From lexical categories to acceptable lexical choices:
A discriminative learning approach

Giovanni Cassani[1], Robert Grimm[1], Harald Baayen[2], Walter Daelemans[1], Steven Gillis[1]
[1]*name.surname*@uantwerpen.be, [2] harald.baayen@uni-tuebingen.de

1. Why and how

Is it possible to learn lexical categories without explicitly trying to learn them?

We trained a single layer perceptron using the Rescorla-Wagner update rule to discriminate lexical items given phonological cues and tested whether it was possible to learn lexical categories

- from phonological cues only
- from distributional similarities derived from phonological cues

in two training regimes

- on *isolated words*, where distributional similarities only depend on word similarities
- on full utterances, where distributional similarities can rely on context

HYPOTHESIS: phonological cues will provide some information - especially with small vocabulary, but distributional similarities will provide more, and scale to larger vocabulary. This will only happen when training in context and after some exposure to the input, while sound sequences will be usable from the earlier on.

4. Conclusions & future work

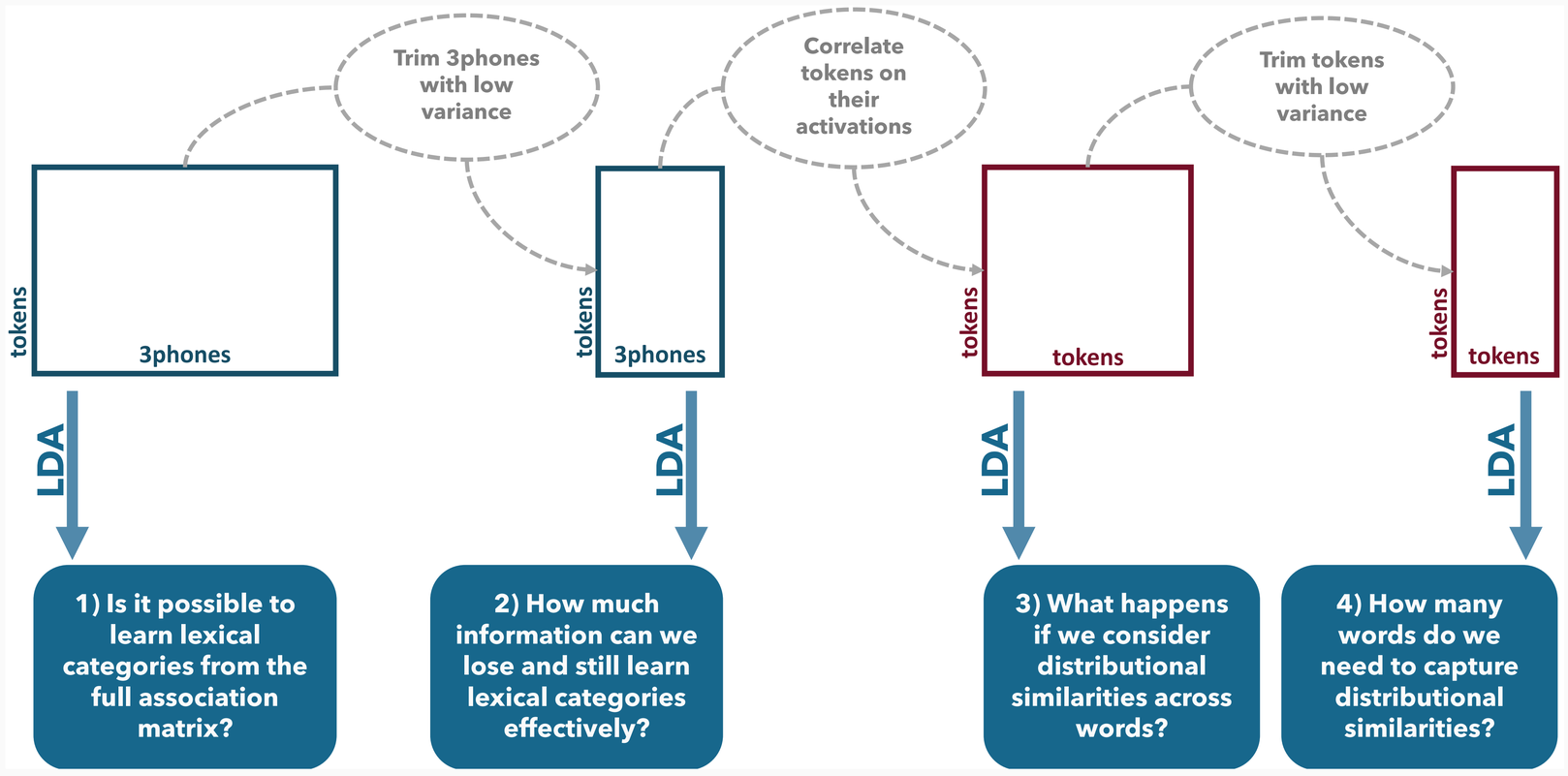
There's information supporting lexical categories in symbolic phonological cues. There's **much more information** supporting lexical categories **in distributional similarities** derived from symbolic phonological cues.

