Recitation 3, 6.S077
Babytalk Part II

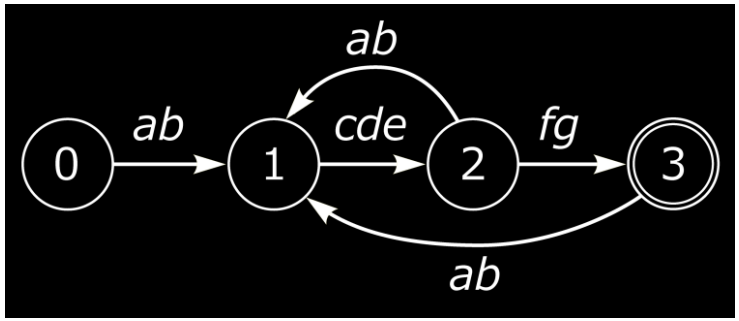
Statistical Learning by 8-Month-Old Infants

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

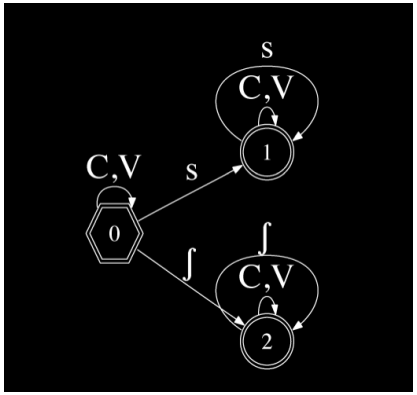
Learners rely on a combination of experience-independent and experience-dependent mechanisms to extract information from the environment. Language acquisition involves both types of mechanisms, but most theorists emphasize the relative importance of experience-independent mechanisms. The present study 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.

SCIENCE • VOL. 274 • 13 DECEMBER 1996

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].

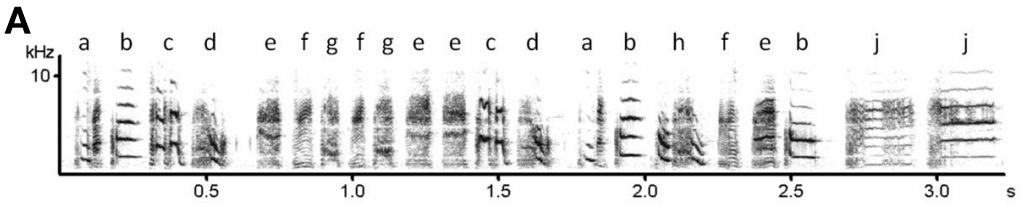
1. [s,z,ts,ts',dz] are never preceded by [ʃ,ʒ,tʃ,tʃ',dʒ].
2. [ʃ,ʒ,tʃ,tʃ',dʒ] are never preceded by [s,z,ts,ts',dz].

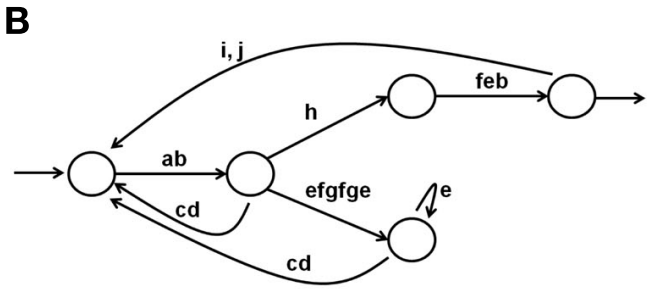
Examples (Sapir and Hojier 1967):

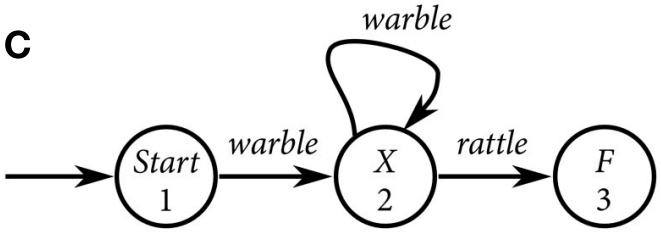
1. ʃi:te:ʒ 'we (dual) are lying'
2. dasdo:lis 'he (4th) has his foot raised'
3. *ʃi:te:z (hypothetical)
4. *dasdo:liʃ (hypothetical)

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)

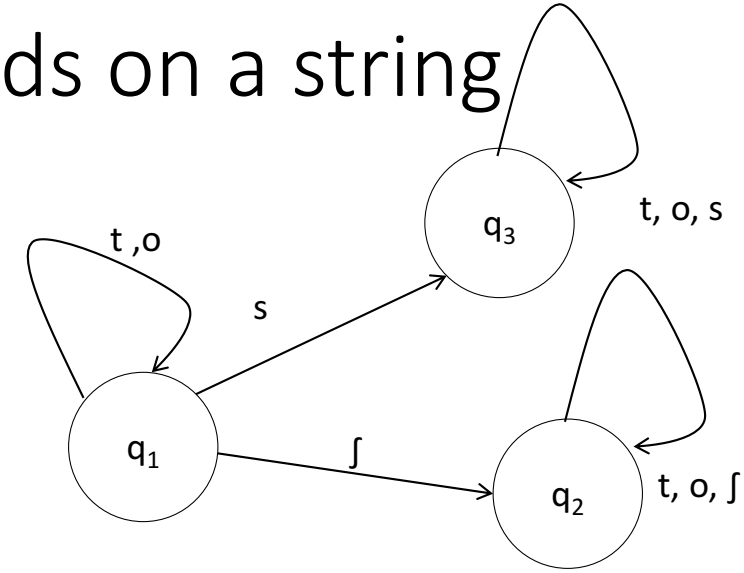
= A finite-state automaton

Categorical production and perception

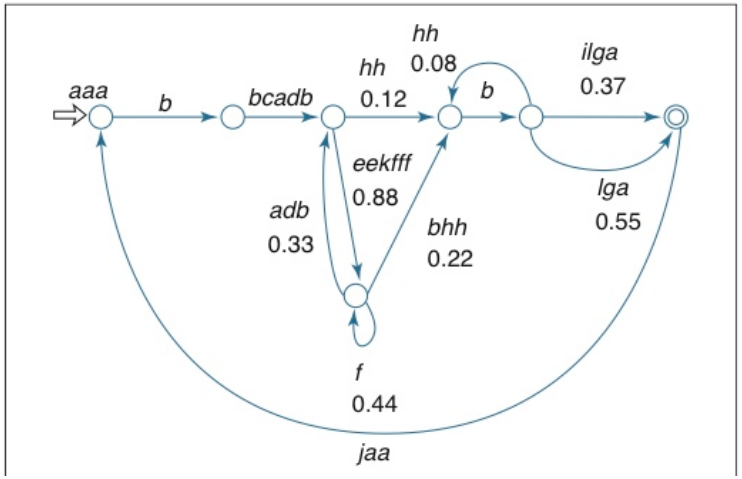
Address just one part of that: how do we find the "chunks" in the input?

