

# Phonological and Lexical Influences on Expressive Vocabulary: Predicting the Ages at Which Individual Words are Spoken

Lisa Ryckbost & Tim Brackenbury  
Bowling Green State University

Presented at the 2006 Symposium  
on Research in Child Language  
Disorders. Madison, WI

## Abstract

Previous research has shown that phonological and lexical characteristics of individual words have an impact on their acquisition and retrieval. This study examines the potential influences of four characteristics on the ages at which individual words are produced by infants and toddlers. Normative data from the expressive vocabulary items on the MacArthur-Bates CDI were examined for age of acquisition (AoA), word length, phonotactic probability, word frequency, and neighborhood density. Three linear regressions revealed separate equations for predicting AoA. Two of the equations were based on only one significant contributor (neighborhood density and word length). All of the equations were equally accurate at predicting AoA.

## Purpose & Questions

- To replicate and expand on Storkel (2004a) by including phonotactic probability in the analysis.
- Does phonotactic probability have an influence on the early word production of infants? If so, what is the influence?
- As a group, how do the four characteristics influence early word production?
- Can a mathematical formula (based on the characteristics) be developed to accurately predict the AoA of specific words?

## Methods

- Created a spreadsheet that included
  - 139 nouns from the MacArthur-Bates CDI
    - equally distributed across AoA's
  - Their AoA's
    - the earliest age at which  $\geq 50\%$  of children from the CDI normative pool produced the word
  - Numeric values for each of the four characteristics

## Correlations

	Word Length	Segment Summary <sup>a</sup>	Biphone Summary <sup>a</sup>	Word Frequency	Neighborhood Density
AoA	.338**	.258**	.255**	-.110	-.355**
Word Length		.772**	.641**	-.272**	-.693**
Segment Summary <sup>a</sup>			.812**	-.187*	-.405**
Biphone Summary <sup>a</sup>				-.079	-.340**
Word Frequency					.195*

<sup>a</sup> measures of phonotactic probability

\*\*Correlations significant at the 0.05 level (2-tailed)

\*Correlations significant at the 0.01 level (2-tailed)

## Phonological & Lexical Characteristic Investigated

- Word length:** a count of the number of phonemes in a word
- Phonotactic probability:** the likelihood that a sound or sound sequence within an individual word occurs across the words in the language
- Word frequency:** a measure of how often a given word occurs in the language in daily conversational speech
- Neighborhood density:** the number of words that differ from a given word by one phoneme substitution, deletion, or addition

Sample Word	Word length	Phonotactic Probability	Word frequency*	Neighborhood Density
Cat	3 phonemes: /k/, /æ/, /t/	High (.2381)	131 occurrences	35 neighbors: (e.g., at, kite, scat)
Vision	5 phonemes: /v/, /ɪ/, /ʒ/, /e/, /n/	Low (.2029)	0 occurrences	2 neighbors: (version, villain)

\*based on 286,108 word tokens reported by Moe, Hopkins, & Rush (1982)

## Findings from Previous Child Studies

- Storkel and Rogers (2000) found a significant influence of phonotactic probability on word learning in 10-13 year old children. High probability nonsense words were correctly mapped more often than low probability words.
- Storkel (2001) examined the influence of phonotactic probability on 3-6 year old children. Novel words with high phonotactic probability were acquired faster than those with low probability.
- Storkel (2004a) examined the influence of word length, frequency, and neighborhood density on the vocabulary development of infants and toddlers through analysis of the MacArthur-Bates CDI.
  - Linear regression revealed significant effects of a) word frequency, b) neighborhood density, and c) interaction between neighborhood density and word length on AoA.
  - Shorter words with higher density and frequency values were acquired earlier.
- Hollich, Jusczyk, and Luce (2000) found 17 month old infants learned low density words faster than high density words.

