Distributional learning & lexical category acquisition:

What makes words easy to categorize?

Giovanni Cassani, Robert Grimm, Steven Gillis, Walter Daelemans

28th July 2017 – CogSci 2017

CLiPS
Computational Linguistics & Psycholinguistics
University of Antwerp

Introduction

Words can be categorized into lexical categories using the contexts in which they occur [Harris, 1954], but some words more easily than others.

What distributional properties of a word makes it easier to categorize?

CAVEAT

The statistical analysis has been refined since the submission. We present new results, that are sometimes different than what appears in the paper.

Differences are **highlighted.**

Got evidence it works

We know there is information in distributional co-occurrences that supports learning of lexical categories [Cartwright & Brent, 1997; Redington & al 1998].

Behavioral experiments have confirmed that children use co-occurrences to group words along syntactic dimensions [Frost & al, 2016; Mintz & al, 2014; Reeder & al, 2013; van Heugten & Johnson, 2010; Zhang & al, 2014].

Contrasted contexts

- **Frequent Frames:** *you_X_the*
  [Mintz, 2003]
- **Flexible Frames:** *you_X + X_the*
  [St. Clair & al, 2010]
- **Bigrams vs trigrams:** *you_X vs you_X_the*
  [Monaghan & al, 2004]
- **Utterance boundaries:** *the_X vs the_X_#end*
  [Freudenthal & al, 2008]

Evaluated learning mechanisms

- Incremental Bayesian clustering [Parisien, 2008]
- Incremental Entropy-based clustering [Chrupała & Alishahi, 2010]
- MOSAIC [Freudenthal & al, 2016]

The evaluation concerns whether **good categories** are learned and whether learning follows aspects of the **developmental pattern.**

A concept of *easiness*

Children categorize certain words better than others.

- What causes words to be categorized better?
- Are words that are easier to categorize using distributional information also those that children categorize better?