Beads on a string

Navajo phonotactics:
s, ʃ cannot precede one another
(Source: Heinz, 2007; 2010)



Bengalese finch song



The simplest linear patterns = regular

ba:d → bat; de:g → dek (Heinz, 2007)

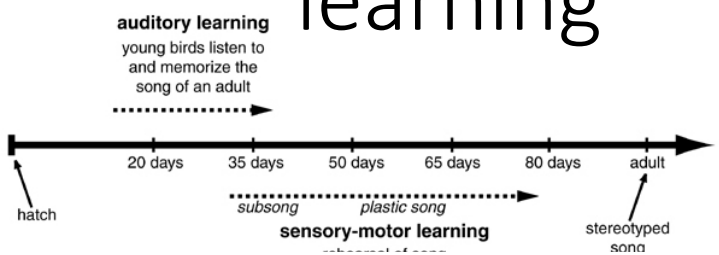
ʃi:tʒ, *ʃi:te:z
(Chandlee & Jardine, 2013)

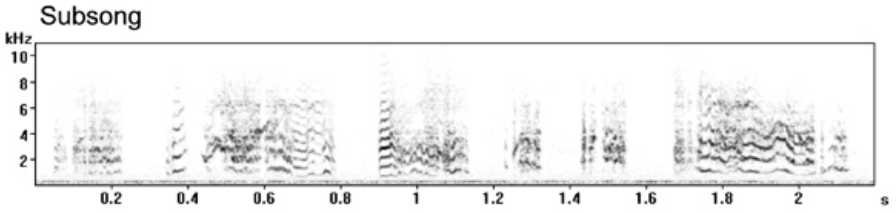
Lonchura striata domestica. Source: K. Okanoya, 2003

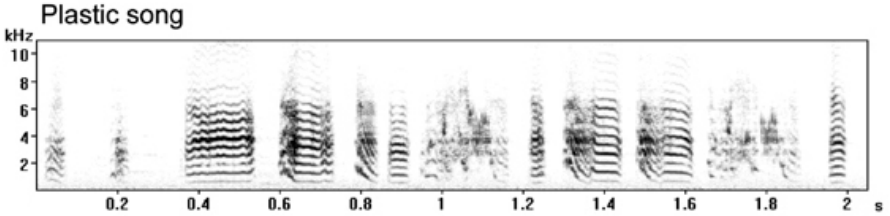
What's the same?

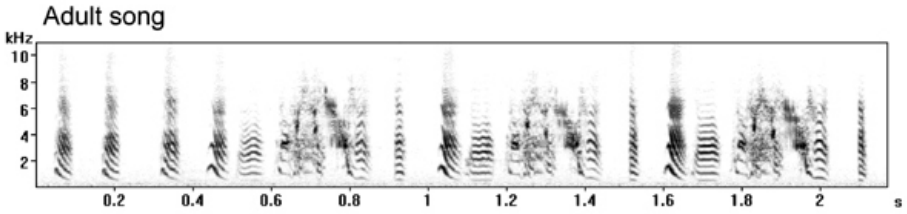
- "Critical period" for learning from external experience
- Babbling (subsong), practice & self-practice
- Plasticity frozen at puberty (by hormonal change – testosterone)
- Left-lateralization for system
- Brain circuitry control
- Beads on a string structure

Songbirds – Zebra finch “critical period” learning









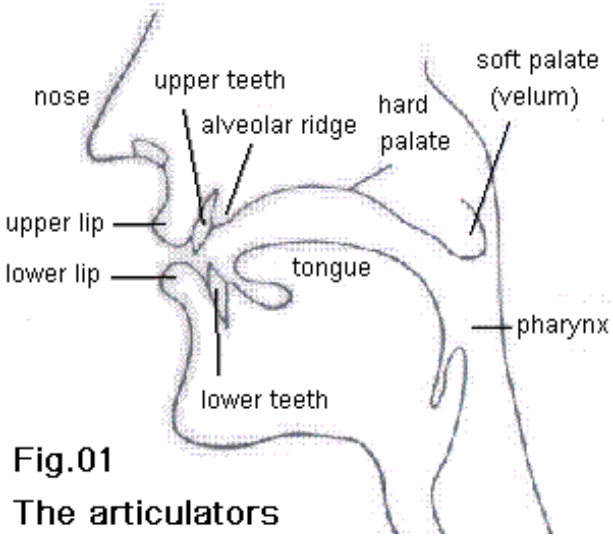


Fig.01

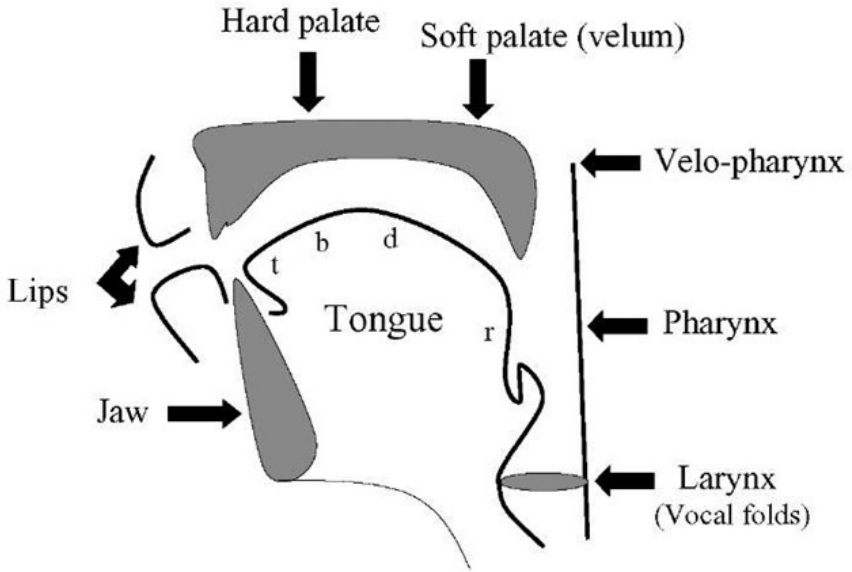
The articulators

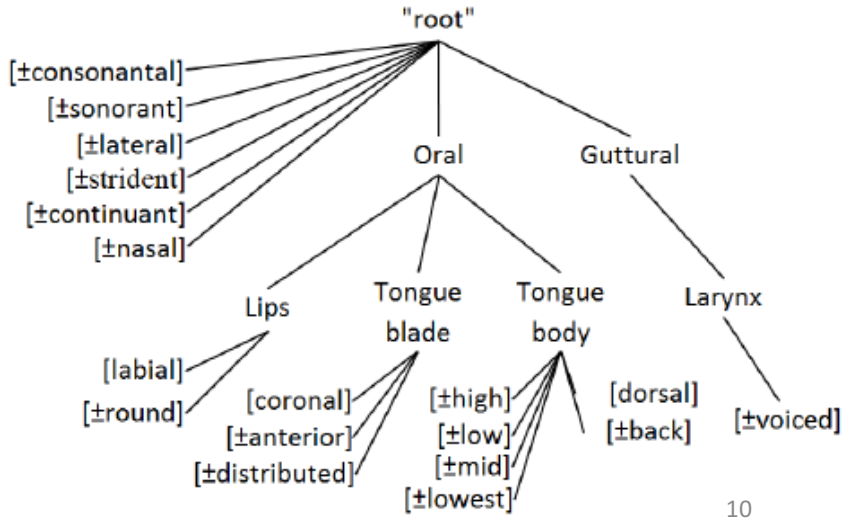
Table 1. Distinctive Features of American English Consonants

	p	b	m	f	v	θ	ð	t	d	n	s	z	l	r	ʃ	ʒ	ʧ	ʤ	j	ɹ	k	g	ŋ	w	ʔ	h
Back	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+
High	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	-	-
Coronal	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-
Anterior	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-
Labial	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-
Continuant	-	-	-	+	+	+	+	-	-	-	+	+	+	-	+	+	-	-	+	+	-	-	-	+	-	+
Lateral	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-
Nasal	-	-	+	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-
Sonorant	-	-	+	-	-	-	-	-	-	+	-	-	+	+	-	-	-	-	+	+	-	-	+	+	-	-
Strident	-	-	-	+	+	-	-	-	-	-	+	+	-	-	+	+	+	+	-	-	-	-	-	-	-	-
Voiced	-	+	+	-	+	-	+	-	+	+	-	+	+	+	-	+	-	+	+	+	-	+	+	+	-	-

