Psych 156A/ Ling 150: Acquisition of Language II

Lecture 7

Speech segmentation II

Announcements

Be working on HW2

Be working on speech segmentation review questions

Midterm on Tuesday, 5/3/16

Computational problem

Divide fluent speech into individual words

tuðəkæsəlbijándðəgáblɪnsíɾi

tu ðə kæsəl bijánd ðə gáblɪn síɾi

to the castle beyond the goblin city

Computational modeling (Working with "digital children")

Computational model: a program that simulates the mental processes occurring in a child's mind (usually implementing a set of mathematical equations that describe those processes). This requires knowing what the input and output are, and then testing the strategies that can take the given input and transform it into the desired output.

Goal: Figure out how the acquisition process works in children.

Computational modeling (Working with "digital children")

Important: We want to make the model match what we know about humans as much as possible.

Why? So that the model can be used to help us understand how humans work. For example, if it matches what we know about the input infants use and the way they use it, it can be used to predict what we expect to see infants doing.

Computational modeling (Working with "digital children")

For example, in speech segmentation, the input could be a sequence of syllables and the desired output is words (i.e., groups of syllables that are useful for various other language things).

Input: "un der stand my po si tion"
Desired Output: "understand my position"

How good is transitional probability on real data?

Gambell & Yang (2006): Computational model goal

Realistic input

Realistic input is important to use since the experimental study of Saffran, Aslin, & Newport (1996) used artificial language data, and it's not clear how well the results they found will map to real language.

How good is transitional probability on real data?

Gambell & Yang (2006): Computational model goal

Psychologically plausible learning algorithm

A psychologically plausible learning algorithm is important since we want to make sure whatever strategy the model uses is something a child could use, too. (Something based on transitional probability would probably work, since Saffran, Aslin, & Newport (1996) showed that infants can track this kind of information in the artificial language.)

How do we measure segmentation performance?

Perfect adult-like segmentation:
identify all the words in the speech stream (*recall*)
only identify syllables groups that are actually words (*precision*)

ðəbígbǽdwʌ́lf

ðə bíg bǽd wʌ́lf
the bɪg bad wolf

How do we measure segmentation performance?

Perfect adult-like segmentation:
identify all the words in the speech stream (*recall*)
only identify syllables groups that are actually words (*precision*)

ðəbígbǽdwʌ́lf

ðə bíg bǽd wʌ́lf
the bɪg bad wolf

Recall calculation:
# of real words found / # of actual words
Identified 4 real words: the, big, bad, wolf
Should have identified 4 words: the, big, bad, wolf
Recall Score: 4 words found/4 should have found = 1.0

How do we measure segmentation performance?

Perfect adult-like segmentation:
identify all the words in the speech stream (*recall*)
only identify syllables groups that are actually words (*precision*)

ðəbígbǽdwʌ́lf

ðə bíg bǽd wʌ́lf
the big bad wolf

Precision calculation:
# of real words found / # of words guessed
Identified 4 real words: the, big, bad, wolf
Identified 4 words total: the, big, bad, wolf
Precision Score: 4 real words found/4 words found= 1.0

How do we measure segmentation performance?

Perfect adult-like segmentation:
identify all the words in the speech stream (*recall*)
only identify syllables groups that are actually words (*precision*)

ðəbígbǽdwʌ́lf

Error

ðəbíg bǽd wʌ́lf
thebig bad wolf

How do we measure segmentation performance?

Perfect adult-like segmentation:
identify all the words in the speech stream (*recall*)
only identify syllables groups that are actually words (*precision*)

ðəbígbǽdwʌ́lf

Error

ðəbíg bǽd wʌ́lf
thebig bad wolf

Recall calculation:
Identified 2 real words: bad, wolf
Should have identified 4 words: the, big, bad, wolf
Recall Score: 2 real words found/4 should have found = 0.5

How do we measure segmentation performance?

Perfect adult-like segmentation:
identify all the words in the speech stream (*recall*)
only identify syllables groups that are actually words (*precision*)

ðəbígbǽdwʌ́lf

Error

ðəbíg bǽd wʌ́lf
thebig bad wolf

Precision calculation:
Identified 2 real words: bad, wolf
Identified 3 words total: thebig, bad, wolf
Precision Score: 2 real words/3 words identified = 0.666…

How do we measure segmentation performance?

Perfect adult-like segmentation:
identify all the words in the speech stream (*recall*)
only identify syllables groups that are actually words (*precision*)

Want good enough scores on both of these measures
in order to be sure that segmentation is really working

One score that combines precision and recall: F-score
- This is the harmonic mean of precision and recall

$$F-score = 2 * \frac{recall * precision}{recall + precision}$$

How do we measure segmentation performance?

