*
Low frequency words	Density	0.11	0.07	0.01	1.64	0.10
	Length	0.06	0.06	0.01	0.98	0.33
High frequency words	Density	0.09	0.07	0.01	1.20	0.23
	*Length	0.23	0.07	0.05	3.30	0.001*
Short words	*Density	0.14	0.06	0.01	2.14	0.03*
	*Frequency	-0.21	0.05	0.04	-3.87	<0.001*
Long words	Density	0.07	0.08	<0.01	0.87	0.39
	Frequency	-0.01	0.08	<0.01	-0.09	0.93

- ▶ Allison
  - ▷ Segment average (Phonotactic probability):  $t = -2.02, p = 0.05, r^2 < 0.03$ 
    - High probability words were produced at earlier ages than low probability words.
  - ▷ Length:  $t = 2.89, p = 0.005, r^2 < 0.05$ 
    - Short words were produced at earlier ages than long words.

Table 2. Results of the follow-up regression analyses by word types for Allison

		$\beta$ estimate	Standard Error	$r^2$	$t$	$p$
Low probability words	Length	0.15	0.08	0.05	1.84	0.07
	*Length	0.28	0.09	0.09	3.11	0.002*
Short words	Segment average	-0.12	0.07	0.04	-1.69	0.10
	Segment average	-0.04	0.08	<0.01	-0.55	0.59

## Summary

Table 3. Summary of predictors for each child (with early acquired word types noted in parentheses) and of differences across children

	April	Peter	Allison
Overall analysis	Length (short)	Density (sparse)	Segment average (high)
		Frequency (high)	Length (short)
		Length (short)	
Follow-up analyses			
Low probability	NA	NA	No significant predictors
High probability	NA	NA	Length (short)
Sparse	NA	Frequency (high)	NA
Dense	NA	Frequency (high)	NA
		Length (short)	
Low frequency	NA	No significant predictors	NA
High frequency	NA	Length (short)	NA
Short	NA	Density (sparse)	No significant predictors
		Frequency (high)	
Long	NA	No significant predictors	No significant predictors

## Discussion

- ▶ The predictors of age of first production varied across children and across word types
- ▶ Possible accounts for the observed variability
  - ▷ Lexical variability
    - Differences in size of the lexicon and frequency of words for each child did not predict the variability.
  - ▷ Use and availability of cues
    - Emergentist Coalition Model (Hirsh-Pasek, Golinkoff, & Hollich, 2000)
      - A range of cues are available for a child to learn words.
      - Availability of these cues varies over time.
      - Word length may be one of the earliest cues used to learn words.

## Conclusion and future investigation

- ▶ Theories of lexical development need to be able to account for individual differences in the predictors of word learning across children, including other populations such as speakers of English as a second language, bilingual speakers, or monolingual Japanese children.

## Reference

- Hirsh-Pasek, K., Golinkoff, R. M., and Hollich, G. (2000). An emergentist coalition model for word learning: Mapping words to objects is a product of the interaction of multiple cues. In *Becoming a word learner*. New York, NY: Oxford University Press.
- Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University.
- Storkel, H. L. (2001). Learning new words: Phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research*, *44*, 1321-1337.
- Storkel, H. L. (2004). Do children acquire dense neighborhoods? An investigation of similarity neighborhoods in lexical acquisition. *Applied Psycholinguistics*, *25*, 201-221.
- Storkel, H. L. & Maekawa, J. (in press). A comparison of homonym and novel word learning: The role of phonotactic probability and word frequency. *Journal of Child Language*.

**For more information please contact:**  
[www.ku.edu/~wrdlrng/](http://www.ku.edu/~wrdlrng/)  
[junko@ku.edu](mailto:junko@ku.edu)