Table 2. Distinctive Features of American English Vowels

i	ɪ	e	ɛ	æ	u	ʊ	o	ɔ	a	ʌ	ə	
+	+	-	-	-	+	+	-	-	-	-	-	**high**
-	-	-	-	+	-	-	-	-	+	+	-	**low**
-	-	-	-	-	+	+	+	+	+	-	-	**back**
-	-	-	-	-	+	+	+	+	-	-	-	**rounded**
+	-	+	-	-	+	-	+	-	-	-	-	**ATR**





All English sounds

Table 1. Distinctive Features of American English Consonants

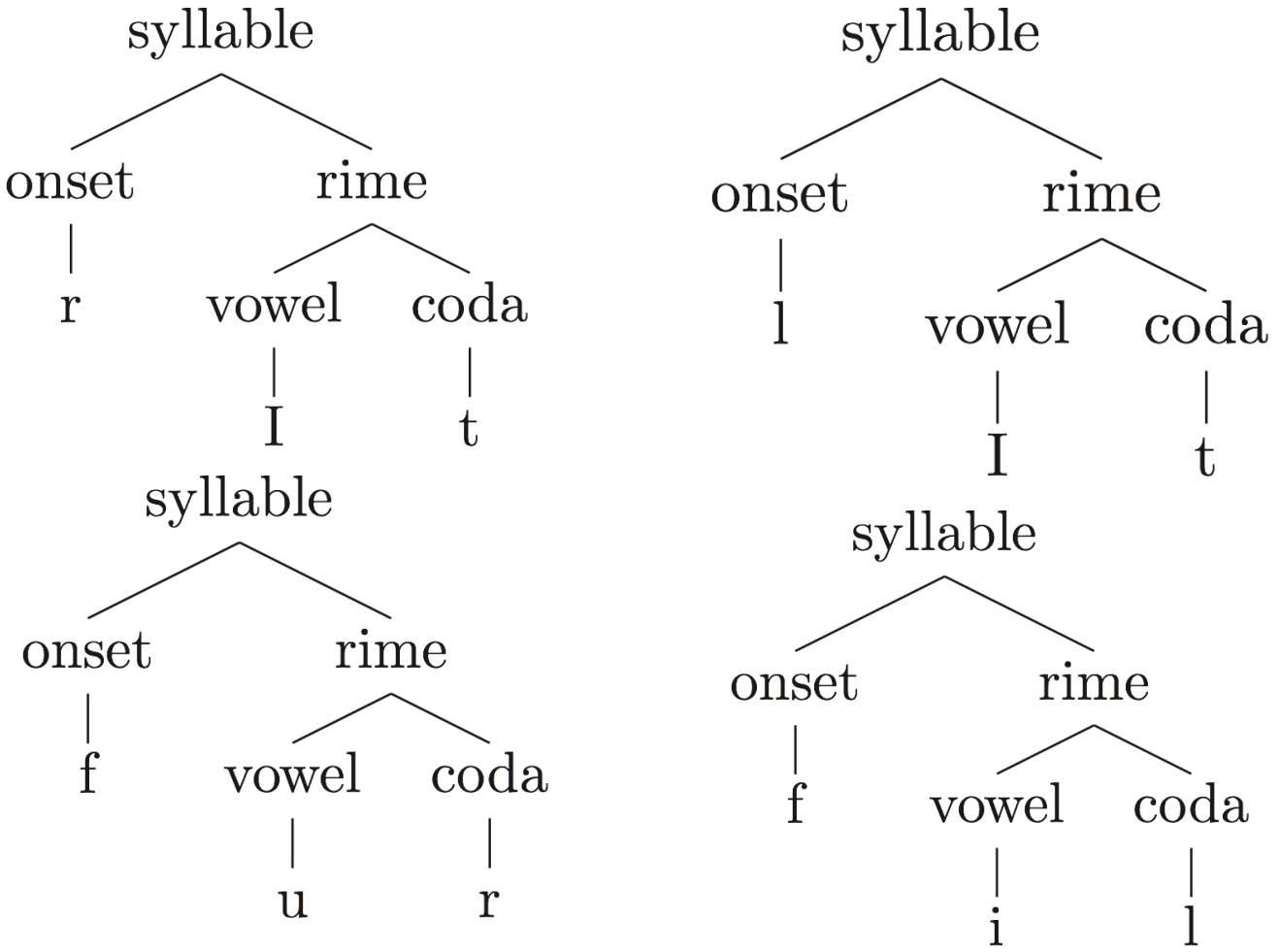
	p	b	m	f	v	θ	ð	t	d	n	s	z	l	ɾ	ʃ	ʒ	tʃ	dʒ	j	ɹ	k	g	ŋ	w	ʔ	h
Back	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+
High	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	-	-
Coronal	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-
Anterior	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-
Labial	+	+	+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

Table 2. Distinctive Features of American English Vowels

i	ɪ	e	ɛ	æ	u	ʊ	o	ɔ	a	ʌ	ə	
+	+	-	-	-	+	+	-	-	-	-	-	**high**
-	-	-	-	+	-	-	-	-	+	+	-	**low**
-	-	-	-	-	+	+	+	+	+	-	-	**back**
-	-	-	-	-	+	+	+	+	-	-	-	**rounded**
+	-	+	-	-	+	-	+	-	-	-	-	**ATR**

“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: Ko**r**ea-Seou**l** – 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…

(sounds)

pigola daropi tudaro

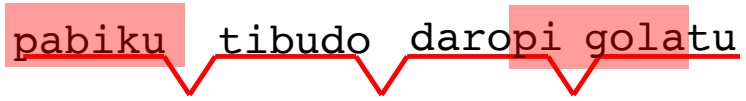
Challenge: Combining Inference with Cognitive Constraints (How real people solve real problems can help real computers)

Problem: *twasbrilligandtheslithytovesdidgyreandgimble*

"Standard" solution: *prettybaby* *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)




```
pabikutibudodaropigolatu…
Pr(bi|pa)=1.0; Pr(ku|bi)=1.0; Pr(ti|ku)=0.3,
Pr(bu|ti)=1; Pr(do|bu)=1.0; Pr(da|do)=0.3
Pr(ro|da)=1; Pr(pi|ro)=1.0; Pr(go|pi)=0.3
Pr(la|go)=1.0; Pr(tu|la)=1.0 …
```

