Recitation 3, 6.S077
Babytalk Part II
Statistical Learning by 8-Month-Old Infants

Jenny R. Saffran, Richard N. Aslin, Elissa L. Newport

Learners rely on a combination of experience-independent and experience-dependent mechanisms to extract information from the environment. Language acquisition involves both types of mechanisms, but most theorists emphasize the relative importance of experience-independent mechanisms. The present study shows that a fundamental task of language acquisition, segmentation of words from fluent speech, can be accomplished by 8-month-old infants based solely on the statistical relationships between neighboring speech sounds. Moreover, this word segmentation was based on statistical learning from only 2 minutes of exposure, suggesting that infants have access to a powerful mechanism for the computation of statistical properties of the language input.
Birdsong & human sound systems: what’s the same?

Bengalese finch
(*Lonchura striata domestica*)
Source: K. Okanoya, 2003

In well formed words, sibilants agree in the feature [anterior].

1. \([s, z, t, s', t', s, t, t']\) are never preceded by \([f, s, t, t, t, t']\).
2. \([f, t, t, f, t', t']\) are never preceded by \([s, z, t, s', t']\).

Examples (Sapir and Hoijer 1967):

1. *fite³*  ‘we (dual) are lying’
2. *dasdo:lis*  ‘he (4th) has his foot raised’
3. *fite:z*  (hypothetical)
4. *dasdo:lj*  (hypothetical)
An animal model for human learning?

Bengalese finch
(*Lonchura striata domestica*)
Source: K. Okanoya, 2003
Sound system components: birds & people

“Beads on a string” model:

1. Beads – chunks or “states” that are categorical classes (remember: “s-sh”)

2. Linear sequence – one state follows another, in constrained way (e.g., “slo” starts a possible English word, but “rdz” does not)

= A finite-state automaton

Categorical production and perception

Address just one part of that: how do we find the ”chunks” in the input?
Navajo phonotactics: s, ∫ cannot precede one another
(Source: Heinz, 2007; 2010)

Bengalese finch song
The simplest linear patterns = regular

ba:d → bat; de:g → dek (Heinz, 2007)
ji:tʒ, *ji:teːz
(Chandlee & Jardine, 2013)

What’s the same?

- “Critical period” for learning from external experience
- Babbling (subsong), practice & self-practice
- Plasticity frozen at puberty (by hormonal change – testosterone)
- Left-lateralization for system
- Brain circuitry control
- Beads on a string structure
Songbirds – Zebra finch “critical period” learning
Table 1. Distinctive Features of American English Consonants

<table>
<thead>
<tr>
<th>Back</th>
<th>High</th>
<th>Coronal</th>
<th>Anterior</th>
<th>Labial</th>
<th>Continuant</th>
<th>Lateral</th>
<th>Nasal</th>
<th>Sonorant</th>
<th>Strident</th>
<th>Voiced</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
### All English sounds

Table 1. Distinctive Features of American English Consonants

|        | p | b | m | f | v | θ | ð | t | d | n | s | z | l | r | s | z | j | tʃ | dʒ | ɹ | ʃ | ɾ | η | w | ʔ | h |
| **Back** | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | + | + | + | + | + | + | + | + | + |
| **High** | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | + | + | + | + | + | + | + | + | + |
| **Coronal** | - | - | - | - | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | - | - | - |
| **Anterior** | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | - | - | - | - | - | - | - | - | - |
| **Labial** | + | + | + | + | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| **Continuant** | - | - | - | + | + | + | + | - | - | + | + | - | + | + | - | - | - | + | + | + | + | + | + | + | + | + |
| **Lateral** | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| **Nasal** | - | + | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | + | + | + | + | + | + | + | + |
| **Sonorant** | - | + | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| **Strident** | - | - | + | + | - | - | - | - | - | + | + | - | - | - | - | - | - | - | + | + | + | + | - | - | - | - |
| **Voiced** | - | + | + | - | + | - | - | - | + | + | - | + | + | - | - | - | - | + | + | + | + | + | + | + | + | + |

Table 2. Distinctive Features of American English Vowels

<table>
<thead>
<tr>
<th>i</th>
<th>ι</th>
<th>e</th>
<th>ɛ</th>
<th>æ</th>
<th>u</th>
<th>o</th>
<th>ɔ</th>
<th>a</th>
<th>ʌ</th>
<th>ə</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
“Use it or lose it” Learning