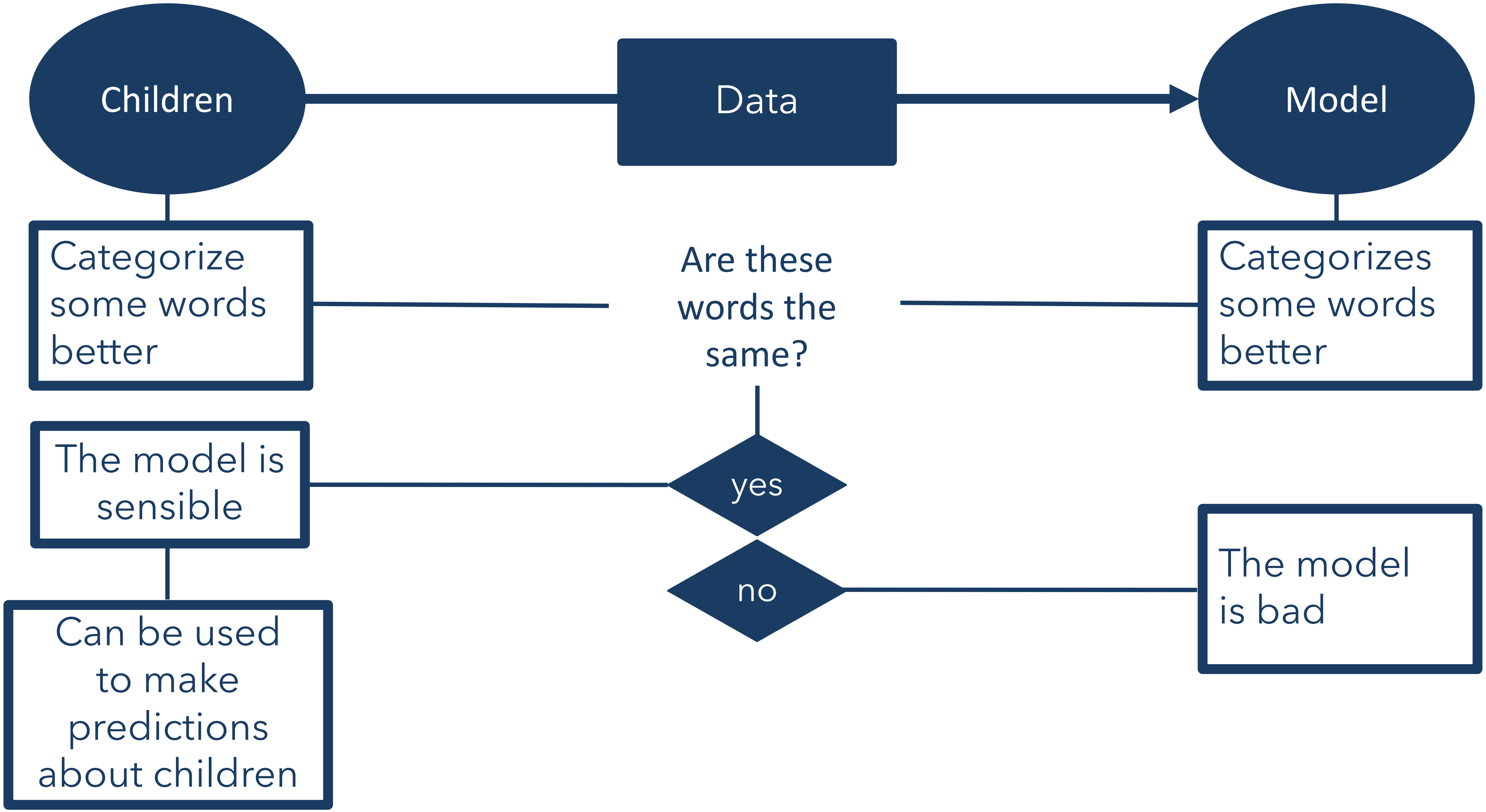
Many potential predictors

Frequency is not enough in accounting for lexical category acquisition [Matthews & Bannard, 2010].

Diversity, **predictability**, and **entropy** are other pieces of distributional information that children can track and might contribute to explain easiness.

Computational simulation

The logic of the experiment



Experimental setting

- **Unsupervised** PoS tagging experiment (5 tags)
- Transcribed English Child-directed speech (13 individual corpora)
- Bigrams and trigrams (with utterance boundaries) as contexts [*b_X*; *X_c*; *a_b_X*; *b_X_c*; *X_c_d*]
- Exemplar-based clustering (TiMBL: IB1, cosine, 1 NN, **no feature weighting**)
- Incremental training (40 to 70% of the input corpus)

Statistical analysis

- Logistic mixed-effects models (**with crossed-effects**):
  - Random intercepts for each corpus and word **+ random slopes for both random effects**
- Predictors included based on **reduction in AIC**
- **Interactions** between each predictor and time were tested and included if they improved the fit

Predictors and outcomes

- *Token frequency*
- *Diversity*
- *Average conditional probability*
- *Entropy (normalized)**
- *Time*
- ***PoS tag of the word***