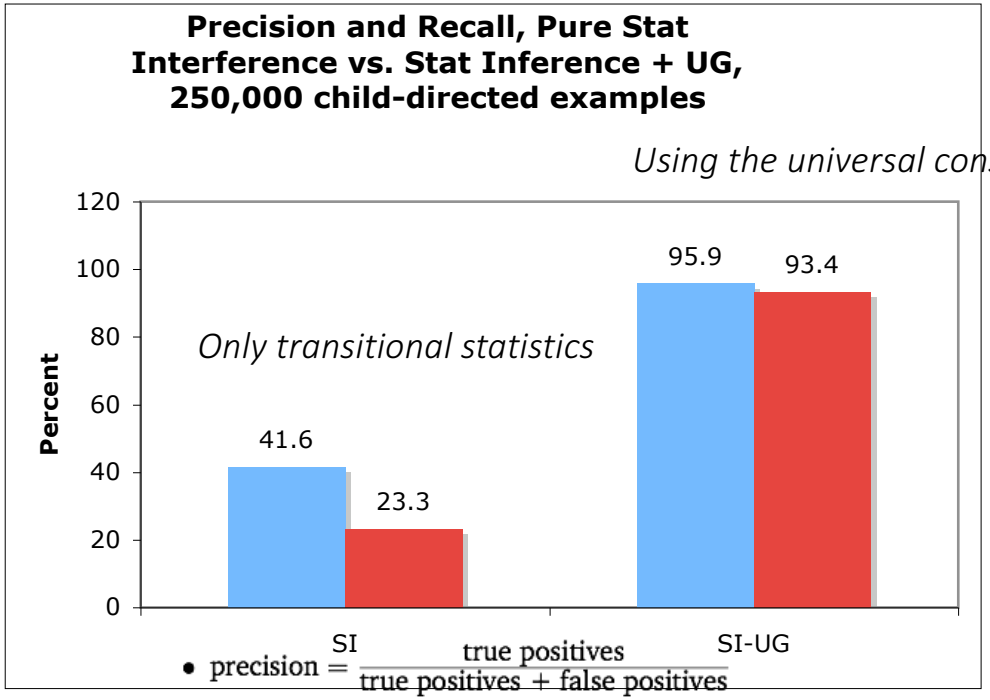
pigola→ pi golatu

tudaro→ tu daropi

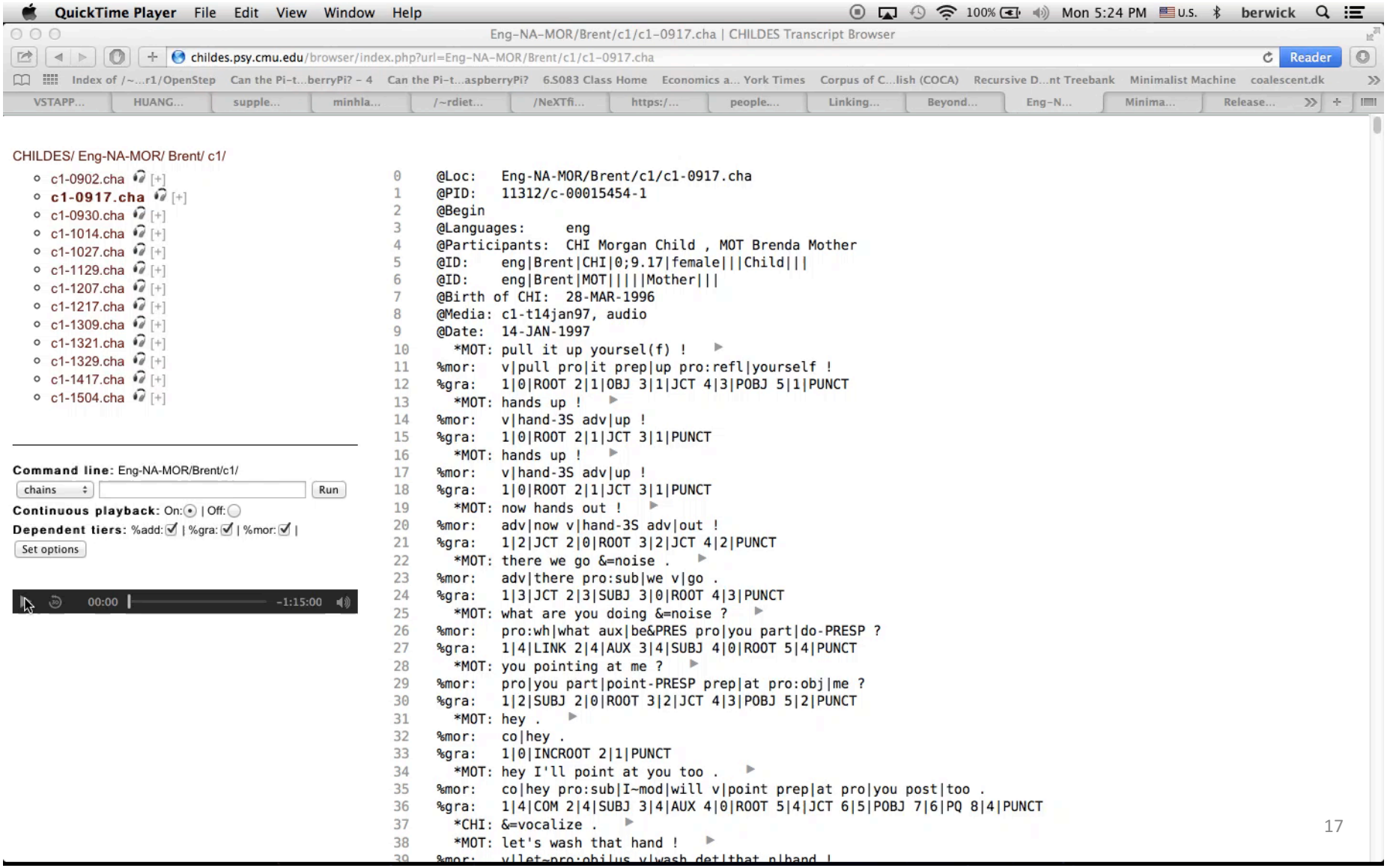
Works great? NO!!!

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!