• In English, we have words like these: right-light; fry-fly; fur-fill
• So, English baby must retain this contrast – it is the difference in 2 distinctive features, lateral and continuant
• What about other languages?
  • Korean: Korea-Seoul – not contrastive
  • Result: Korean babies lose r/l distinction, lose the ability to discriminate
  • Use of categories and rules results in decline of perceptual abilities
  • No animals do this with human speech; Korean dogs and monkeys do not lose the l/r contrast
r, l in both the onset and the coda, so must be distinguished (also: fly/fry)

Other languages?
Challenge: segmentation
twasbrilligandtheslithytovesdidgyre

\{pabiku, tibudo, daropi, golatu\}

pabikutibudodaropipabiku
tibudodaropitibudodaropi
pabikudaropipabikugolatu
tibudogolatutibudogolatu
golatudaropipabikutibudo
daropigolatudaropipabiku
tibudogolatudaropigolatu
daropigolatupabikutibudo
pabikutibudodaropigolatu…

pigola  daropi  tudaro

(sounds)
Challenge: Combining Inference with Cognitive Constraints
(How real people solve real problems can help real computers)

Problem: \textit{twasbrillig} and the \textit{sithy toves} did gyre and gimble

“Standard” solution: prettybaby \textit{pre-ty-ba-by}

Graph of transition probabilities: Pr(x_{i+1}|x_i) & look for local minima

“Standard” claim: works great; “stats is all you need” (Science, 1996)

\begin{align*}
\text{Pr}(b_1|p_a) &= 1.0; \quad \text{Pr}(k_u|b_i) = 1.0; \quad \text{Pr}(t_i|k_u) = 0.3, \\
\text{Pr}(b_u|t_i) &= 1; \quad \text{Pr}(d_o|b_u) = 1.0; \quad \text{Pr}(d_a|d_o) = 0.3 \\
\text{Pr}(r_o|d_a) &= 1; \quad \text{Pr}(p_i|r_o) = 1.0; \quad \text{Pr}(g_o|p_i) = 0.3 \\
\text{Pr}(l_a|g_o) &= 1.0; \quad \text{Pr}(t_u|l_a) = 1.0 \quad \text{...}
\end{align*}

\textit{pabiku} \quad \textit{tibudo} \quad \textit{daropi} \quad \textit{golatu} \quad \textit{daropi}

\textit{pabikutibudodaropigolatu}... 
Works great? NO!!!
Actual results on actual speech to children: works lousy
What’s the answer? But, add a **ONE** universal constraint about human language and it works GREAT!

---

Precision and Recall, Pure Stat
Interference vs. Stat Inference + UG,
250,000 child-directed examples

Using the universal constraint

Only transitional statistics

What IS this ONE universal constraint???? HINT: you all **know** it!
17

@Loc: Eng-NA-MOR/Brent/c1/cl-0917.cha
@PID: 11312/c-08015454-1
@Begin 1
@Languages: eng
@Participants: CHI Morgan Child, MOT Brenda Mother
@ID: eng/Brent[CHI]0.9.17|female|Child
@ID: eng/Brent[MOT]|Mother
@Birth of CHI: 28-MAR-1996
@Media: c1-t14jan97, audio
@Date: 14-JAN-1997

"MOT: pull it up yourself(f)!

"MOT: hands up!

"MOT: hands up!

"MOT: pull profit prep|up pro|refl|yourself!

"MOT: hands up!

"MOT: hands up!

"MOT: hands up!

"MOT: now hands out!

"MOT: what you are doing & noise!

"MOT: pro|what aux|be|PRES pro|you part|do|PRES?

"MOT: hey .

