Distributional Bootstrapping: a Memory-Based Learning Approach

Giovanni Cassani
CLiPS @ Uantwerpen
18th December 2015
CLIN26@Amsterdam
Outline

- Background
- Distributional bootstrapping
- Frame-based approaches
- Experiment
- Conclusions
Outline

- Background
- Distributional bootstrapping
- Frame-based approaches
- Experiment
- Conclusions
A chicken-and-egg problem: understanding a system when you only have access to the system itself.

What information might provide the starting point? How does a child bootstrap her language learning?
A chicken-and-egg problem: understanding a system when you only have access to the system itself.

What information might provide the starting point? How does a child bootstrap her language learning?

- Semantics? (Pinker, 1984)
A chicken-and-egg problem: understanding a system when you only have access to the system itself.

What information might provide the starting point? How does a child bootstrap her language learning?

- Semantics? (Pinker, 1984)
- Prosody? (Christophe et al, 2008)
A chicken-and-egg problem: understanding a system when you only have access to the system itself.

What information might provide the starting point? How does a child bootstrap her language learning?

- Semantics? (Pinker, 1984)
- Prosody? (Christophe et al, 2008)
- Syntax? (Gleitman, 1990)
A chicken-and-egg problem: understanding a system when you only have access to the system itself.

What information might provide the starting point? How does a child bootstrap her language learning?

- Semantics? (Pinker, 1984)
- Prosody? (Christophe et al, 2008)
- Syntax? (Gleitman, 1990)
- Distributional co-occurrences? (Maratsos, 1980)
Outline

- Background
- Distributional bootstrapping
- Frame-based approaches
- Experiment
- Conclusions
Distributional bootstrapping proposes that children track co-occurrences among linguistic units in order to understand a great deal of how language works.

Distributional information has been shown to be useful to segment words (Saffran et al. 1996, Goldwater et al., 2009), identify phonemes (Maye et al., 2002), build a semantic space (Baroni et al., 2007), and more.
A specific task in which distributional cues are very useful is the **grouping of words** into lexical categories, which are in turn essential to master language. Different proposals so far have investigated

if certain **types of contexts** are more useful than others  
(Mintz, 2002)

if a handful of **lexically specific** cues can do the job  
(Mintz, 2003)

what **computational mechanisms** the child can use  
(Parisien, 2008)
The interest in lexically specific cues stems from the questionable assumption that children cannot evaluate all possible distributional cues and thus have to focus on just some of them. However, it also gives a very transparent, usage-based account of lexical category acquisition.

The question shifts to what such cues are and how they can be identified.
Outline

- Background
- Distributional bootstrapping
- Frame-based approaches
- Experiment
- Conclusions
The most frequent \textit{a\_X\_b} tri-grams in which two specific words flank a slot that can be taken by any other word, are a very good cue to group the words occupying the empty slot.

Mintz defined most frequent as 45, called such cues \textit{frequent frames}, and showed that they are \textit{extremely precise} in grouping words. However, precision comes at the expense of recall.
the most frequent \textit{a} \textit{X} and \textit{X} \textit{b} bi-grams are better

St Clair and colleagues took the 45 most frequent words, built 90 (left and right) bi-grams, called them \textit{flexible frames}, implemented a feed-forward NN and showed that such cues were better than frequent frames: a bit less precision, but far better recall.
Where do frames come from?

If the picklock to language is in distributional co-occurrences, frames’ saliency should be defined distributionally, with no a-priori restrictions.

What frames become salient when no assumptions are built-in and only distributional information is considered?
Outline

- Background
- Distributional bootstrapping
- Frame-based approaches
- Experiment
- Conclusions
Evaluate learning of broad lexical categories
(i) using **lexically specific** cues
(ii) under different **combinations of restrictions** on available types of contexts and distributional information
(iii) on **typologically different** languages, to see how important morphological complexity is.
Experiment

Goal

Evaluate learning of broad lexical categories
(i) using \textbf{lexically specific} cues
(ii) under different \textit{combinations of restrictions} on available types of contexts and distributional information
(iii) on \textit{typologically different} languages, to see how important morphological complexity is.

Contexts can be only \textit{bi-grams}, \textit{only tri-grams}, or \textbf{both}. Distributional information include \textit{token frequency}, \textit{type frequency} and average conditional probability of context given word, $P(\text{context}|\text{word})$. 

Giovanni Cassani
CLiPS
Amsterdam
18/12/2015
We mapped gold-standard PoS tags to a five categories tag-set, consisting of **Nouns**, **Verbs**, **Adjectives**, **Adverbs**, and **Function** words, to focus on content words categorization.

We evaluate **recall and precision** on the words that were uttered by all children and all caregivers.
Each corpus was divided in sections according to the age in months; we selected cues and trained the MBL incrementally on the first $n$ months and tested on the last $m$ months ($n$ and $m$ vary depending on the language).
Experiment

Results – English, recall

![Graphs showing recall for different types and training sections](image-url)
Experiment

Results – English, precision
Experiment

Results – French, recall

![Graphs showing recall results for different conditions over training sections.](image-url)
Experiment

Results – French, precision
Experiment

Results – Hebrew, recall

![Graphs showing recall rates for different types of Hebrew expressions](image)
Experiment

Results – Hebrew, precision
Outline

- Background
- Distributional bootstrapping
- Frame-based approaches
- Experiment
- Conclusions
Conclusions

Context restrictions

When bi-grams and tri-grams are both considered, the curve is similar to bi-grams at first and become more similar to tri-grams later, suggesting that it might take more time for longer regularities to emerge when no restriction is imposed.
Generally, when more information is made available, learning is better, both relatively – stronger increase from first to last stage, and absolutely – higher figures at the end of learning.

However, more information comes with more salient cues being stored, challenging the idea of lexical specificity as a solution to reduced capacities.
We clearly see that irrespective of the model performance goes down and learning fades when the morphological complexity of the language increases.

One can always run the same models using morphemes, but then it’s not clear how the child decides upon the right level of analysis.
We should stop

- assuming language-specific linguistic knowledge in models of bootstrapping (words vs morphemes)
- assuming that higher-order tasks are tackled only after lower-order tasks are accomplished (categorization after segmentation)

We should

- move to perceptually motivated cues and linguistically motivated outcomes (Baayen et al, 2015).
- isolate which pieces of information makes cues salient, and test their sufficiency and necessity


Conclusions

References


Conclusions

Thank you!

- Questions?
- Suggestions!
- Criticisms!?