Prosody as a means of acquiring syntactic categories and building a syntactic skeleton

Gutman A.*, Dautriche I.,# Crabbé B. & Christophe A.,#

*Zukunftscolleg, Konstanz University, Germany #LSCP, Ecole Normale Supérieure/CNRS, Paris *Alpage, Paris Diderot University

Background & Objective

How can children learn a language without explicit training?

Children acquiring their native language have to learn abstract syntactic categories (nouns, verbs, adjectives etc.) - these may facilitate the acquisition of word meanings [1] as well as the syntactic structures of the language. Two-year-olds already have an abstract representation of nouns and verbs [2]

What underlies this feat? We propose to simulate this process with a computational model based on experimental evidence.

A model for bootstrapping syntactic acquisition

We already know that:

- 9-month-olds perceive prosodic boundaries [3]
- Function words are frequent and situated at the border of prosodic boundaries [4]
- Toddlers use function words to categorize the next content words [5]

Given this knowledge, we may construct an approximate syntactic structure [6]:

- Prosodic structure is treated as an approximate shallow syntactic structure.
- A prosodic phrase = a syntactic phrase
- Function words and content words serve to label categorically the prosodic phrases (NP, VP).

Result: A syntactic skeleton.

Goal: Testing the feasibility of the labeling component of the model.

Overview of the computational model

Input: Transcribed corpus of child directed speech augmented with prosodic boundaries following the algorithm of [7].

Learning simulation: A prosodic phrase is seen as a probabilistic event, to which we associate features:

- The observed features are words at the edge of the phrase (i.e., E).
- The hidden feature is the syntactic category of the phrase (NP, VP).

We predict the category $\phi$ of each prosodic phrase from its features $\mathbf{f}$ using Naive Bayes:

$$\phi = \text{arg max}_{\phi} p(\phi) \prod_{i} p(f_i | \phi)$$

The prior probability of the category $p(\phi)$ and the probability of the different features given that category $p(f_i | \phi)$ are estimated via the Expectation-Maximization algorithm [7]:

- Initialization: Initially assign the prosodic phrases to syntactic clusters.
- Maximization: Estimate the parameters of the model.
- Expectation: Re-assign the prosodic phrases to new syntactic clusters.
- Reiterate the two last steps until convergence.

Experiment 1 - Unsupervised clustering using Function words

Assumptions

- Children use function words to classify phrases.
- Function words are frequent and often the first word of a phrase [3].

Implementation

- Assign each prosodic phrase to a cluster labelled by its first word (the feature $f$).
- Retain only the clusters corresponding to the most frequent function words.
- Re-assign the rest of the corpus using the EM algorithm.

Experiment 2 - Semi-supervised classification using Content words

Assumptions

- The child knows already some of the most frequent content words [8], and can identify them as objects (nouns) or actions (verbs) [6].
- Based on this knowledge she can identify the class of the current prosodic phrase.

Implementation

- Pre-define a list of known frequent content words based on the corpus: a semantic seed
- Use the last word of each phrase (the feature $f$) to assign it to one of three classes: Nominal, Verbal or Other.
- Re-assign the unclassified part of the corpus using EM algorithm.

Evaluation & Results

The quality of each cluster is evaluated by examining the largest syntactic category which it has captured: purity $(C)$ Number of bits of largest category

<table>
<thead>
<tr>
<th>Cluster size</th>
<th>Pure-Word Int. Avg. 100% Purity</th>
<th>Random Int. Avg. 100% Purity</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.75</td>
<td>0.25</td>
</tr>
<tr>
<td>4</td>
<td>0.80</td>
<td>0.30</td>
</tr>
<tr>
<td>5</td>
<td>0.85</td>
<td>0.35</td>
</tr>
</tbody>
</table>

Comparison with a uniform random clustering of 3 clusters:

- The precision is very good even with a small semantic seed (highly precise VP and NP clusters, which span about 50% of these categories).
- The recall is lower since the known words do not cover all contents.
- But precision is more important than recall for learning!
- Although based on content words, the learning algorithm ultimately relies on function words: knowing a few content words allows the learner to discover the function words associated with the category label.

Conclusion

- The syntactic skeleton hypothesis is corroborated: The language learner can rely on prosodic boundaries, function words and content words in order to construct a syntactic skeleton.
- Experiment 1: While the prosodic boundaries give the structure of the prosodic skeleton, our baseline random EM shows that the prosodic structure alone does not permit the inference of meaningful phrasal categories. However relying on function words permits the construction of good categories (high purity).
- Experiment 2: Relying on knowledge of frequent content words can lead to emergence of abstract syntactic categories. These abstract categories (VP and NP) are grounded in linguistic experience (frequency of words) and semantic experience (verbs = actions and nouns = objects).
- Our model shows that a small semantic seed allows the discovery of function words and that the knowledge of function words may help in the classification of novel words.

References


Implementation

- Experiment 1 - Unsupervised clustering using Function words

Assumptions

- The child knows already some of the most frequent content words [8], and can identify them as objects (nouns) or actions (verbs) [6].
- Based on this knowledge she can identify the class of the current prosodic phrase.

Implementation

- Pre-define a list of known frequent content words based on the corpus: a semantic seed
- Use the last word of each phrase (the feature $f$) to assign it to one of three classes: Nominal, Verbal or Other.
- Re-assign the unclassified part of the corpus using EM algorithm.

Evaluation & Results

We vary the size of the semantic seed (known words) starting from (3 Nouns, 1 Verb) and compare to a uniform random clustering of 3 clusters:

- The precision is very good even with a small semantic seed (highly precise VP and NP clusters, which span about 50% of these categories).
- The recall is lower since the known words do not cover all contents.
- But precision is more important than recall for learning!
- Although based on content words, the learning algorithm ultimately relies on function words: knowing a few content words allows the learner to discover the function words associated with the category label.

Conclusion

- The syntactic skeleton hypothesis is corroborated: The language learner can rely on prosodic boundaries, function words and content words in order to construct a syntactic skeleton.
- Experiment 1: While the prosodic boundaries give the structure of the prosodic skeleton, our baseline random EM shows that the prosodic structure alone does not permit the inference of meaningful phrasal categories. However relying on function words permits the construction of good categories (high purity).
- Experiment 2: Relying on knowledge of frequent content words can lead to emergence of abstract syntactic categories. These abstract categories (VP and NP) are grounded in linguistic experience (frequency of words) and semantic experience (verbs = actions and nouns = objects).
- Our model shows that a small semantic seed allows the discovery of function words and that the knowledge of function words may help in the classification of novel words.

* Corresponding author: ariel.gutman@uni-konstanz.de