"MOT: let's wash that hand!
Strategies for learning words: 6 methods

1. Use isolated words, e.g., “ball”, “hey”
   • What does corpus analysis show?
   • Mother-to-child speech: 9% of all utterances are isolated words
   • This strongly correlates with timing of child learning that word – good!
   • What’s the big open question?
   • How? – bad!
   • Does length of utterance work?
   • lsee vs. spaghetti
   • NO workable algorithm proposed for extracting isolated words...
Strategies for learning words

2. Use statistics
   • Transitional probabilities (TPs) between adjacent syllables, A, B
   • $TP(A \rightarrow B) = \frac{\text{Prob}(AB)}{\text{Prob}(A)}$, where probabilities are estimated by frequencies
   • Word boundaries at points of local minima
   • E.g., $TP(\text{pre} \rightarrow \text{tty})$ & $TP(\text{ba} \rightarrow \text{by})$ both $> TP(\text{tty} \rightarrow \text{ba})$, so “tty-ba” local minimum and so likely word break
   • This is the essence of the Saffran, Aslin, Newport experiment w/ 8.5 month old babies exposed to 2 minutes of artificial speech
Strategies for learning words

2. Statistical methods, continued:  
Evolutionary: probably old? Hauser et al. 2001, cotton-top tamarin monkeys

<table>
<thead>
<tr>
<th></th>
<th>Language A</th>
<th>Language B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Words</td>
<td>tupiro, golabu, bidaku, padoti</td>
<td>tudaro, pigola, bikuti, budopa</td>
</tr>
<tr>
<td>Test words</td>
<td>tupiro, golabu</td>
<td>tudaro, pigola</td>
</tr>
<tr>
<td>Test non-words</td>
<td>dapiku, tilado</td>
<td>tigobu, kudabi</td>
</tr>
<tr>
<td>Test part-words</td>
<td>tibida, kupado</td>
<td>pabiku, tibudo</td>
</tr>
</tbody>
</table>
Hypothesis:
Like babies; orient to novel stimuli
Fig. 3. Mean (standard error) percent of word versus partword test trials on which subjects responded, for Language A (left) and Language B (right). Black bars indicate responses to word trials, stippled bars indicate responses to partword trials.
Strategies for learning words

3. Metrical segmentation
   • 90% of English content words (? What’s that?) are stress initial in conversational speech (Cutler & Carter, 1987)
   • So maybe stressed syllable = beginning of word
   • Back to crying - Evidence for metrical detection: 7.5 month old babies detect strong-weak pattern in English fluent speech better than weak-strong pattern
   • “taris” extracted by babies as word from “guitaris” – why?
   • What are the problems?
     • Language specific (Consider French vs. German again)
     • Bootstrapping: How does infant know the metrical pattern for their language?
     • Use known words, but where do these come from?
4. Phonotactic constraints

- What makes a well-formed syllable?
- Pight, clight, zight vs. flight, dnight, ptight. Which are “possible” English words, which are not?
- Only certain consonant clusters are valid “onsets” in English (Halle, 1978)
- Language specific, so must come from experience (plus any initial templates)
- How might this be useful?
  - Sound sequence “vt”, break word between ”v” and “t”
  - Problem: sometimes clusters that don’t occur in onsets are in fact parts of words
  - Can you think of one?
  - “embed” → mb
Syllabification in a sense logically prior – infant keeps track of tp’s over syllables

r, l in both the onset and the coda, so must be distinguished (also: fly/fry)

Other languages?
Strategies for learning words

5. Allophonic constraints
   • Say what?
   • “tab” vs. “cat” – what’s the difference in the “t”?
   • Aspirated vs. unaspirated: word boundaries can have articulatory diffs
   • Again assumes infant can pick these out
   • Doesn’t this assume infant can first find the boundaries?
   • Nitrates vs. night rates
Strategies for learning words

6. Memory

- Sound patterns extracted and stored in memory for later use – helps with new words
- 8-month old infants can store “python” “vine” “peccaries” and remember them as familiar when embedded in stories with speaker and word order variation, even though it’s highly unlikely they know what these words are
- Can then use these patterns to extract new words: e.g., if you learn “savory” you can use that to learn “unsavory”

No one factor at work – let’s see how they can be put together

Use linguistic representations in conjunction with “small” processing power

Now let’s evaluate some models – first a word about measuring performance
all word segmentations in (test) set done correctly = all word segmentations, done incorrectly

correct segmentations in test set reported by the program = incorrect segmentations reported by the program

(selected elements) = segmentations returned by the program

(Segmented test set)

Precision and Recall (0-100%)

Precision

Recall

Usually give weighted average of Precision and Recall, “F-measure”
The input: mother’s speech to children, from “Brown corpus” in CHILDES