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


QuickTime Player File Edit View Window Help
Mon 5:24 PM
berwick
Eng-NA-MOR/Brent/c1/c1-0917.cha | CHILDES Transcript Browser
childes.psy.cmu.edu/browser/index.php?url=Eng-NA-MOR/Brent/c1/c1-0917.cha
Reader
CHILDES/ Eng-NA-MOR/ Brent/ c1/
c1-0902.cha [+]
c1-0917.cha [+]
c1-0930.cha [+]
c1-1014.cha [+]
c1-1027.cha [+]
c1-1129.cha [+]
c1-1207.cha [+]
c1-1217.cha [+]
c1-1309.cha [+]
c1-1321.cha [+]
c1-1329.cha [+]
c1-1417.cha [+]
c1-1504.cha [+]
Command line: Eng-NA-MOR/Brent/c1/
chains
Run
Continuous playback: On: | Off:
Dependent tiers: %add: | %gra: | %mor: |
Set options
00:00
-1:15:00
0 @Loc: Eng-NA-MOR/Brent/c1/c1-0917.cha
1 @PID: 11312/c-00015454-1
2 @Begin
3 @Languages: eng
4 @Participants: CHI Morgan Child , MOT Brenda Mother
5 @ID: eng|Brent|CHI|0;9.17|female|||Child|||
6 @ID: eng|Brent|MOT|||||Mother|||
7 @Birth of CHI: 28-MAR-1996
8 @Media: c1-t14jan97, audio
9 @Date: 14-JAN-1997
10 *MOT: pull it up yoursel(f) !
11 %mor: v|pull pro|it prep|up pro:refl|yourself !
12 %gra: 1|0|ROOT 2|1|OBJ 3|1|JCT 4|3|POBJ 5|1|PUNCT
13 *MOT: hands up !
14 %mor: v|hand-3S adv|up !
15 %gra: 1|0|ROOT 2|1|JCT 3|1|PUNCT
16 *MOT: hands up !
17 %mor: v|hand-3S adv|up !
18 %gra: 1|0|ROOT 2|1|JCT 3|1|PUNCT
19 *MOT: now hands out !
20 %mor: adv|now v|hand-3S adv|out !
21 %gra: 1|2|JCT 2|0|ROOT 3|2|JCT 4|2|PUNCT
22 *MOT: there we go &=noise .
23 %mor: adv|there pro:sub|we v|go .
24 %gra: 1|3|JCT 2|3|SUBJ 3|0|ROOT 4|3|PUNCT
25 *MOT: what are you doing &=noise ?
26 %mor: pro:wh|what aux|be&PRES pro|you part|do-PRESP ?
27 %gra: 1|4|LINK 2|4|AUX 3|4|SUBJ 4|0|ROOT 5|4|PUNCT
28 *MOT: you pointing at me ?
29 %mor: pro|you part|point-PRESP prep|at pro:obj|me ?
30 %gra: 1|2|SUBJ 2|0|ROOT 3|2|JCT 4|3|POBJ 5|2|PUNCT
31 *MOT: hey .
32 %mor: co|hey .
33 %gra: 1|0|INCROOT 2|1|PUNCT
34 *MOT: hey I'll point at you too .
35 %mor: co|hey pro:sub|I~mod|will v|point prep|at pro|you post|too .
36 %gra: 1|4|COM 2|4|SUBJ 3|4|AUX 4|0|ROOT 5|4|JCT 6|5|POBJ 7|6|PQ 8|4|PUNCT
37 *CHI: &=vocalize .
38 *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?
   - Isee 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→B) = Prob(AB)/Prob(A), where probabilities are estimated by frequencies
   - Word boundaries at points of *local minima*
   - E.g., TP(pre→tty) & TP(ba→by) both > TP(tty→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

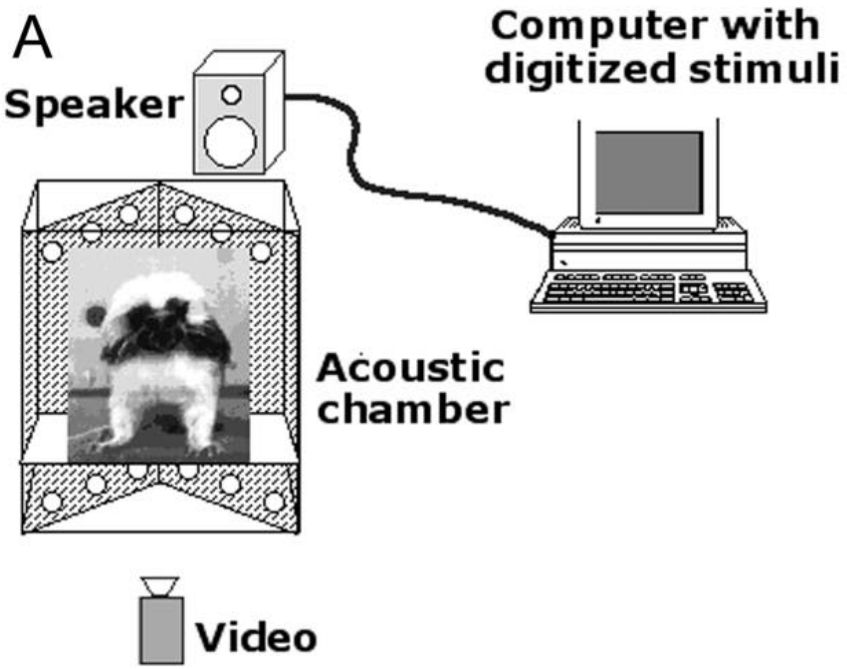
Strategies for learning words

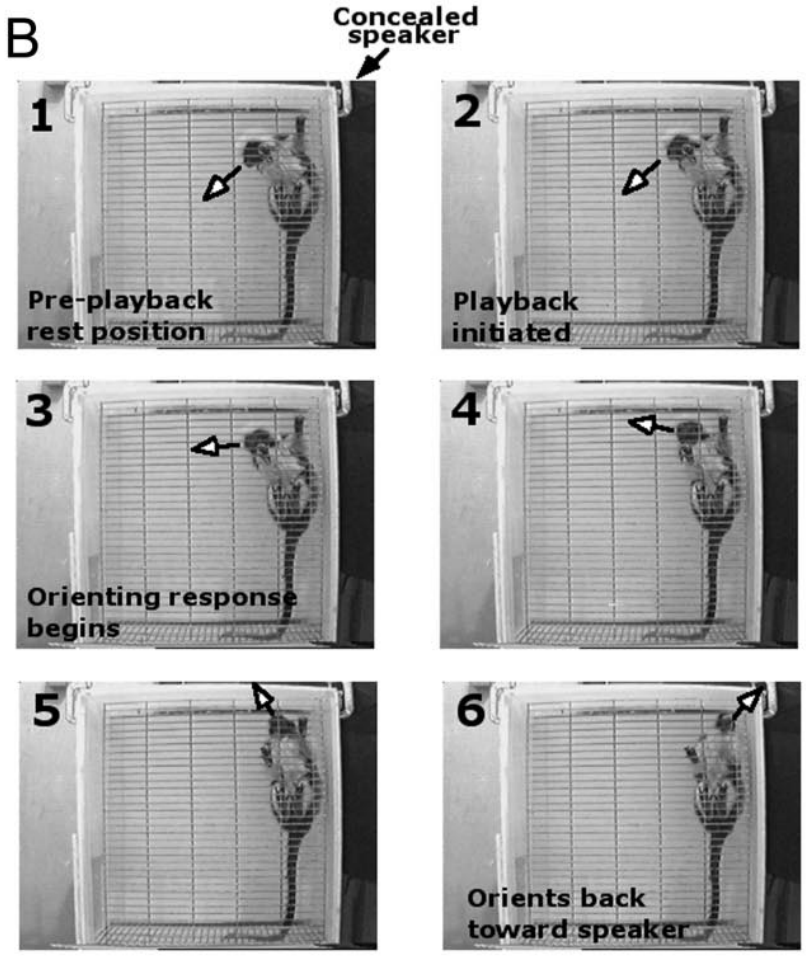
2. Statistical methods, continued:

Evolutionary: probably old? Hauser *et al.* 2001, cotton-top tamarind monkeys

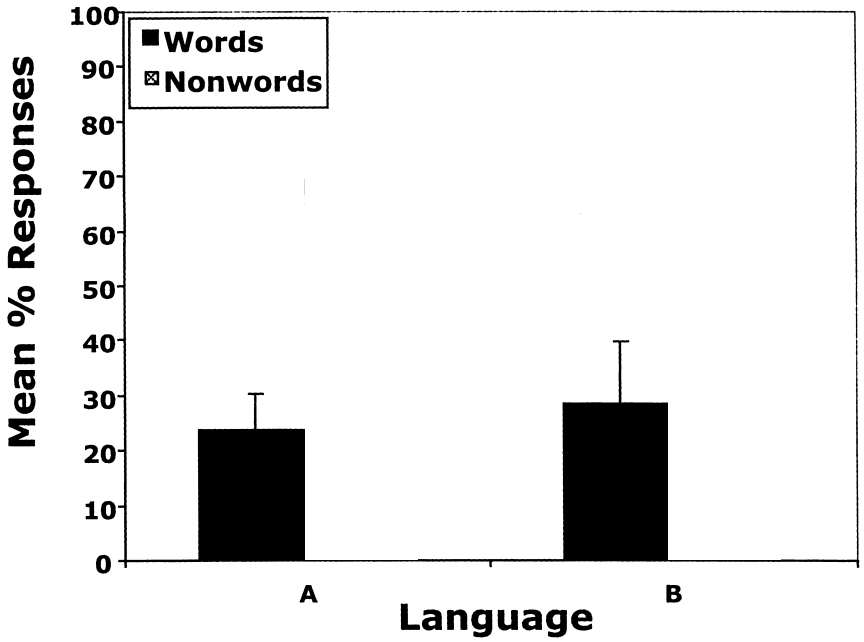
Table 1
Design of Languages A and B and test items comparing words versus non-words or words versus part-words