## Linear Regressions & Predictive Equations

Original data, all characteristics in equation

- Overall significant ( $p < 0.001$ ,  $r^2 = 11.80$ )
  - $AoA = 22.6 + (0.242 * \text{word length}) - (1.05 * \text{segment summary}) + (48.0 * \text{biphone summary}) - (0.00105 * \text{frequency}) - (0.124 * \text{neighborhood density})$

Original data, stepwise

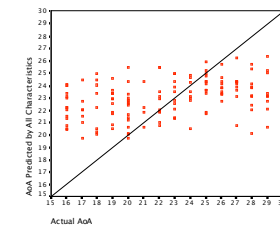
- Overall significant ( $p < 0.001$ ,  $r^2 = 12.61$ )
  - Neighborhood density** was the only significant variable
  - $AoA = 24.345 - (0.175 * \text{the word's density})$

Transformed data, stepwise

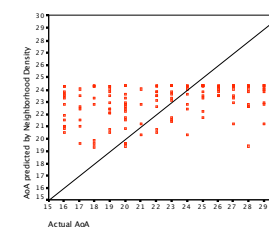
- Transformations done to reduce the relatedness between the variables (Storkel 2004b)
  - Overall significant ( $p < 0.001$ ,  $r^2 = 11.44$ )
    - Word length** was the only significant variable
    - $AoA = 19.21 + (0.88 * \text{the word's length})$

## Scatterplots from the Predictive Equations

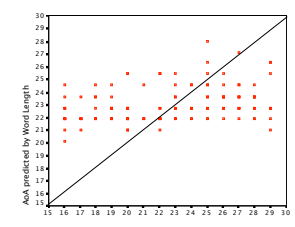
All Characteristics Equation



Neighborhood Density Equation



Word Length Equation



Each scatterplot is based on the 139 words analyzed.

## Validity of Predictive Equations

- Each equation is restricted in its predictive range
  - neighborhood density equation predicts at or below 24 months
  - word length equation predicts at or above 19 months
- Correlations with actual AoA's
  - All characteristics equation  $r = .310$   $p = 0.006^*$
  - Neighborhood density equation  $r = .327$   $p = 0.009^*$
  - Word length equation  $r = .208$   $p = 0.084$
- No significant differences between the three equation's predictive accuracies ( $p \geq 0.439$ )

## Discussion

- Influence of phonotactic probability on AoA
  - Not a significant contributor to any of the regression models
  - Strongly correlated with word length and neighborhood density
  - Its influence on word learning appears to occur at later stages of development
- Interactions of phonological and lexical characteristics on AoA
  - Neighborhood density and word length were the only significant predictors
  - Similar to the results of Storkel (2004a)
- Predictive Equations
  - The all characteristics and the neighborhood density equations significantly correlated with actual AoA.
    - $\sim 10\%$  of shared variance between actual and predicted AoA's for both
  - Neighborhood density appears to be the best equation
    - based on parsimony and predictive accuracy

# Phonological and Lexical Influences on Expressive Vocabulary: Predicting the Ages at Which Individual Words are Spoken

(Page 2)

Presented at the 2006 Symposium  
on Research in Child Language  
Disorders. Madison, WI

## References

- Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., Pethick, S., & Reilly, J.S. (1994). *The MacArthur Communicative Development Inventories: User's guide and technical manual*. San Diego, CA: Singular Publishing.
- Hollich G., Jusczyk, P. W., & Luce, P. A. (2002). Lexical neighborhood effects in 17-month-old word learning. In B. Skarabela, S. Fish, & A. H.-J. Do (Eds.), *Proceedings of the 26<sup>th</sup> Annual Boston University Conference on Language Development* (Vol.1). Somerville, MA: Cascadilla Press.
- Moe, A. J., Hopkins, K. J., & Rush, R. T. (1982). *The Vocabulary of First Grade Children*. Springfield, IL: Thomas.
- Storkel, H. (2004a). Do children acquire dense neighborhoods? An investigation of similarity neighborhoods in lexical acquisition. *Applied Psycholinguistics*, 25, 201-221.
- Storkel, H. (2004b). Methods for minimizing the confounding effects of word length in the analysis of phonotactic probability and neighborhood density. *Journal of Speech, Language, and Hearing Research*, 47, 1454-1468.
- Storkel, H. & Rogers, M. (2000). The effect of probabilistic phonotactics on lexical acquisition. *Clinical Linguistics & Phonetics*, 14, 407-425.