Phonological information becomes less useful when the vocabulary increases. Distributional information stays useful **even with a larger vocabulary**.

Crucially, distributional information is **only useful in context** and takes longer to become reliable - showing a U-shape curve.

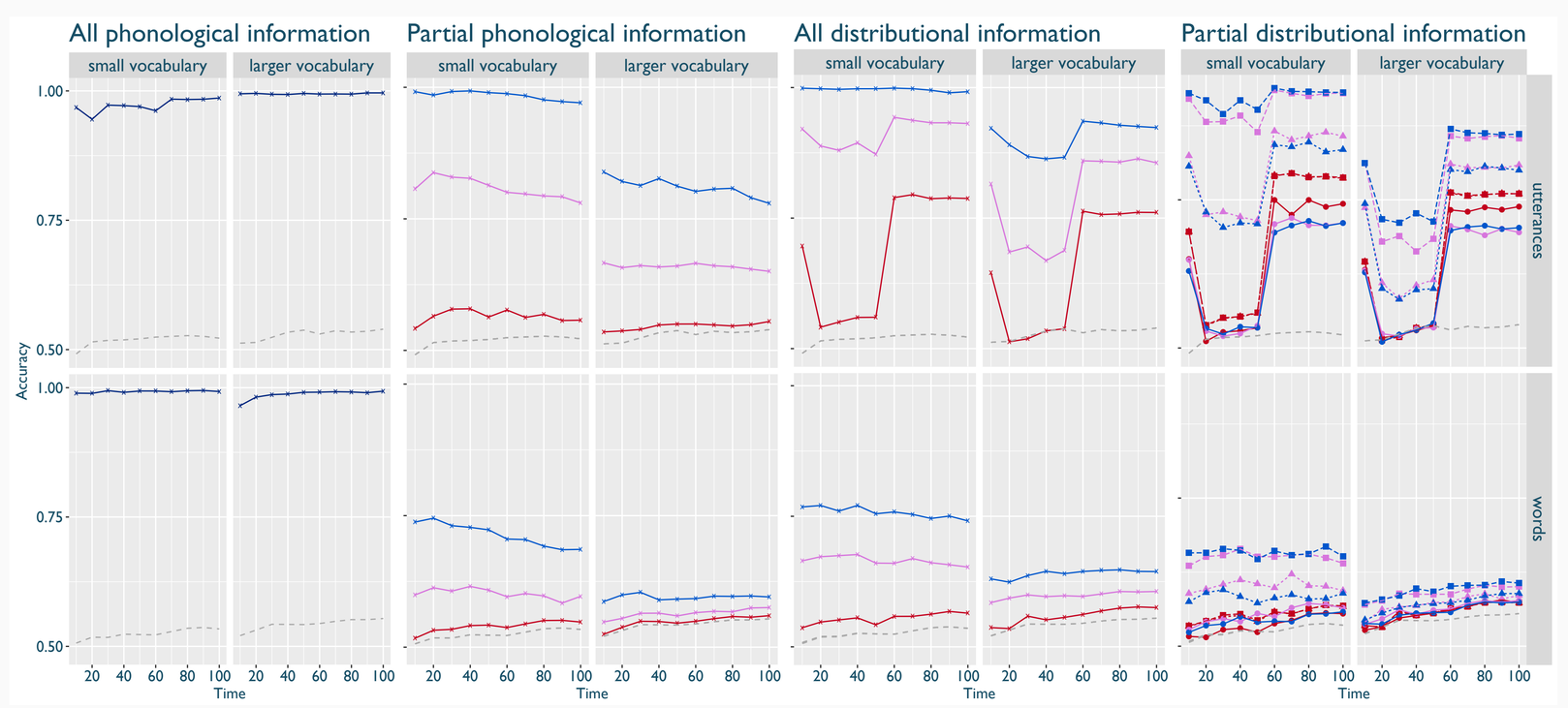
Lexical categories are not present during training, but the model still learns information that supports them: **lexical categories are side effects of language learning**.