	Language A	Language B
Words	tupiro, golabu, bidaku, padoti	tudaro, pigola, bikuti, budopa
Test words	tupiro, golabu	tudaro, pigola
Test non-words	dapiku, tilado	tigobu, kudabi
Test part-words	tibida, kupado	pabiku, tibudo


A
Speaker
Computer with digitized stimuli
Acoustic chamber
Video



Hypothesis:
Like babies; orient to novel stimuli


Mean % Responses
100
90
80
70
60
50
40
30
20
10
0
Words
Nonwords
A
B
Language

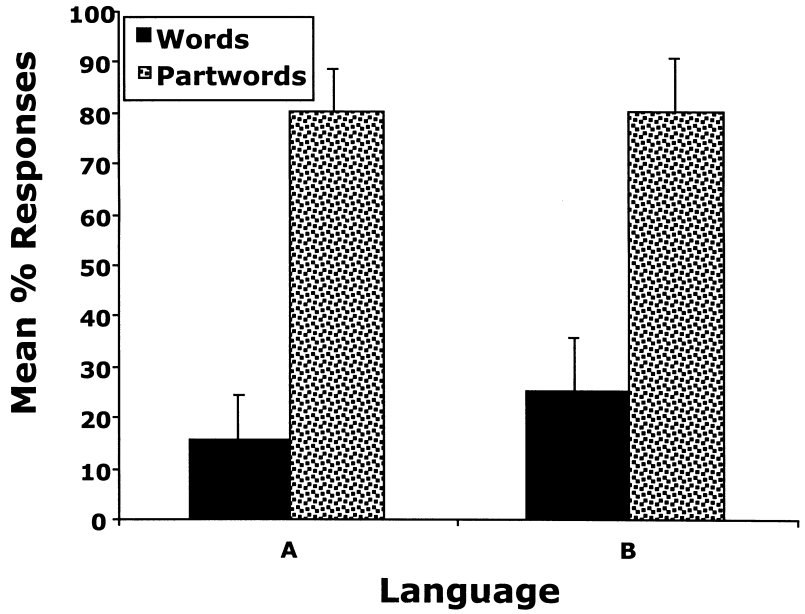


Fig. 3. Mean (standard error) percent of *word* versus *partword* test trials on which subjects responded, for Language A (left) and Language B (right). Black bars indicate responses to *word* trials, stippled bars indicate responses to *partword* trials.

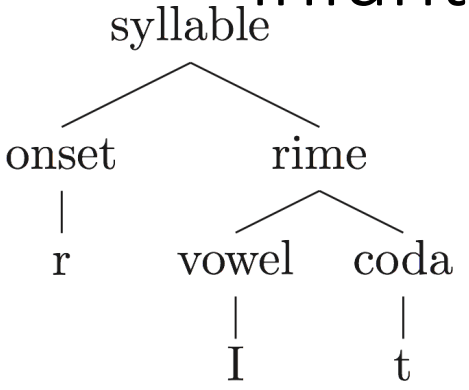
Strategies for learning words

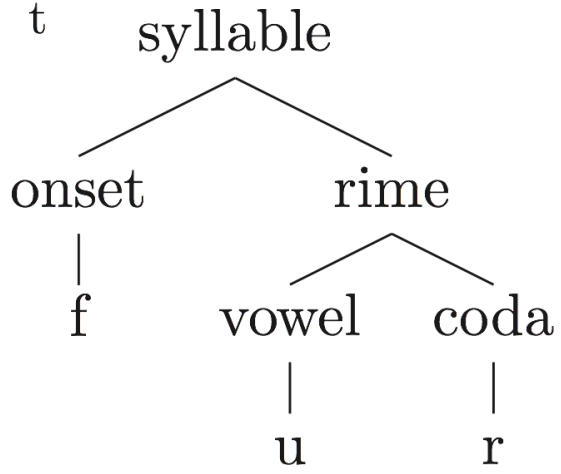
3. Metrical segmentation
   - 90% of English content words (? What's that?) are stress initial in conversational speech (Cutler & Carter, 1987)
   - So maybe stressed syllable = beginning of word
   - Back to crying - Evidence for metrical detection: 7.5 month old babies detect strong-weak pattern in 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?

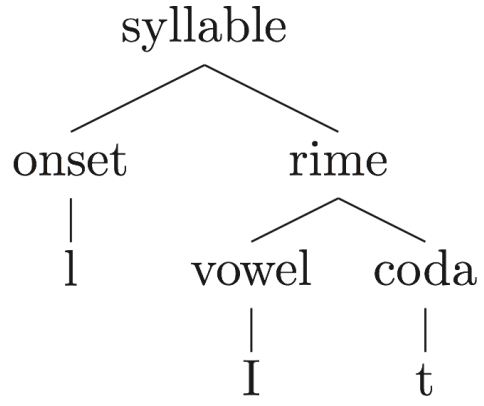
Strategies for learning words

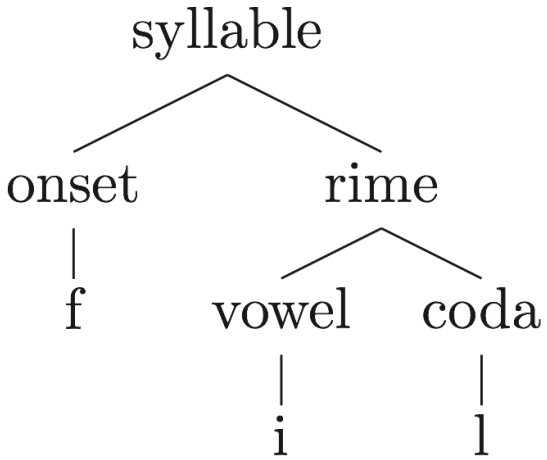
4. Phonotactic constraints
   - What makes a well-formed syllable?
   - Pight, clight, zight vs. flight, dnight, ptight. Which are “possible” English words, which are not?
   - Only certain consonant clusters are valid “onsets” in English (Halle, 1978)
   - Language specific, so must come from experience (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 <u>syllables</u>









