

The bootstrapping toolbox:

Which cues are more useful to learn
lexical categories and why

What's it about?

The analysis of what information in the linguistic input may support lexical category acquisition

Main contributions

- Item-level analysis of distributional and phonological cue usefulness
- Exploratory study of the relation between sound and meaning and its role in lexical category acquisition
- Comparison of two different classes of learning models
- Proof of concept that categories need not be the goal of learning for categorically organized representations to emerge

The wampimuk jumps and smirgles.



CLIPS

Computational Linguistics & Psycholinguistics
University of Antwerp

Giovanni Cassani

16/05/2019 Antwerp



Let's break it up

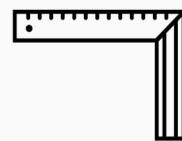


CLIPS

Computational Linguistics & Psycholinguistics
University of Antwerp

Giovanni Cassani

16/05/2019 Antwerp



The neighbors



The sound

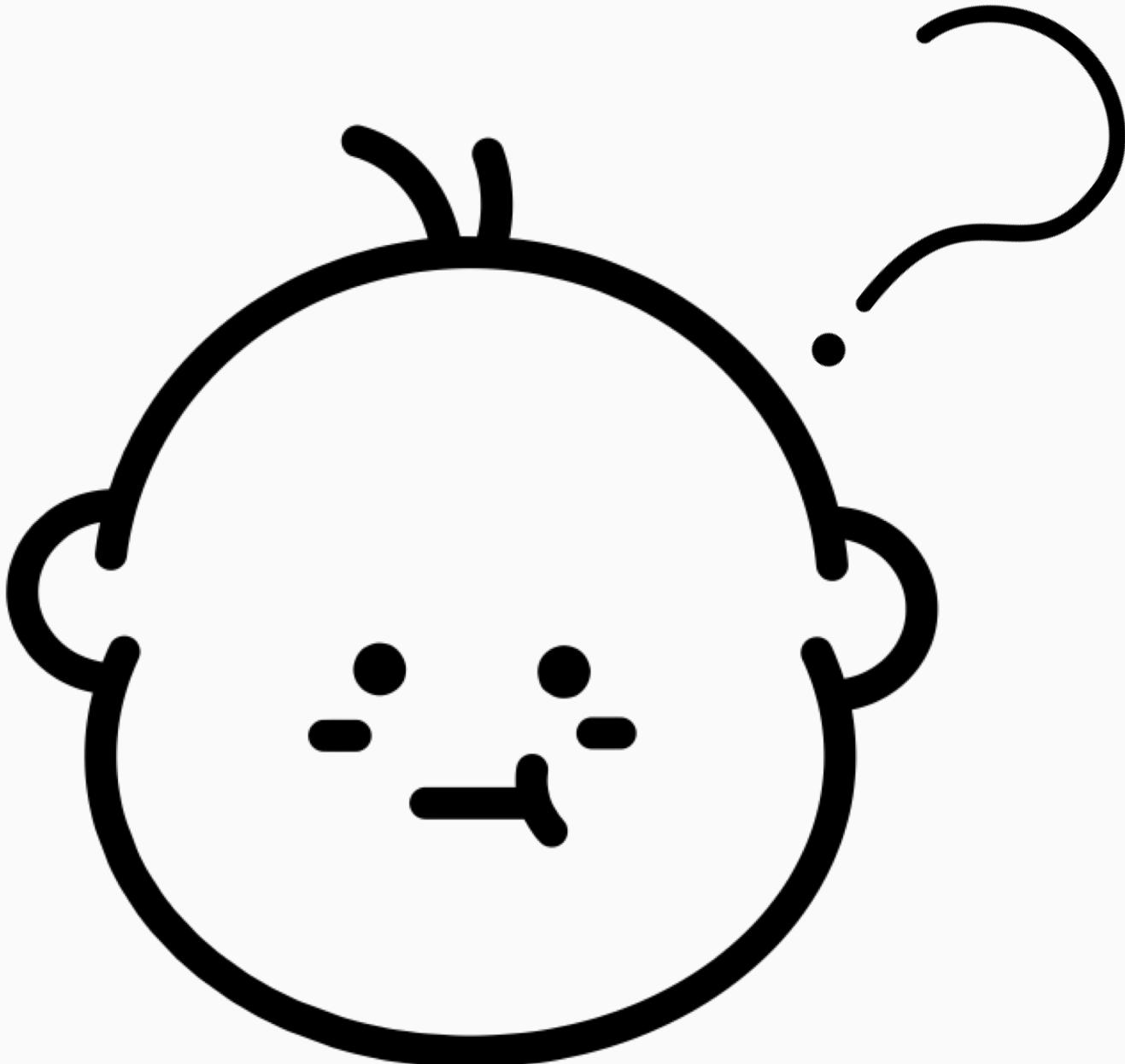


CLIPS

Computational Linguistics & Psycholinguistics
University of Antwerp

Giovanni Cassani

16/05/2019 Antwerp



CLIPS

Computational Linguistics & Psycholinguistics
University of Antwerp

Giovanni Cassani

16/05/2019 Antwerp

From the ground up: what is useful and how to spot it



CLIPS

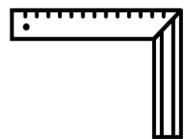
Computational Linguistics & Psycholinguistics
University of Antwerp

Giovanni Cassani

16/05/2019 Antwerp

**Usefulness is
information.**

**Information
is variance.**



The neighbors



The sound



CLIPS

Computational Linguistics & Psycholinguistics
University of Antwerp

Giovanni Cassani

16/05/2019 Antwerp

You shall know a
word by the
company it keeps.

J.R. Firth 1957

Why is *the* more informative than *and*?



CLIPS

Computational Linguistics & Psycholinguistics
University of Antwerp

Giovanni Cassani

16/05/2019 Antwerp

Diverse Frequent Interesting



CLIPS

Computational Linguistics & Psycholinguistics
University of Antwerp

Giovanni Cassani