Perfect adult-like segmentation:
identify all the words in the speech stream (*recall*)
only identify syllables groups that are actually words (*precision*)

Perfect segmentation

Recall = 100% (1.0)
Precision = 100% (1.0)
F-score = 2*(1.0 * 1.0)/(1.0 + 1.0) = 1.0

$$F-score = 2 * \frac{recall * precision}{recall + precision}$$

How do we measure segmentation performance?

Perfect adult-like segmentation:
identify all the words in the speech stream (*recall*)
only identify syllables groups that are actually words (*precision*)

Not-so-perfect segmentation

Recall = 50% (0.50)
Precision = 67% (0.67)
F-score = 2*(0.50 * 0.67)/(0.50 + 0.67) = 0.57

$$F-score = 2 * \frac{recall * precision}{recall + precision}$$

Where does the realistic data come from?

CHILDES

Child Language Data Exchange System
http://childes.psy.cmu.edu/

Large collection of child-directed speech data (usually parents interacting with their children) transcribed by researchers. Used to see what children's input is actually like.

CHILDES Child Language Data Exchange System

Where does the realistic data come from?

Gambell & Yang (2006)

Looked at Brown corpus files in CHILDES (226,178 words made up of 263,660 syllables).

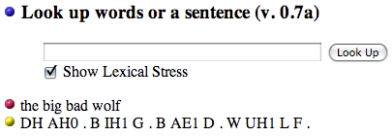
Converted the transcriptions to pronunciations using a pronunciation dictionary called the CMU Pronouncing Dictionary.

http://www.speech.cs.cmu.edu/cgi-bin/cmudict

The CMU Pronouncing Dictionary

Where does the realistic data come from?

Converting transcriptions to pronunciations



Gambell and Yang (2006) tried to see if a model learning from transitional probabilities between syllables could correctly segment words from realistic data.

the	big	bad	wolf
DH AH0 .	B IH1 G .	B AE1 D .	W UH1 L F .
ðə	bíg	bǽd	wʌ́lf

Segmenting realistic data

Gambell and Yang (2006) tried to see if a model learning from transitional probabilities between syllables could correctly segment words from realistic data.

ðə	bíg	bǽd	wʌ́lf
DH AH0	B IH1 G	B AE1 D	W UH1 L F

Specific strategy implemented:
Place a boundary at a **transitional probability minimum**.

"There is a word boundary AB and CD if
TrProb(A --> B) > TrProb(B-->C) < TrProb(C --> D)."

Segmenting realistic data

Gambell and Yang (2006) tried to see if a model learning from transitional probabilities between syllables could correctly segment words from realistic data.

Desired segmentation

ðə	bíg	bǽd	wʌ́lf
DH AH0	B IH1 G	B AE1 D	W UH1 L F
the	big	bad	wolf

Modeling results for transitional probability

Precision: 41.6%

Recall: 23.3%

F-score: 29.9%

A learner relying only on transitional probability does not reliably segment words such as those in child-directed English.

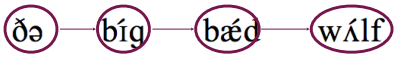
About 60% of the words posited by the transitional probability learner are not actually words (41.6% precision) and almost 80% of the actual words are not identified (23.3% recall).

Why such poor performance?

"We were surprised by the low level of performance. Upon close examination of the learning data, however, it is not difficult to understand the reason….a sequence of monosyllabic words requires a word boundary after each syllable; a [transitional probability] learner, on the other hand, will only place a word boundary between two sequences of syllables for which the [transitional probabilities] within [those sequences] are higher than [those of surrounding the sequences]…" - Gambell & Yang (2006)

Why such poor performance?

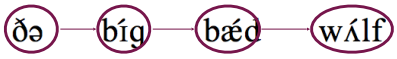
"We were surprised by the low level of performance. Upon close examination of the learning data, however, it is not difficult to understand the reason….a sequence of monosyllabic words requires a word boundary after each syllable; a [transitional probability] learner, on the other hand, will only place a word boundary between two sequences of syllables for which the [transitional probabilities] within [those sequences] are higher than [those of surrounding the sequences]…" - Gambell & Yang (2006)



TrProb1 TrProb2 TrProb3

Why such poor performance?

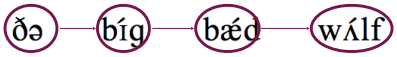
"We were surprised by the low level of performance. Upon close examination of the learning data, however, it is not difficult to understand the reason….a sequence of monosyllabic words requires a word boundary after each syllable; a [transitional probability] learner, on the other hand, will only place a word boundary between two sequences of syllables for which the [transitional probabilities] within [those sequences] are higher than [those of surrounding the sequences]…" - Gambell & Yang (2006)



0.6 0.3 0.7

Why such poor performance?