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

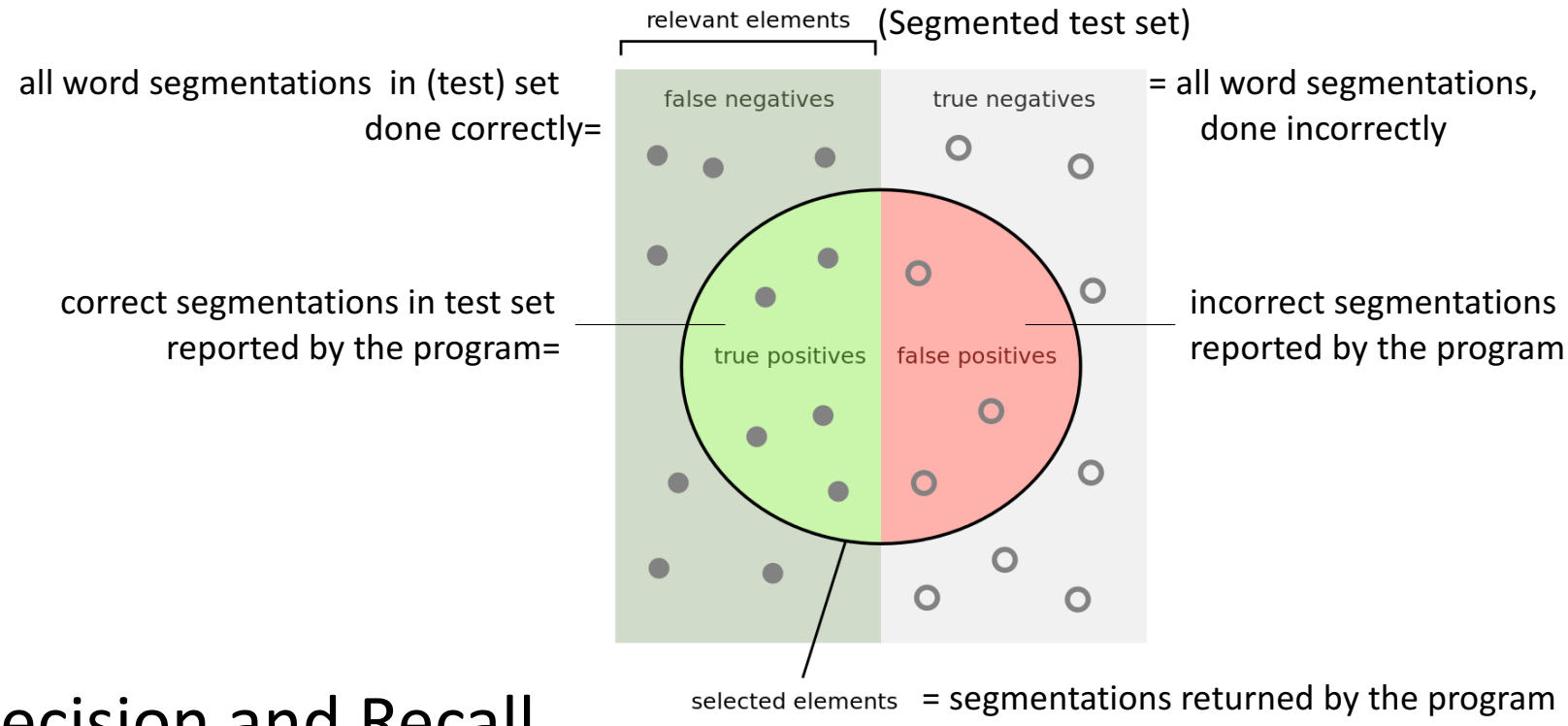
Strategies for learning words

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



Precision and Recall (0-100%)

Precision

How many selected items are relevant?

$$\text{Precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}$$

Recall

How many relevant items are selected?

$$\text{Recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$

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

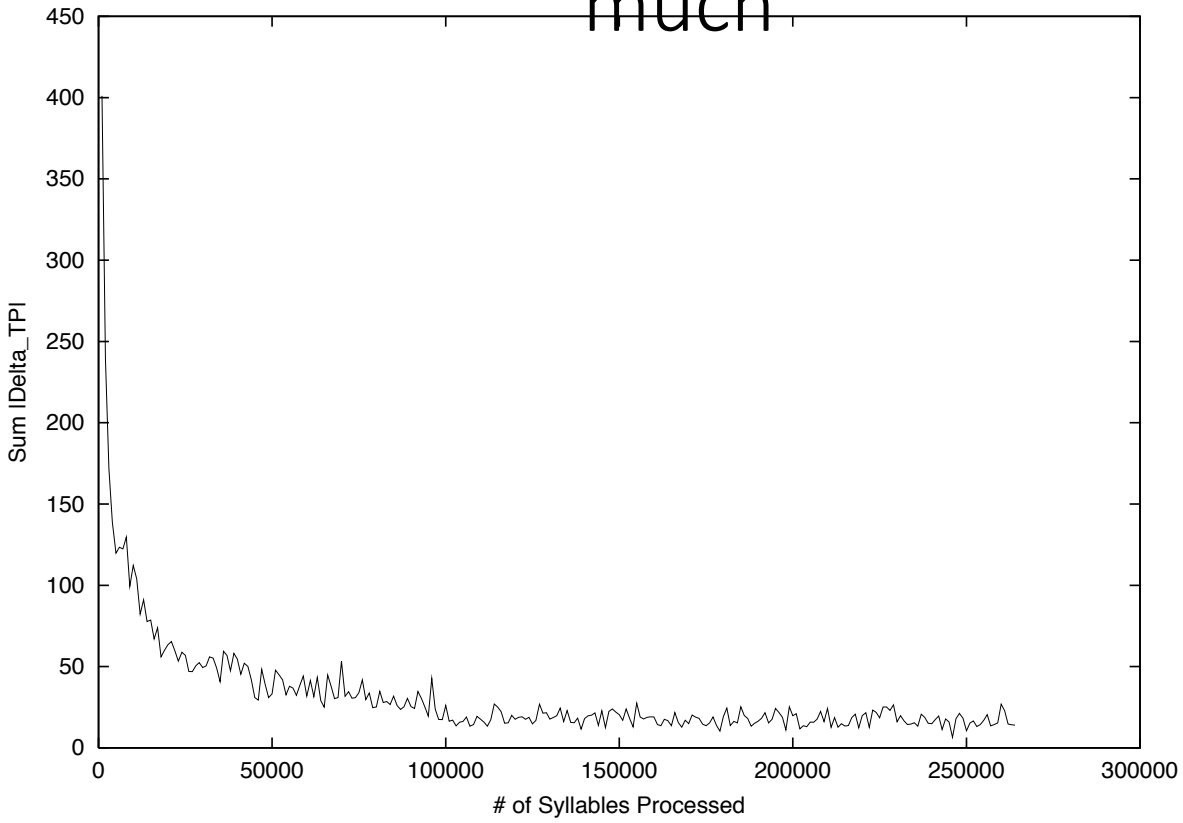
- There is a word boundary between syllables AB and CD

  if TP(A→B) >TP(B→C) < TP(C→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



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
- $W_1S_1S_2S_3W_2 \Rightarrow$ 3 words W_1S_1 S_2 S_3W_2
- Can help statistical learning: $S_1W_1W_2W_3S_2$ 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:

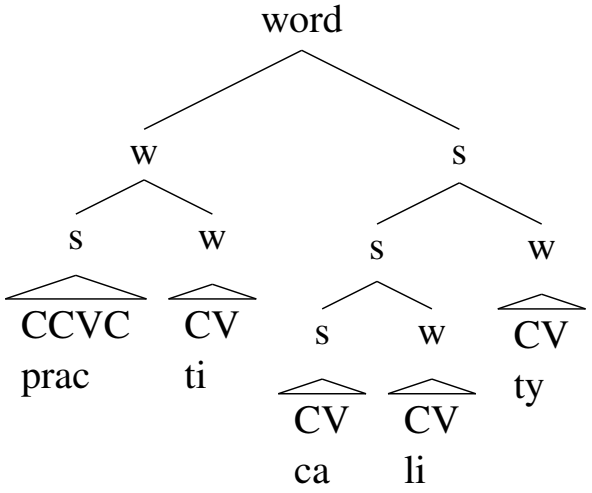
a) Recognize strong vs. weak syllables
b) A collection of reliably segmented words
c) A computation that finds the dominant pattern in the set of words

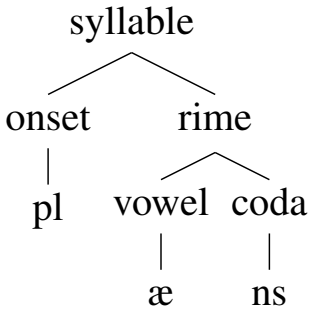
For USC, only (a) required

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

But how do kids pick up stress? We seem to hear it, buthow?

