

Clustering of words in young children's vocabularies: Comparison of English and Spanish



Mariza Rosales and Holly L. Storkel, Ph.D., CCC-SLP
University of Kansas



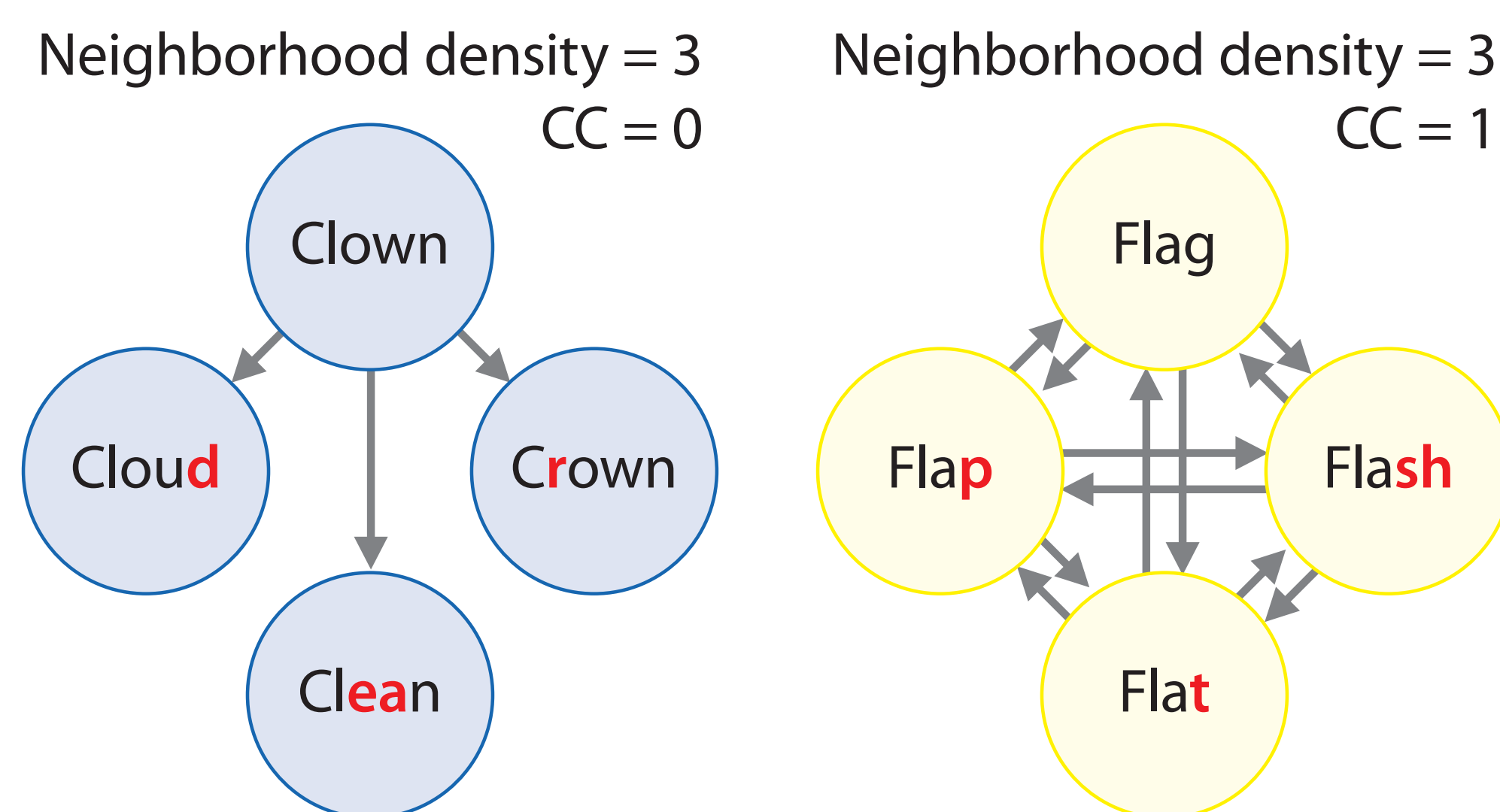
Introduction

Neighborhood density is the number of words that sound similar to a given word

Neighbors of the word sit				
pit	sat	sick	sits	_it

In both English and Spanish, neighborhood density affects word learning

- Words with many neighbors learned earlier than words with few neighbors (Storkel 2004)
- It is easier to learn and remember words in English and Spanish that are added to large neighborhoods than to small neighborhoods.
- Specifically, existing neighbors reinforce the representation of new words.
- However, density does not capture the interconnectedness of the neighbors. It's merely the size of the neighborhood.
- The clustering coefficient (CC) does $CC = \frac{\text{number of triangles in the network}}{\text{number of possible triangles}}$. Adult neighborhoods are highly clustered (i.e., have high CC, Vitevitch, 2006)



Purpose

To determine whether clustering is observed in young children's vocabulary

Hypothesis 1:

More interconnections among neighbors may further enhance learning and memory

Hypothesis 2:

This effect will be the same across English and Spanish, as observed for neighborhood density

Questions

- When does clustering emerge in acquisition?
- Does the emergence of clustering vary across languages?

