Background

Category Learning Problem in Natural Language
How do kids discover that they have multiple categories (of words, nouns)?
How do they learn what belongs in what class, and learn to assign new items to a class?

Correlated Cues in Artificial Language Learning

Lexicon: 2 classes of words (1 and 2), and 2 classes of dependents (1 and 2)
Dependency: words from Class 1 can only cooccur with Class 2 / dependents

To learn regular dependencies between items, learners need partially correlating information on some members of each class. The dependency and the correlating information make up the correlated cue.

Investigating the correlated cue in natural language acquisition:
(1) do they exist in natural language?
(2) are they in children’s input?
(3) are children sensitive to the correlating information?
(4) is category learning dependent on the correlated cue?

Future work will determine how the correlated cue works (computational models), and whether the artificial language results really parallel what appears to go on in natural language

Tsez (Dido)

Nakh-Dagestani language with 4 noun classes spoken by about 6,000 people in Dagestan

Dependency

Class 1 | Class 2 | Class 3 | Class 4
---|---|---|---
-igu uži | -igu kid | -igu k'r'tu | -igučča
-igood boy(/l) | -il-good girl(/i) | -ill-good cat(i) | -ll-good soup(t)

Regular noun class agreement over most vowel initial verbs, adjectives and adverbs

Is there Partially Correlating Information for each class?

Real Words:

Classification Experiment to elicit classification of Real and Nonce words containing different predictive features by native Tsez speakers

Nonce Words:

Subjects: 32 Native Tsez speakers in Shamil and Kizilyurt, Dagestan (10 young children (~ 6y), 12 older children (~ 9y), 10 adults)

Eat/Don’t Eat Task: -i (~eat, intransitive), and -ač (~eat, transitive) show overt agreement with the subject and object respectively

Assistant introduces each character and item, participant tells the character to eat, then what to eat/not eat. Class assignment is evident in the agreement.

(1) Do Correlated Cues Exist in Tsez?

Is there information on a subset the nouns in each class that correlates with class? This, in conjunction with agreement, would constitute a correlated cue.

Decision Trees are built by a supervised learning algorithm that takes words with specified features and determines which features are most predictive of class

Plaster & Harizanov (2009): Built decision trees classifying Tsez nouns from a dictionary (Xalilov, 1999), but we want to use input that reflects actual language use.

Need a corpus, Build a corpus

-3,000 lines (10 hours) of child directed Tsez speech, transcribed with the help of native speakers

Use nouns from the corpus instead of the dictionary to build a decision tree reflecting words actually in use => set of highly predictive features

(2) Do Correlated Cues Exist in the Children’s Input?

- From Corpus: predictive feature on nouns triggering overt agreement
- Class 1: 100%, Class 2: 52%, Class 3: 51%, Class 4: 45%
- 84% of verb tokens and 77% of adjective tokens show overt agreement
- Are there enough examples to be useful?
- future work with artificial languages will help determine what is enough

(3) Are Speakers Sensitive to the Correlating Information?

Critical Findings

- Correlated Cues exist in Tsez
- Not all statistically predictive information is used equally
  - not just an effect of differences in predictiveness
  - for children phonological cues on real words more powerful than semantic cues
  - children ignore weak semantic cues entirely
- Differences between children and adults suggest a bias to use certain kinds of cues when discovering classes

Experimental Results

Older Children

Adults

Younger Children