

Distributional nuggets for lexical categories: Finding the useful information in co-occurrence patterns

Giovanni Cassani, Robert Grimm, Steven Gillis, Walter Daelemans
name.surname@uantwerpen.be

What makes a distributional context useful?

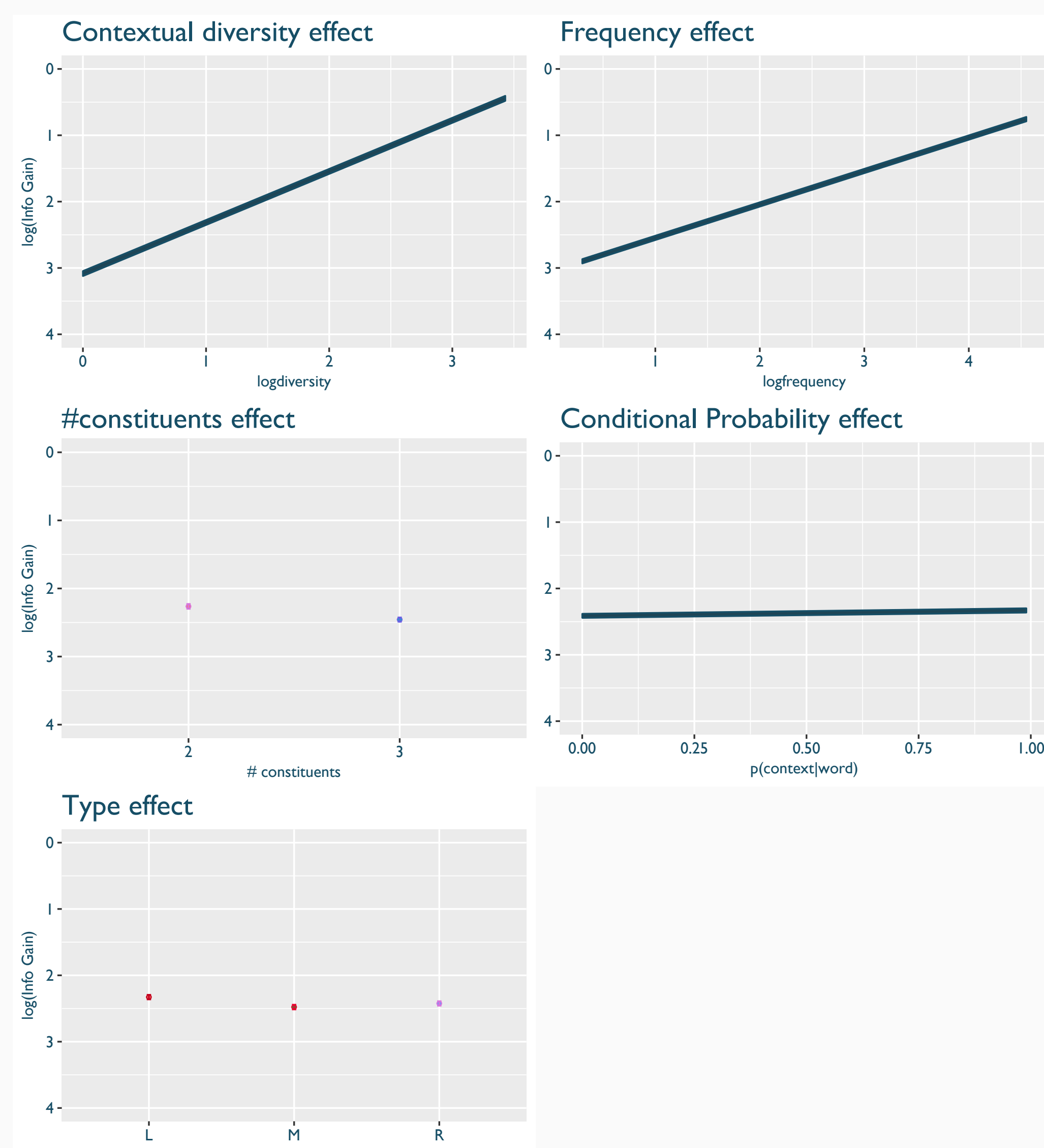
Design

13 corpora of child-directed speech
Longitudinal design
Linear Mixed Models
Usefulness $\simeq \log(\text{Information Gain})$
Five predictors: Frequency, Lexical diversity,
Conditional probability of contexts given
words, Number of constituents, Type of
context