Methods

Data taken from the MacArthur Communicative Development Inventories database of English-speaking toddlers and Spanish-speaking toddlers between the ages of 16-30 months. (<http://130.191.235.141>)

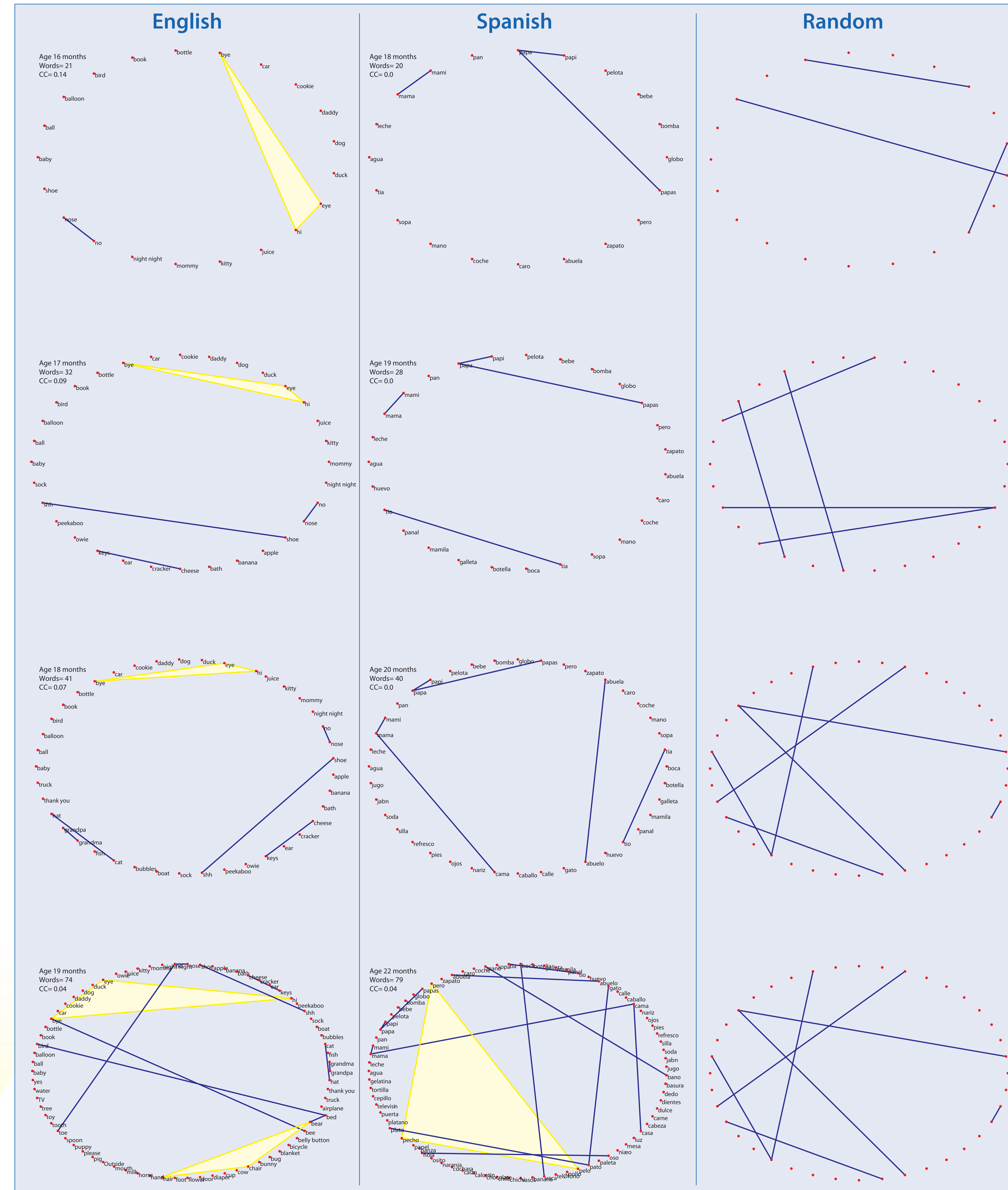
Parents indicated which words their toddler knew. The age-of-acquisition of each word was defined as the earliest age when 50% or more of the children tested reportedly knew the word.

English vocabularies at 16, 17, 18, and 19 months were analyzed. Spanish vocabularies were matched in size to the English with 18, 19, 20, and 22 months being analyzed.

Networks of the vocabularies were generated and CC was computed using the program PAJEK

Similar sized random networks also were generated using PAJEK for comparison

Results



Summary

- English vocabularies show clustering at the earliest point examined (16 months)
- Spanish vocabularies at 18, 19, and 20 months show no clustering (similar to random)
- At 22 months, Spanish vocabularies show evidence of clustering (differ from random)

Conclusions

- Clustering was observed in both languages which leads to the possibility that it might influence word learning
- Future study needed to examine learning of words with High CC vs. Low CC
- Age of emergence of clustering varies across languages (Spanish & English)
- May be related to difference in word length across Spanish and English
- Emergence of clustering in language may be dependent on the structure of the language
- Future study needed to examine impact of emergence of clustering on vocabulary growth in individual children. Clustering could be an important developmental milestone.

Acknowledgements

NIH R25 GM62232, NIH K23 DC04781, NIH R01 DC08095

Undergraduate Research Award from the University of Kansas Honors Program

Dr. Holly Storkel,
Dr. Mike Vitevitch,
Dr. Marian Hukle, and
The Word and Sound Learning Lab

