Useful contexts & easy words: Effects of distributional factors on lexical category acquisition

Giovanni Cassani

10 February 2017 – CLIN27

CLiPS
Computational Linguistics & Psycholinguistics
University of Antwerp
Patterns of distributional co-occurrences are informative about lexical categories [Redington & al, 1998] but some more than others. **What distributional properties of a context make it more useful?**

Words can be syntactically categorized using the contexts in which they occur [Harris, 1954] but some more easily than others. **What distributional properties of a word makes it easier to categorize?**
Overview

- Distributional bootstrapping
  - What have we done?
  - What do we miss?
- Computational simulation
- Results
  - Easy words
  - Useful contexts
- Conclusion
Overview

- Distributional bootstrapping
  - What have we done?
    - What do we miss?
  - Computational simulation
  - Results
    - Easy words
    - Useful contexts
  - Conclusion
Since Cartwright & Brent [1997] and Redington & al [1998], we know there is information in distributional co-occurrences that supports learning of lexical categories.

**Behavioral experiments** have confirmed that children are sensitive to this information and use it 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]
Evaluating 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.
Overview

- Distributional bootstrapping
  - What have we done?
  - What do we miss?
- Computational simulation
- Results
  - Easy words
  - Useful contexts
- Conclusion
A concept of *usefulness*

Not all contexts are equally informative:
- What does it mean for a context to be useful?
- How can a child determine this?
- **Using what information?**

Models work on too many starting assumptions that are not yet well motivated and grounded in experimental evidence.
A concept of easiness

Not all words are equally important: children are better at categorizing certain words than others.

- What causes certain words to be categorized better?
- Are words that are easier to categorize using distributional information also the words that children categorize better?
While definitely important, frequency is not enough in accounting for lexical category acquisition [Matthews & Bannard, 2010].

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

- Distributional bootstrapping
  - What have we done?
  - What do we miss?
- Computational simulation
  - Results
    - Easy words
    - Useful contexts
  - Conclusion
Computational simulation

Experimental setting

- **Unsupervised** PoS tagging experiment (5 tags)
- Transcribed English Child-directed speech
- Bigrams and trigrams (with utterance boundaries) as contexts \([b_{\_X}; \_c; a_{\_b_{\_X}}; b_{\_X_{\_c}}; \_c_{\_d}]\)
- Exemplar-based clustering (TiMBL: IB1, cosine, 1 NN, **no feature weighting**)
- Incremental training (40 to 70% of the input corpus)
- Mixed-effects models
Predictors and outcomes

- Context type (left, right, non-adj)
- # constituents (bigrams vs trigrams)
- Token frequency
- Diversity
- Average conditional probability
- Entropy (normalized)
- Time

- Gain ratio (contexts)
- Hits (words)
First, 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*.

Then, we computed Gain Ratio from co-occurrence statistics in the training. GR values were used as dependent variables to assess *usefulness*. 
Overview

- Distributional bootstrapping
  - What have we done?
  - What do we miss?
- Computational simulation
- Results
  - Easy words
    - Useful contexts
- Conclusion
Results - Easy words

Main effects
Overview

- Distributional bootstrapping
  - What have we done?
  - What do we miss?
- Computational simulation
- Results
  - Easy words
  - Useful contexts
- Conclusion
Results - Useful contexts

Main effects

- Frequency effect
- Diversity effect
- Time effect
- Bigram-Trigram effect
- Entropy effect
- Conditional probability effect
- Type effect
Results - Useful contexts

Interactions
Results - Useful contexts

Time-frequency interaction

Giovanni Cassani (CLiPS) - 10/02/2017 Leuven
Time-probability interaction
Results - Useful contexts

Time-entropy interaction

Giovanni Cassani (CLiPS) - 10/02/2017 Leuven
Results - Useful contexts

Time-diversity interaction

Giovanni Cassani (CLiPS) - 10/02/2017 Leuven
Overview

- Distributional bootstrapping
  - What have we done?
  - What do we miss?
- Computational simulation
- Results
  - Easy words
  - Useful contexts
- Conclusion
Useful contexts need to be **highly systematic**, and the more so with more exposure to the input:

- occur frequently
- with many different words
- occur a comparable amount of time with all the words they co-occur with
- **but** be hard to predict given the words they occur with

\[
\text{it}_X\#\text{end}; \ x\_\text{the}; \ you\_X; \ldots
\]
Beware of the noise

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

- occur with fewer contexts
- have low entropy distributions over contexts
- are hard to predict given the contexts in which they occur (cf. positive effect of diversity of usefulness)

*apple; forget; table; door; …*
Words that make good contexts are harder to categorize, while poor contexts consist of words that are categorized more effectively $\Rightarrow$ Children categorize content words better and earlier, and use function words to do this.

A full distributional learning account can effectively explain lexical category acquisition.
Thank you!
Questions?
References

- Harris, Z. (1954). Distributional structure. Word, 10(2-3), 146-152
References