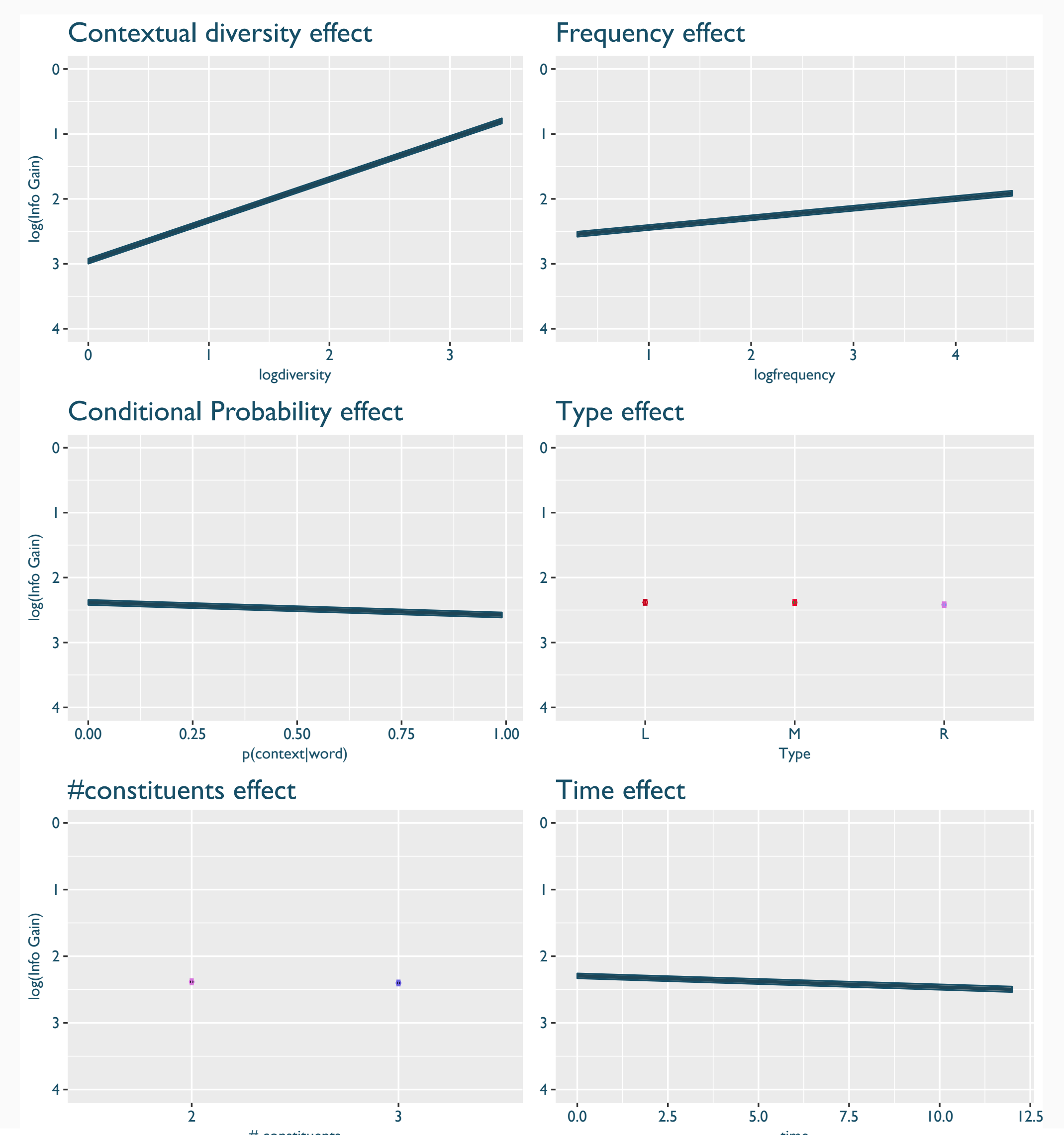
Useful contexts are diverse, frequent,
slightly unpredictable

HYPOTHESIS: if a model selects useful
contexts considering diversity and
conditional probability on top of frequency,
it should discover more useful contexts and
achieve better categorization

