Synergies in learning words and their referents

Mark Johnson
Katherine Demuth
Michael Frank
Bevan Jones

- Input: unsegmented utterances tagged with contextual objects
- Output: word segmentation and word to object mapping

PIG|DOG ▲ I ▲ Z ▲ ð ▲ æ ▲ t ▲ ð ▲ œ ▲ p ▲ I ▲ g ▲
Word to object “topic models” as PCFGs

- Objects in non-linguistic context $\approx$ sentence topics
- Such topic models can be expressed as *Probabilistic Context-Free Grammars*
- PCFG rules *choose a topic* from possible topic marker and *propagate it through sentence*
- Each word is either generated by sentence topic or a special null topic
- Requiring *at most one topic per sentence*:
  - improves accuracy
  - can be expressed by PCFG
Adaptor grammars for word segmentation

- Adaptor grammars (AGs) generalise PCFGs by learning probability of entire subtrees
  - Prob. of adapted subtree $\propto$ number of times tree was previously generated + $\alpha \times$ PCFG prob. of subtree
  - AGs are hierarchical Dirichlet or Pitman-Yor Processes

- AG for unigram word segmentation:

  Words $\rightarrow$ Word $|$ Word Words
  Word $\rightarrow$ Phons
  Phons $\rightarrow$ Phon $|$ Phon Phons

- Segmentation accuracy improves if AG learns collocations
Joint segmentation and object-mapping

- Combine word-object “topic PCFGs” with word segmentation AGs
- Synergies in learning:
  - improving topic detection improves word segmentation
    70% $\rightarrow$ 75% f-score
  - improving word segmentation improves topic detection
    50% $\rightarrow$ 74% f-score
- Joint (rather than staged) learners can exploit these synergies
- Are there similar synergies in other aspects of language acquisition?
- Do human learners exploit such synergies?