How do children figure out stress for the word segmentation problem?
Consonant-Vowel pattern in babbling: universal

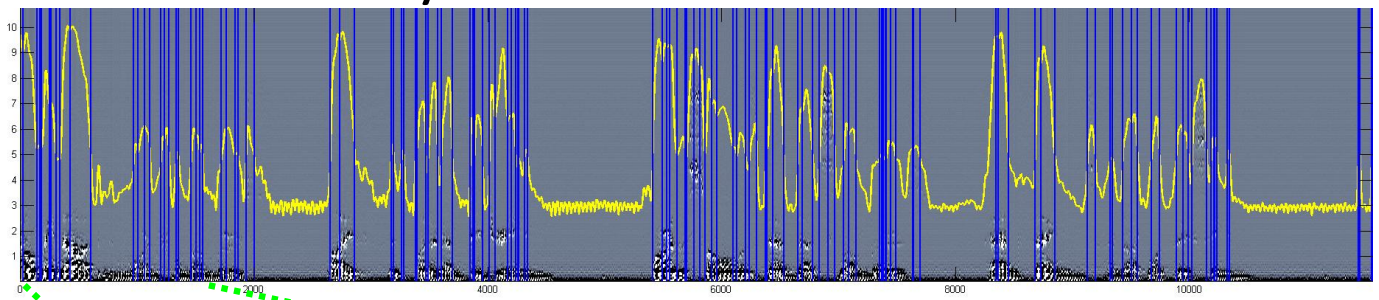




(a) Syllable structure of the word *plans.*

The right representation is <u>combinatoria</u>l

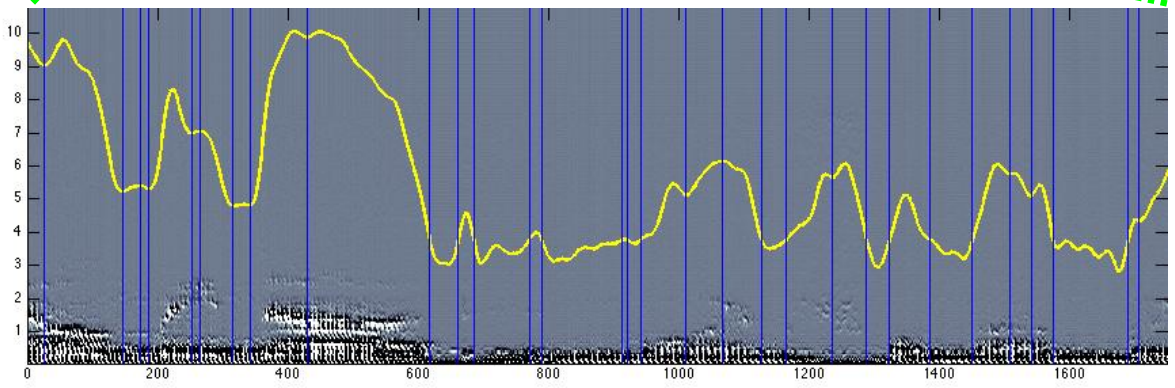
This way? The Beat Generation



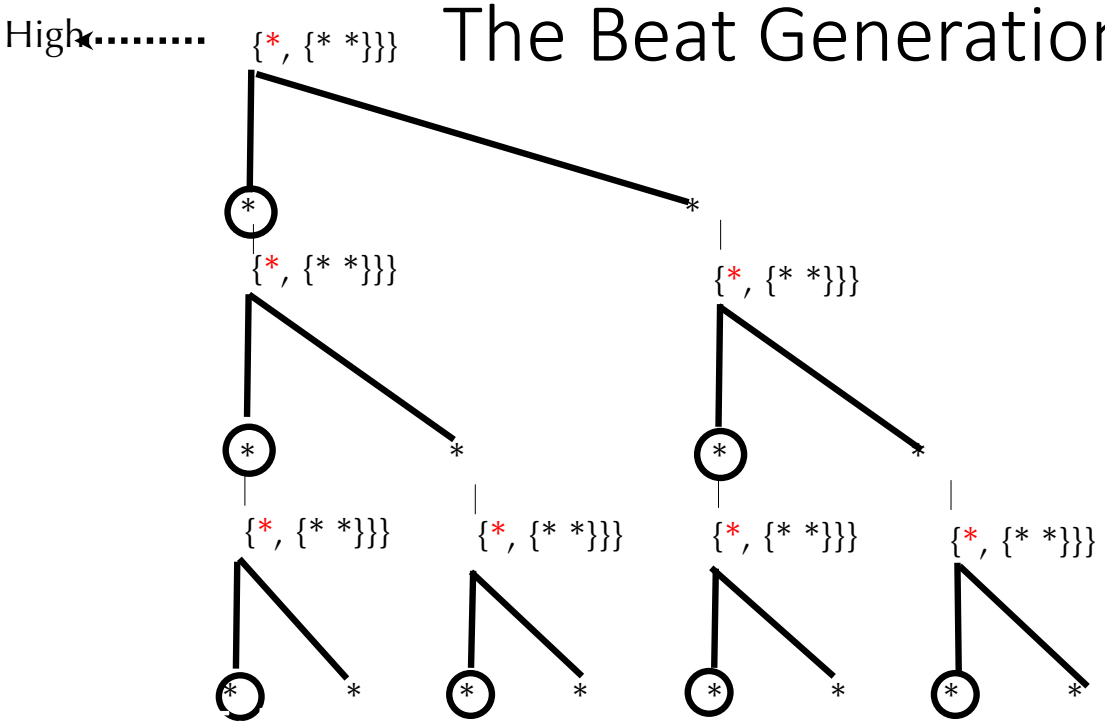
/ ∪ / ∪ / ∪ / ∪

Tell me not in mournful numbers…

Yields beat pattern



The Beat Generation



Iambic: mark <u>left</u> as “head” & project to next level

Suggests: there is an operation that takes <u>two</u> 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 (S_1S_2) , a word boundary is posited in between
   b) If there are more than 1 weak syllable between 2 strong syllables ($S_1W…WS_2$) 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_iW_{i-1} and $W_{j+1,\ l} < j$ 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 <u>two</u> words, one containing S_1 the other one S_2. But in both cases, no word is added to dictionary (learner is unsure)

Results

Model	Precision	Recall	F-measure ($\alpha = 0.5$)
SL	41.6%	23.3%	0.298
SL + USC	73.5%	71.2%	0.723
Algebraic agnostic	85.9%	89.9%	0.879
Algebraic random	95.9%	93.4%	0.946

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