❑ **Hits**

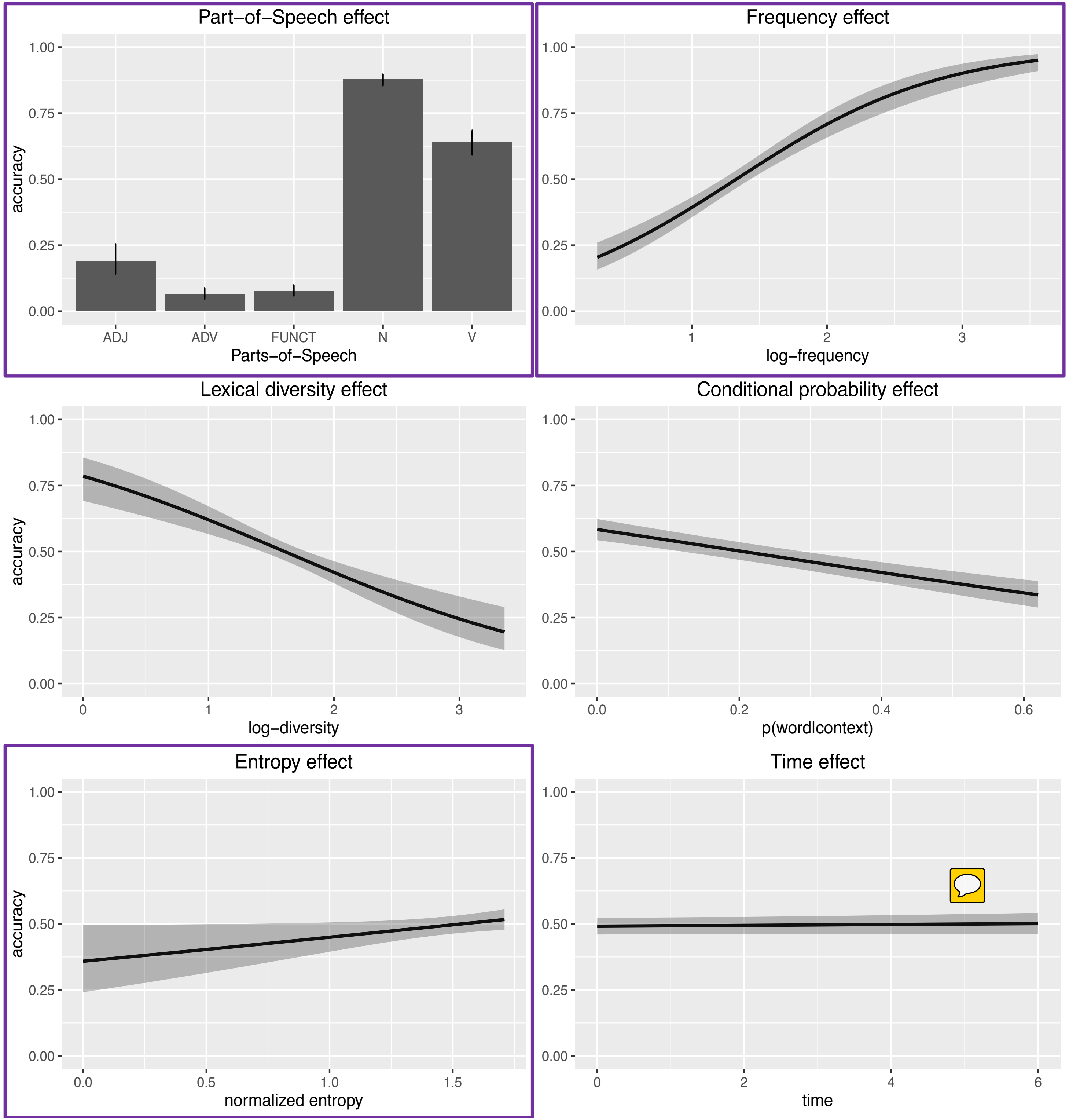
**exponentiated*

Operationalization

We ran the clustering experiment, finding the nearest neighbor in the training set for target words in the test set. Categorization accuracy was used as a dependent variable to assess ***easiness***.