2. Logical flow

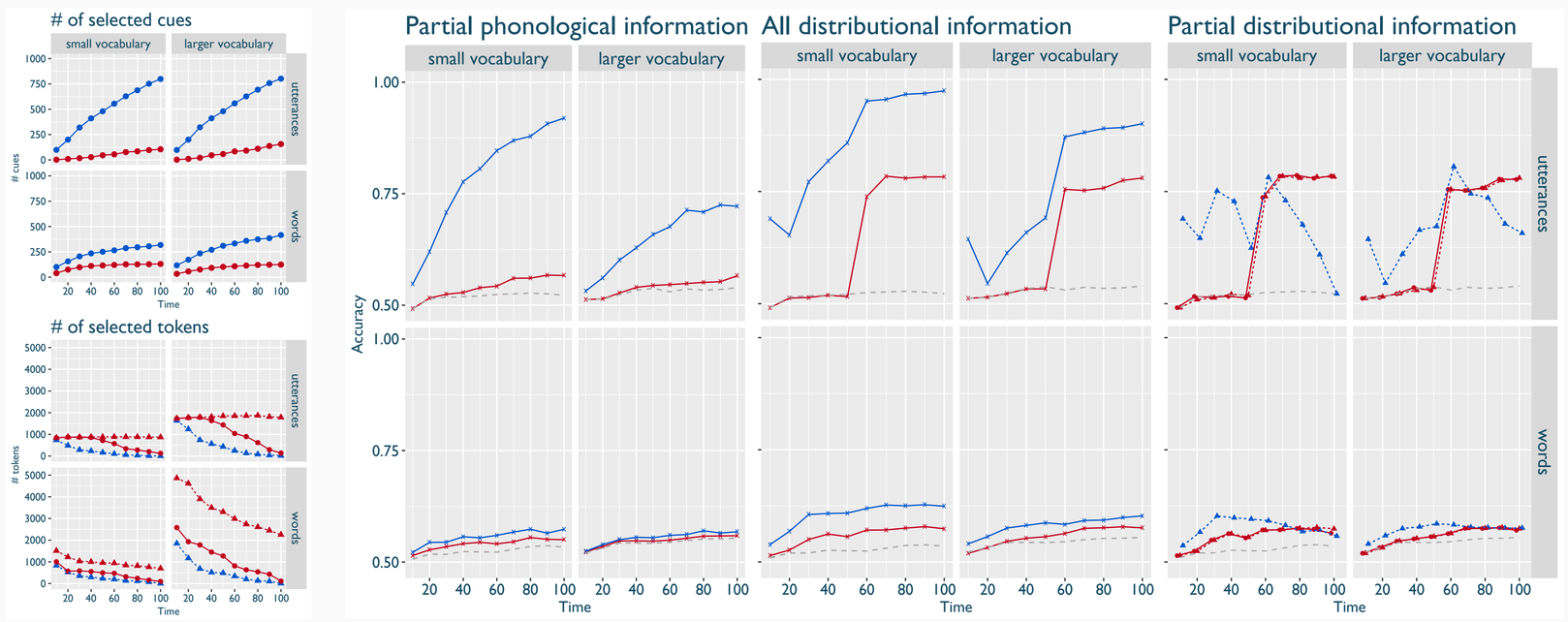


3. Results

Set the dimensionality of the reduced phonological and distributional spaces



Set a threshold on the variance a phonological cue or a token need to have



Acknowledgements

This research was supported by a **BOF/TOP grant (ID 29072) of the Research Council of the University of Antwerp**. We thank Konstantin Sering for the invaluable discussions about how to design the experiments and Stéphan Tulkens for the help in writing the code. The poster was designed on Overleaf using the baposter template by Brian Amberg.

CLiPS
Computational Linguistics & Psycholinguistics
University of Antwerp