16/05/2019 Antwerp

Categories emerge from uncertainty



CLIPS

Computational Linguistics & Psycholinguistics
University of Antwerp

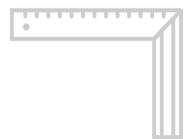
Giovanni Cassani

16/05/2019 Antwerp

**Direct way to
quantify usefulness**

**Word and context-
level analysis**

Beyond frequency



The neighbors



The sound



CLIPS

Computational Linguistics & Psycholinguistics
University of Antwerp

Giovanni Cassani

16/05/2019 Antwerp

Why does *smirgles* sound more like a verb?



CLIPS

Computational Linguistics & Psycholinguistics
University of Antwerp

Giovanni Cassani

16/05/2019 Antwerp

What are the properties of useful sound combinations?



CLIPS

Computational Linguistics & Psycholinguistics
University of Antwerp

Giovanni Cassani

16/05/2019 Antwerp

Frequent Diverse Interesting

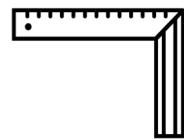


CLIPS

Computational Linguistics & Psycholinguistics
University of Antwerp

Giovanni Cassani

16/05/2019 Antwerp



The neighbors

v.



The sound

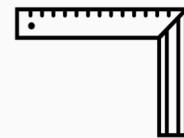


CLIPS

Computational Linguistics & Psycholinguistics
University of Antwerp

Giovanni Cassani

16/05/2019 Antwerp



Distributional cues:

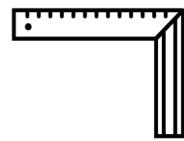
- **Diversity driven**
- **Only work with rich context**
- **Useful later in development**
- **Scale up to large vocabulary**



Phonological cues:

- **Frequency driven**
- **Relatively context insensitive**
- **Useful early in development**
- **Break up with large vocabulary**





The neighbors



The sound



CLIPS

Computational Linguistics & Psycholinguistics
University of Antwerp

Giovanni Cassani

16/05/2019 Antwerp

Whatever it takes: First sound, then neighbors



CLIPS

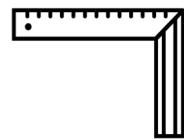
Computational Linguistics & Psycholinguistics
University of Antwerp

Giovanni Cassani

16/05/2019 Antwerp

**Phonological information
is widespread.
Distributional information
is concentrated.**

Learning to ignore.