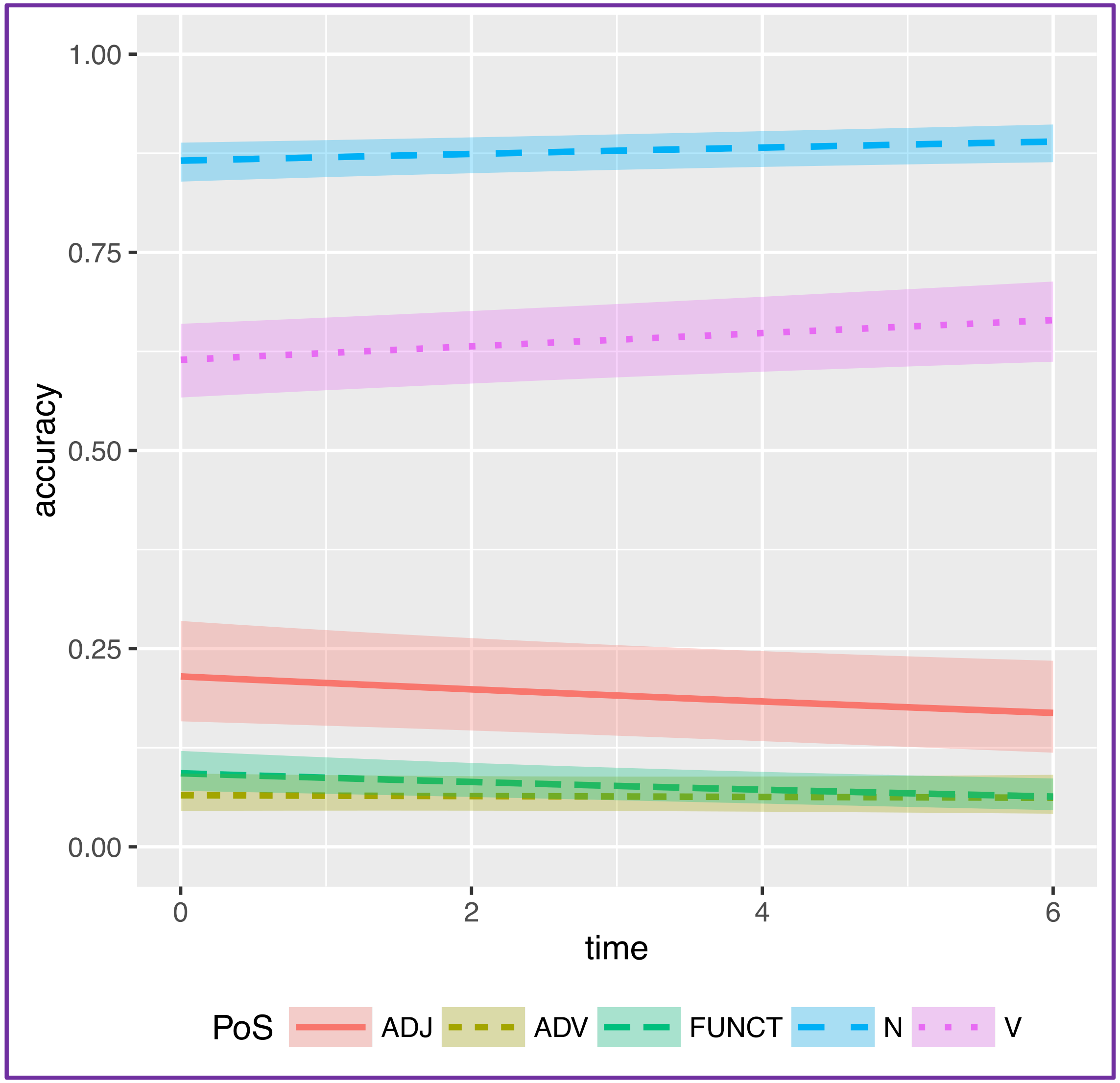
Results

Main effects



CLiPS
Computational Linguistics & Psycholinguistics
University of Antwerp

Interactions



Beware of the noise

Words are easier to categorize when **highly specific**:

- > are frequent
- > occur with fewer contexts
- > are hard to predict given the contexts in which they occur
- > are nouns or verbs

apple; forget; table; door; …

Next?

Quantify **how useful a context** is to categorization and assess which distributional properties affect it

Test **other learning mechanisms** than Memory Based Learning (e.g. neural nets, Bayesian inference)

Extend this approach **beyond distributional properties**

Thank you!

Questions?