• How to make training data? Run this through CMU Pronouncing Dictionary
• Divides word into syllables and tells us stress
• “cat” → K AE1 T
• Stress runs from 0 (stress free), 1 (primary stress), 2 (secondary), through 9
• “catalog” → K AE1 T AH0 L AO0 G”,
• “catapult” → K AE1 T AH0 P AH2 L T
• Then group phonetic segments into syllables
• Easy in English: maximize length of onset so long as it is a valid consonant cluster
• Example. “Einstein” is “AY1 N ST AY0 N” by CMU, in syllables: AY1N STAY0N because /st/ is longest onset; /nst/ is longer but violates English phonotactics
The training corpus

• Finally, remove punctuation and word boundaries, but keep utterance boundaries between sentences (line breaks in CHILDES)
• Result: 226,178 words, consisting of 263,660 syllables
• OK, let’s see how well the various methods do....first, statistics & tp
Transitional probability in practice

• On the plus side: it is the only language-independent method (so no chicken-and-egg problem)

• Has been shown to be influential in children early (as early as 7 months), compared to, e.g., stress

• Assume: child has syllabified speech perfectly (Why?)

• Assume: child has neutralized effects of stress among variants of syllables (Why? There are 58,884 unique syllables not looking at stress; if you count stress, lots more difft syllables – must compute tp’s for all of the pairs you find)

• Assume: data for training same as data for testing (Why? Unusual ML condition... Why do this?)
Process entire training corpus & then

• There is a word boundary between syllables AB and CD
  if $TP(A \rightarrow B) > TP(B \rightarrow C) < TP(C \rightarrow D)$
How well does this work?

• Lousy. Precision is 41.6%, Recall is 23.3 %

• In other words, about 60% or words posited by statistical learner are not English words, and almost 80% of actual English words are not extracted, even under these favorable learning conditions

• Why?

• Clue: 226,178 words, consisting of 263,660 syllables

• So most words are 1 syllable. What does tp do?

• Most words are 1 syllable, followed by another 1 syllable word 85% of the time
Transitional Probability absolute value of changes declines rapidly as # of syllables processed grows – there are so many syllables the tp can’t change much.

Figure 1: Transitional Probability absolute value of changes during the course of training. Note the rapid stabilization of TPs.
The unique stress constraint (USC)

• The only known mechanism that takes advantage of the abundance of single word utterances

• If the learner hears an utterance that contains exactly one primary stress, she can immediately conclude that such utterance, regardless of its length, can and can only be a single word

• $W_1 S_1 S_2 S_3 W_2 \Rightarrow 3$ words $W_1 S_1 S_2 S_3 W_2$

• Can help statistical learning: $S_1 W_1 W_2 W_3 S_2$ provides cues: at least 2 words, and the string of $W$’s has a word boundary somewhere – perhaps use transitional probability there
USC has fewer assumptions than metrical segmentation learning

Metrical segmentation assumes:

a) Recognize strong vs. weak syllables
b) A collection of reliably segmented words
c) A computation that finds the dominant pattern in the set of words

For USC, only (a) required

It’s universal – no chicken-and-egg problem

But how do kids pick up stress? We seem to hear it, but how?
How do children figure out stress for the word segmentation problem?
Consonant-Vowel pattern in babbling: universal

(a) Syllable structure of the word *plans*.

The right representation is *combinatorial*
This way? The Beat Generation

Tell me not in mournful numbers...

Yields beat pattern
The Beat Generation

Iambic: mark **left** as “head” & project to next level

Suggests: there is an operation that takes **two** items & “merges” them
Why do we say the USC is “innate”

• Where could it come from?
• Statistical learning can’t generate a good candidate set, and it’s the only other language independent method known
• USC is also a “negative” principle – how do you know it’s not violated by some “other” example?
• If child only gets positive examples, then this is hard to figure out (Why?)
• In any case, we can now come up with a variety of models that use the USC
Transitional probabilities + USC

1. Apply usual statistical analysis to get transitional probabilities
   a) If two strong syllables are adjacent (S₁S₂), a word boundary is posited in between
   b) If there are more than 1 weak syllable between 2 strong syllables (S₁W...WS₂), then a word boundary is posited where the pairwise tp is at the local minimum

2. (a) solves monosyllabic problem; (b) has some complications – if multiple local minima (“drinking the champagne”)

3. Results: precision = 73.5%; recall = 71.2% (comparable to best methods in literature which use a very computationally intensive optimization algorithm)
Algebraic learning