The neighbors



The sound



CLIPS

Computational Linguistics & Psycholinguistics
University of Antwerp

Giovanni Cassani

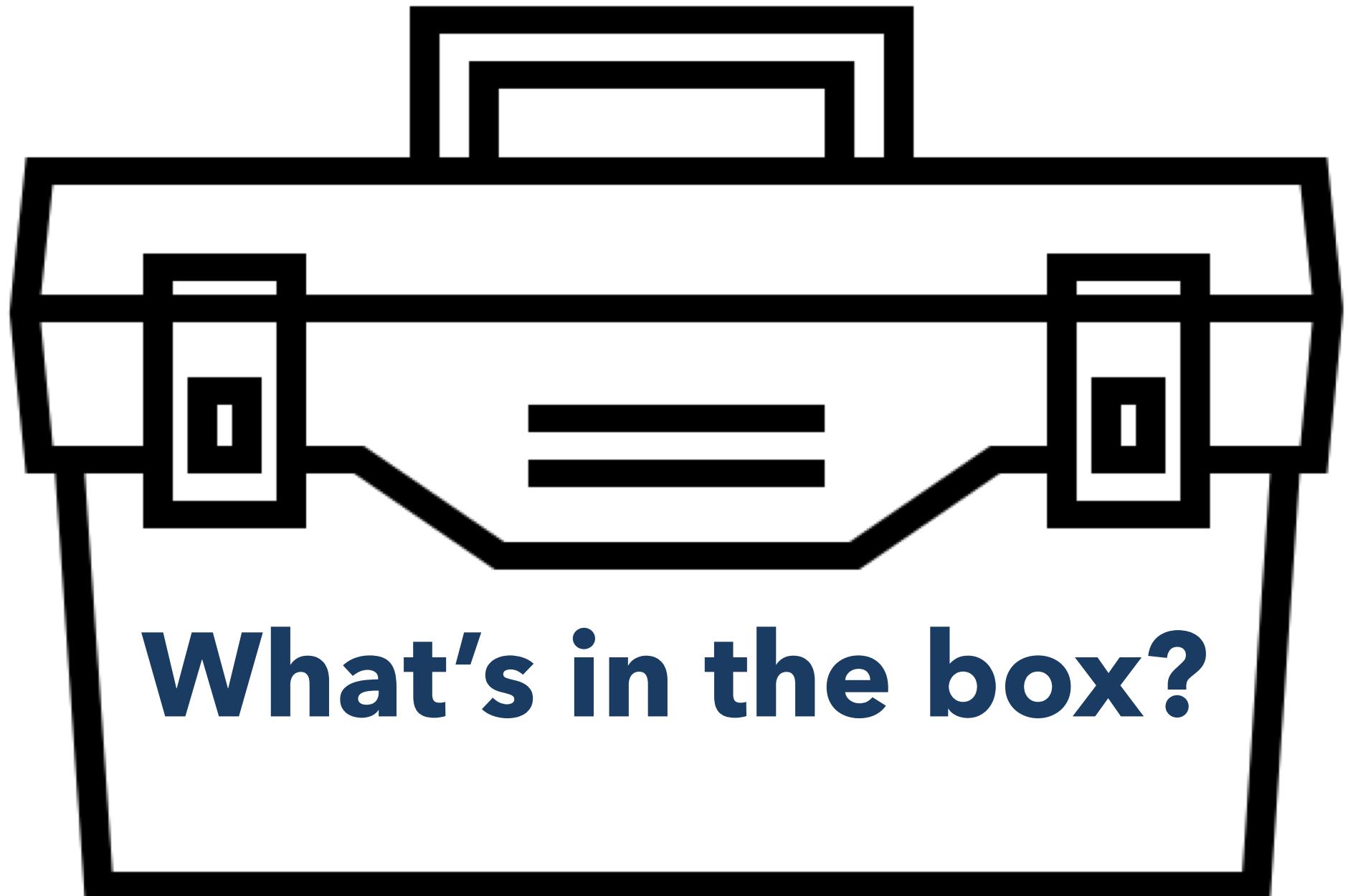
16/05/2019 Antwerp

Phonological and distributional cues map systematically



**Use one to know
about the other.**

**Bonus track:
Discriminate meaning
from sound and you'll
get categories largely
for free.**



Thank you!

Questions?