One predictor at a time

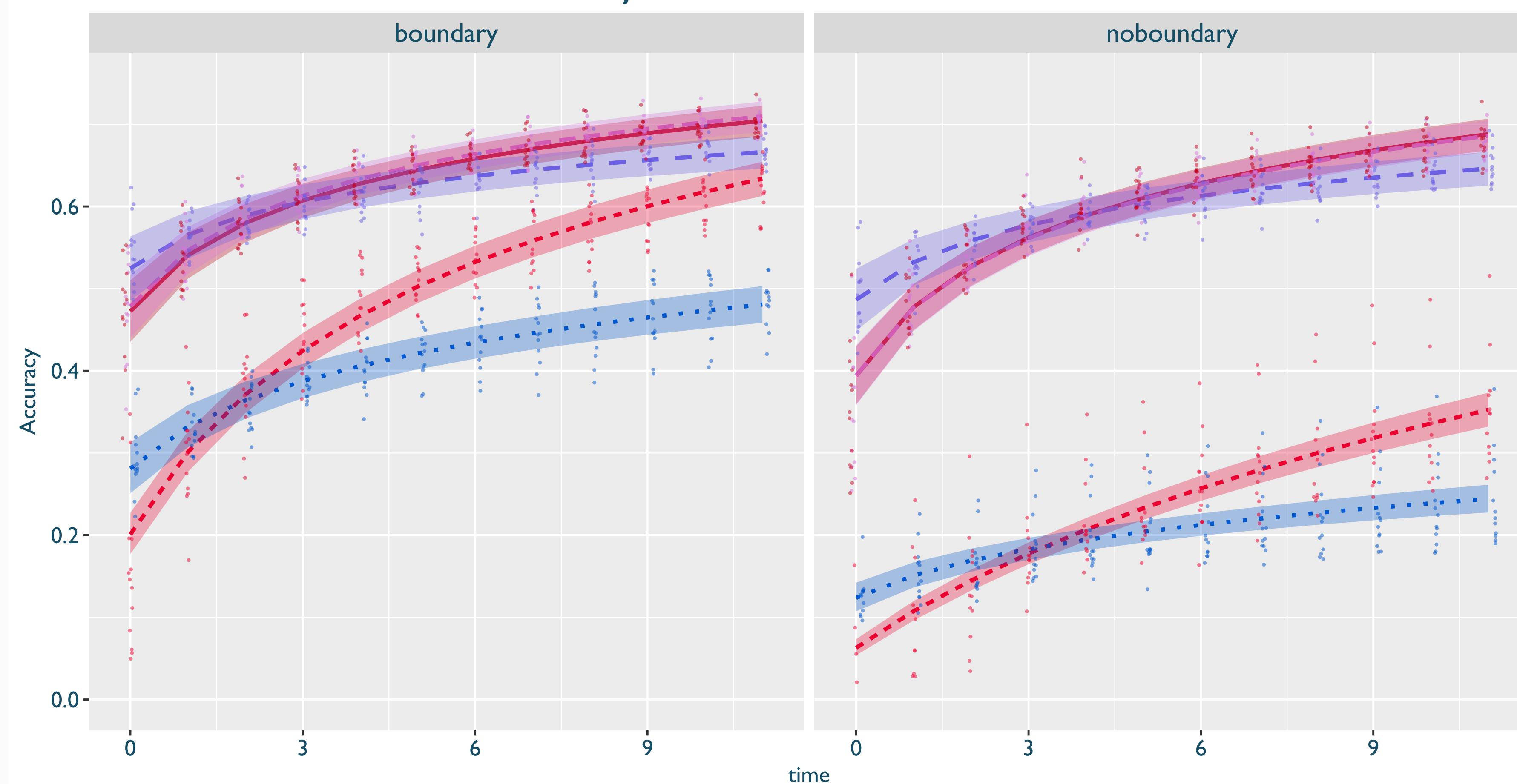


All predictors at once

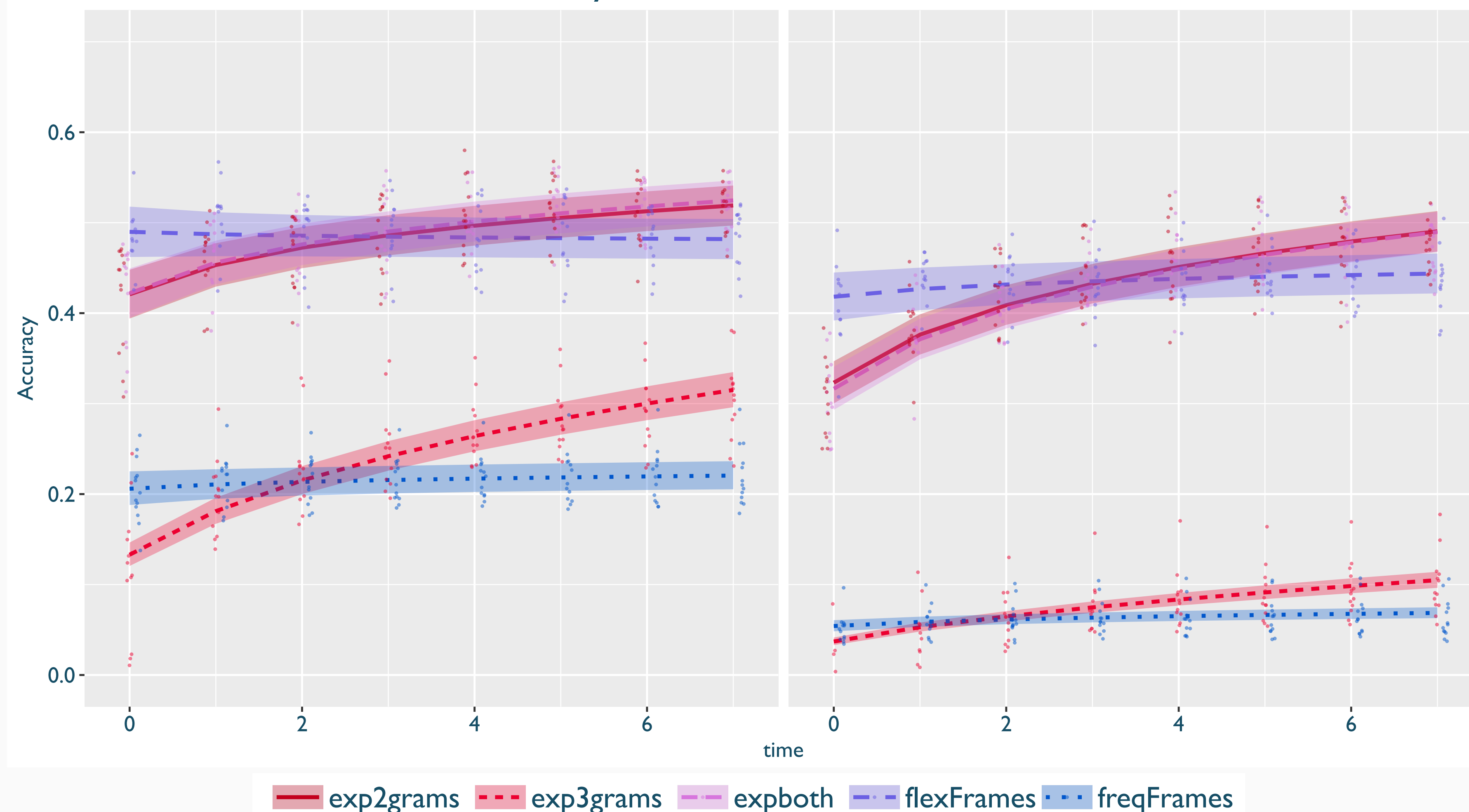


Two more allies: Diversity and predictability help learning lexical categories

Core words: classification accuracy



Rare words: classification accuracy



Learning: Usefulness = $\log(F) * \log(D) * \text{predictability}$

Data: 13 longitudinal corpora of child-directed speech

Evaluation: Classification accuracy using unsupervised KNN ($k = 1$) and cosine as distance metric

Contrasts: the experimental model in 3 context conditions (2grams, 3grams, both) and two frequency-based baselines (frequent frames and flexible frames), both evaluated over time considering/not considering utterance boundaries as context elements

Statistics: Logistic mixed models on classification accuracy with time, boundary condition, and model as predictors

Top figure: Core vocabulary

Target words are words that appear in all 13 corpora
Longitudinal leave-one-out classification: target words are used to categorize other target words

Bottom figure: Generalization

Training items are words that occur in all 13 corpora
Test items are all other words
Training-Testing design: words from training are used to categorize words from the test set. with no overlap between them

CONCLUSIONS

Bigrams are more effective than trigrams, especially when generalizing to new, rare words

The experimental 2gram model outperforms flexible frames, even though they can rely on the exact same type of contexts: *how useful contexts are selected is critical!* Diversity and predictability help lexical category acquisition

Acknowledgements

This research was supported by a BOF/TOP grant (ID 29072) of the Research Council of the University of Antwerp. The poster was designed on Overleaf using the baposter template by Brian Amberg.