References

- Harris, Z. (1954). Distributional structure. Word, 10(2-3), 146-152
- Maratsos, M. P., & Chalkley, M. A. (1980). The Internal Language of Children Syntax: The nature and ontogenesis of syntactic categories. In K. E. Nelson (Ed.), Children's language (Vol. 2, pp. 127-213). New York, NY: Gardner Press.
- Pinker, S. (1987). The bootstrapping problem in language acquisition. In B. MacWhinney (Ed.), Mechanisms of language acquisition. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Gleitman, L. R., & Gillette, J. (1996). The role of syntax in verb-learning. In P. Fletcher & B. MacWhinney (Eds.), The Handbook Of Child Language (pp. 413-427). Oxford, UK: Blackwell.
- Cartwright, T. A., & Brent, M. R. (1997). Syntactic categorization in early language acquisition: Formalizing the role of distributional analysis. Cognition, 63(2), 121-170.
- Christophe, A., Guasti, T., Nespor, M., Dupoux, E., & Van Ooyen, B. (1997). Reflections on phonological bootstrapping: Its role for lexical and syntactic acquisition. Language and Cognitive Processes, 12(5-6), 585-612.
- Redington, M., Chater, N., & Finch, S. (1998). Distributional Information: A Powerful Cue for Acquiring Syntactic Categories. Cognitive Science, 22(4), 425-469.
- Christophe, A., Millotte, S., Bernal, S., & Lidz, J. (2008). Bootstrapping Lexical and Syntactic Acquisition. Language and Speech, 51(1-2), 61-75.
- Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. Cognition, 90(1), 91-117.
- Monaghan, P., & Christiansen, M. H. (2004). What distributional information is useful and usable for language acquisition. In D. G. Kenneth Forbus, Terry Regier (Ed.), Proceedings of the 26th annual conference of the Cognitive Science Society. Austin, TX: Cognitive Science Society.
- Freudenthal, D., Pine, J. M., & Gobet, F. (2008). On the Utility of Conjoint and Compositional Frames and Utterance Boundaries as Predictors of Word Categories. In V. Sloutsky, B. Love, & K. McRae (Eds.), Proceedings of the 30th Annual Conference of the Cognitive Science Society (pp. 1947-1952). Austin, TX: Cognitive Science Society.
- Parisien, C. (2008). An Incremental bayesian Model for Learning Syntactic Categories. In A. Clark & K. Toutanova (Eds.), 12th Conference on Computational Natural Language Learning (CoNLL) (pp. 89-96). Stroudsburg, PA: Association for Computational Linguistics.

References

- Chrupała, G., & Alishahi, A. (2010). Online Entropy-based Model of Lexical category Acquisition. In M. Lapata & A. Sarkar (Eds.), Proceedings of the 14th Conference on Computational Natural Language Learning (CoNLL) (pp. 182-191). Stroudsburg, PA: Association for Computational Linguistics.
- Matthews, D., & Bannard, C. (2010). Children's Production of Unfamiliar Word Sequences Is Predicted by Positional Variability and Latent Classes in a Large Sample of Child-Directed Speech. Cognitive Science, 34(3), 465-488.
- St. Clair, M. C., Monaghan, P., & Christiansen, M. H. (2010). Learning grammatical categories from distributional cues: Flexible frames for language acquisition. Cognition, 116(3), 341-360.
- van Heugten, M., & Johnson, E. K. (2010). Linking infants' distributional learning abilities to natural language acquisition. Journal of Memory and Language, 63(2), 197-209.
- Reeder, P. A., Newport, E. L., & Aslin, R. N. (2013). From shared contexts to syntactic categories: The role of distributional information in learning linguistic form-classes. Cognitive Psychology, 66(1), 30-54.
- Mintz, T. H., Wang, F. H., & Li, J. (2014). Word categorization from distributional information: Frames confer more than the sum of their (Bigram) parts. Cognitive Psychology, 75, 1-27.
- Zhang, Z., Shi, R., & Li, A. (2014). Grammatical Categorization in Mandarin-Chinese-Learning Infants. Language Acquisition, 22(1), 104-115.
- Frost, R. L. A., Monaghan, P., & Christiansen, M. H. (2016). Using statistics to learn words and grammatical categories: How high frequency words assist language acquisition. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), Proceedings of the 38th Annual Conference of the Cognitive Science Society (pp. 81). Austin, TX: Cognitive Science Society.
- Freudenthal, D., Pine, J. M., Jones, G., & Gobet, F. (2016). Developmentally plausible learning of word categories from distributional statistics. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), Proceedings of the 38th Annual Conference of the Cognitive Science Society (pp. 674). Austin, TX: Cognitive Science Society.