Evidence for Prosodic vs Statistical Segmentation

Recall, there are at least two ways to do segmentation:

- **prosodic information** — e.g. English nouns likely to be trochaic
- **statistical information** — e.g. end of word unlikely to be /spl/

Peter Jusczyk ’97:

- **subjects**: English-household 9mo infants
- **stimuli**: Dutch / Mandarin word lists (similar/dissimilar prosody)
- **measure**: fixation time in conditioned head turn procedure
- **results**: segmentation in Dutch, not Mandarin
Evidence for Statistical vs Prosodic Segmentation

Erik Thiessen & Jenny Saffran ’03:

- **subjects:** 7mo / 9mo infants
- **stimuli:** synthesized sequences of nonsense words:
  (a) ‘DObi DApu BUgo DIti …’ (+ random lights, for interest)
  (b) ‘dobi daPU buGO diTI …’ (+ random lights, for interest)

metrical + statistical information \( P(bi \mid do) = 1, \quad P(da \mid bi) = 0.2 \text{ to } 0.4 \)

then:
1. central light until fixation;
2. side light until fixation;
3. side light and audio until look away:
   (a) ‘diti, budo’ (complete words)
   (b) ‘pudo, bida’ (partial words)

- **measure:** fixation time in conditioned head turn procedure
- **results:** 7mo subjects prefer (novel) partial words despite prosody,
  9mo subjects prefer (novel) iambic words, w. effect for statistics
Heather Bortfeld & al ’05: familiar words help babies segment speech

▶ subjects: 6mo infants
▶ stimuli:
  1. 30 seconds of recordings with familiar ‘Mommy’ / unfamiliar ‘Lola’ word:
     (a) ‘Mommy’s feet get sore from standing all day.’
        ‘The doctor wants Mommy’s feet to be clean.’
        ‘The red shoes felt best on Mommy’s feet.’
     (b) ‘Lola’s dog ran around the yard.’
        ‘He patted Lola’s dog on the head.’
        ‘The neighborhood kids played with Lola’s dog.’

     (interrupted if infant looked away from speaker)
  2. central light until fixation (judged by observer with headphone music);
  3. side light until fixation;
  4. side light and audio until look away: repetitions of ‘feet/dog’
▶ measure: fixation time in conditioned head turn procedure
▶ results: facilitation for word ‘feet’ next to familiar word ‘Mommy’
But segmentation is only part of larger problem of grammar induction.

For a while, nativists (discontinuity hypothecists) cited **Gold’s theorem**:  
- learner may hypothesize grammar $G$ with, say, ‘*hat*’ always singular  
- learner may never have seen evidence of ‘*hat*’ used as plural  
- but learner can’t rule out grammar $G' \supset G$, with ‘*hat*’ also plural  
...as evidence that some form of **principles** must constrain grammar.  

Basically, the problem is general lack of **negative examples** in language.  
(This is also called a **poverty of the stimulus** argument.)

Problem with application of Gold’s theorem to poverty of stimulus:  
- learner can statistically induce neg. examples from ‘suspicious absence’!  
  (if ‘*hat*’ can be plural, how come I have never seen it before?)
Top-down (Bayesian) Grammar Induction

Define a simple induction model:

- $\tilde{F}(c \rightarrow d \ e)$: frequency of rule $c \rightarrow d \ e$ hypothesized thus far
- $\tilde{F}(d)$: frequency of category $d$ hypothesized thus far
- $\tilde{F}()$: number of categories hypothesized thus far
- $\alpha, \beta$: prior assumed frequency of new rule / new category

Estimate probability of new/old grammar rule at category $c$ during parsing:

$$P(c \rightarrow d \ e \mid c) \overset{\text{def}}{=} \frac{\tilde{F}(c \rightarrow d \ e)}{\tilde{F}(c) + \alpha}$$

$$+ \frac{\alpha}{\tilde{F}(c) + \alpha} \cdot \left( \frac{\tilde{F}(d)}{\tilde{F}() + \beta} + \frac{\beta}{\tilde{F}() + \beta} \right) \cdot \left( \frac{\tilde{F}(e)}{\tilde{F}() + \beta} + \frac{\beta}{\tilde{F}() + \beta} \right)$$
Top-down (Bayesian) Grammar Induction

\[
P(c \rightarrow d e | c) \overset{\text{def}}{=} \frac{\tilde{F}(c \rightarrow d e)}{\tilde{F}(c) + \alpha} + \frac{\alpha}{\tilde{F}(c) + \alpha} \cdot \left( \frac{\tilde{F}(d)}{\tilde{F}(d) + \beta} + \frac{\beta}{\tilde{F}(d) + \beta} \right) \cdot \left( \frac{\tilde{F}(e)}{\tilde{F}(e) + \beta} + \frac{\beta}{\tilde{F}(e) + \beta} \right)
\]

Parser proceeds top-down, and for each rule application . . .

- uses old rule with probability proportional to \( \tilde{F}(c \rightarrow d e) \)
- uses new rule with probability proportional to \( \alpha \),
and for each category . . .

- uses old category with probability proportional to \( \tilde{F}(d) \) or \( \tilde{F}(e) \)
- uses new category with probability proportional to \( \beta \)

Some interesting effects:

- frequently used rules/categories get more frequently used
- number of rules/categories tends to converge (depending on \( \alpha, \beta \))
- longer, frequently-used categories (verbs) tend to be higher in the tree
Top-down (Bayesian) Grammar Induction

\[
P(c \rightarrow d e | c) \overset{\text{def}}{=} \frac{\tilde{F}(c \rightarrow d e)}{\tilde{F}(c) + \alpha} + \frac{\alpha}{\tilde{F}(c) + \alpha} \cdot \left( \left( \frac{\tilde{F}(d)}{\tilde{F}(d) + \beta} \right) + \frac{\beta}{\tilde{F}(d) + \beta} \right) \cdot \left( \left( \frac{\tilde{F}(e)}{\tilde{F}(e) + \beta} \right) + \frac{\beta}{\tilde{F}(e) + \beta} \right)
\]

For example (from CHILDES corpus, McWhinney’00):

▶ /yuwanttusidēbuk/ : yuwanttusidēbuk → /yuwanttusidēbuk/
▶ /lukdērzēbcviwhizhæt/ : yuwanttusidēbuk → /yuwanttusidēbuk/
  lukdērzēbcviwhizhæt → /lukdērzēbcviwhizhæt/
▶ /ændədægi/ : yuwanttusidēbuk → /yuwanttusidēbuk/
  lukdērzēbcviwhizhæt → /lukdērzēbcviwhizhæt/
  ændədægi → /ændədægi/
▶ /yuwanttsukædis/ : yuwanttusidēbuk → /yuwanttu/ sidēbuk (B-aN)
  sidēbuk → /sidēbuk/
  sidēbuk → /lukædis/
  lukdērzēbcviwhizhæt → /lukdērzēbcviwhizhæt/
  ændədægi → /ændədægi/
For next time... 

Read: 

- Traxler ch 9, pp. 343–360