• Can we do without statistical learning?
• Note that computational burden of tp’s is not trivial
• 58,448 unique syllable pairs
• Whenever learner sees an occurrence of, e.g., A, it has to adjust values of all the B’s in tp(A→B)
• So learner has to adjust values of potentially thousands of tp’s for every syllable processed in input – might be too computationally costly
Algebraic segmentation

• Suppose we use known words to bootstrap novel words
• 8 month olds can retain sound patterns in memory (Juscyk & Holmes, 1997)
• Kid can extract “big” from “bigsnake” and so extract “snake”
• Other evidence kids can do this:
  • “hiccing up” from “hicc-up”
  • “two dults” from “a-adult”
• The method works like this:
  1. Use the USC
  2. At word boundary, this might not work: $S_1W_1...W_nS_2$ (‘languageacquisition’) there are 2 possibilities:
     a) If both $S_1W_{i-1}$ and $W_{j+1} < j$ are part of known owords on both sides, then $W_j$ must be a word
     b) Otherwise, word boundary somewhere in the string of W’s, and USC doesn’t help
  3. In case (b), we can use two strategies: (1) agnostic: skip this one for now; (2) pick random position in the W’s to make two words, one containing $S_1$ the other one $S_2$. But in both cases, no word is added to dictionary (learner is unsure)
The logic behind the agnostic learner is that the learner is non-committal if the learning data contains uncertainty unresolvable by “hard” linguistic constraints such as USC. This could arise for two adjacent long words such as “language acquisition”, where two primary stresses are separated by multiple weak syllables as in the case of (6b). It could also arise when the input data (casual speech) is somewhat degraded such that some primary stresses are not prominently pronounced, as discussed in 5.2. While the agnostic learner does not make a decision when such situations arise, it can be expected that the words in the sequence $S_1Wn_1S_2$ will mostly like appear in combinations with other words in future utterances, where USC may directly segment them out. The random learner is implemented as a baseline comparison, though we suspect that in actual language acquisition, the learner may invoke the language-specific Metrical Segmentation Strategy, rather than choosing word boundaries randomly, in ambiguous contexts such as $S_1Wn_1S_2$.

Note further that in both versions of the algebraic model, no word is added to the lexicon when the learner is unsure about the segmentation; that is, both algebraic learners are conservative and conjecture words only when they are certain. This is important because mis-segmented words, once added to the lexicon, may lead to many more mis-segmentations under the subtraction algorithm. In section 6.1, we discuss ways in which this assumption can be relaxed.

Table 1 summarizes the segmentation results from the two algebraic learners, along with those from earlier sections on statistical learning.

<table>
<thead>
<tr>
<th>Model</th>
<th>Precision</th>
<th>Recall</th>
<th>F-measure ($\alpha = 0.5$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SL</td>
<td>41.6%</td>
<td>23.3%</td>
<td>0.298</td>
</tr>
<tr>
<td>SL + USC</td>
<td>73.5%</td>
<td>71.2%</td>
<td>0.723</td>
</tr>
<tr>
<td>Algebraic agnostic</td>
<td>85.9%</td>
<td>89.9%</td>
<td>0.879</td>
</tr>
<tr>
<td>Algebraic random</td>
<td>95.9%</td>
<td>93.4%</td>
<td>0.946</td>
</tr>
</tbody>
</table>

It may seem a bit surprising that the random algebraic learner yields the best segmentation results but this is not unexpected. The performance of the agnostic learner suffers from deliberately avoiding segmentation in a substring where word boundaries lie. The random learner, by contrast, always picks out some word boundary, which is very often correct. And this is purely due to the fact that words in child-directed English are generally short. Taken 11 A comparable case of this idea is the Structural Triggers Learner (Fodor, 1998) in syntactic parameter setting. We thank Kiel Christianson for pointing out this connection. 24
Summary

• Word segmentation can get off the ground only through use of language-independent means: experience-independent linguistic constraints such as the Universal Stress Constraint (USC) & experience dependent statistical learning are the only candidates we know so far

• Statistical learning does not scale up to realistic settings of language acquisition

• Simple principles drawing on USC can improve statistical learning and improve it, but computational of statistical learning may still be prohibitive

• Algebraic learning under the USC, with trivial computational cost, in principle universal, outperforms all other segmentation models