"We were surprised by the low level of performance. Upon close examination of the learning data, however, it is not difficult to understand the reason….a sequence of monosyllabic words requires a word boundary after each syllable; a [transitional probability] learner, on the other hand, will only place a word boundary between two sequences of syllables for which the [transitional probabilities] within [those sequences] are higher than [those of surrounding the sequences]…" - Gambell & Yang (2006)



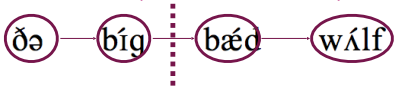
0.6 0.3 0.7

0.6 > 0.3 < 0.7

Why such poor performance?

"We were surprised by the low level of performance. Upon close examination of the learning data, however, it is not difficult to understand the reason….a sequence of monosyllabic words requires a word boundary after each syllable; a [transitional probability] learner, on the other hand, will only place a word boundary between two sequences of syllables for which the [transitional probabilities] within [those sequences] are higher than [those of surrounding the sequences]…" - Gambell & Yang (2006)

learner posits one word boundary at minimum TrProb



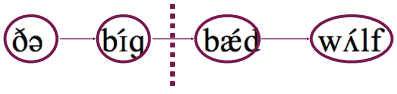
0.6 0.3 0.7

0.6 > 0.3, 0.3 < 0.7

Why such poor performance?

"We were surprised by the low level of performance. Upon close examination of the learning data, however, it is not difficult to understand the reason….a sequence of monosyllabic words requires a word boundary after each syllable; a [transitional probability] learner, on the other hand, will only place a word boundary between two sequences of syllables for which the [transitional probabilities] within [those sequences] are higher than [those of surrounding the sequences]…" - Gambell & Yang (2006)

…but nowhere else



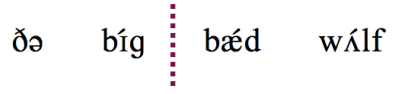
0.6 0.3 0.7

0.6 > 0.3, 0.3 < 0.7

Why such poor performance?

"We were surprised by the low level of performance. Upon close examination of the learning data, however, it is not difficult to understand the reason….a sequence of monosyllabic words requires a word boundary after each syllable; a [transitional probability] learner, on the other hand, will only place a word boundary between two sequences of syllables for which the [transitional probabilities] within [those sequences] are higher than [those of surrounding the sequences]…" - Gambell & Yang (2006)

…but nowhere else



Why such poor performance?

"We were surprised by the low level of performance. Upon close examination of the learning data, however, it is not difficult to understand the reason….a sequence of monosyllabic words requires a word boundary after each syllable; a [transitional probability] learner, on the other hand, will only place a word boundary between two sequences of syllables for which the [transitional probabilities] within [those sequences] are higher than [those of surrounding the sequences]..." - Gambell & Yang (2006)

…but nowhere else

ðəbíg	bædwʌ́lf
thebig	badwolf

Precision for this sequence: 0 words correct out of 2 found
Recall: 0 words correct out of 4 that should have been found

Why such poor performance?

"More specifically, a monosyllabic word is followed by another monosyllabic word 85% of the time. As long as this is the case, [this kind of transitional probability learner] cannot work." - Gambell & Yang (2006)

Additional learning bias

Gambell & Yang (2006) idea
Children are sensitive to the properties of their native language like stress patterns very early on. Maybe they can use those sensitivities to help them solve the segmentation problem.

Hypothesis: Unique Stress Constraint (USC)
Children think a word can bear at most one primary stress.

no stress	stress	stress	stress
ðə	bíg	bǽd	wʌ́lf
the	big	bad	wolf

Additional learning bias

Gambell & Yang (2006) idea
Children are sensitive to the properties of their native language like stress patterns very early on. Maybe they can use those sensitivities to help them solve the segmentation problem.

Hypothesis: Unique Stress Constraint (USC)
Children think a word can bear at most one primary stress.

ðə	bíg	bǽd	wʌ́lf
the	big	bad	wolf

Learner gains knowledge: These must be separate words

Additional learning bias

Gambell & Yang (2006) idea
Children are sensitive to the properties of their native language like stress patterns very early on. Maybe they can use those sensitivities to help them solve the segmentation problem.

Hypothesis: Unique Stress Constraint (USC)
Children think a word can bear at most one primary stress.

húz	ə	fréd	əv	ðə	bíg	bǽd	wʌ́lf
who's	a	fraid	of	the	big	bad	wolf

Get these boundaries because stressed (strong) syllables are next to each other.

Additional learning bias

Gambell & Yang (2006) idea
Children are sensitive to the properties of their native language like stress patterns very early on. Maybe they can use those sensitivities to help them solve the segmentation problem.

Hypothesis: Unique Stress Constraint (USC)
Children think a word can bear at most one primary stress.

húz	ə	fréd	əv	ðə	bíg	bǽd	wʌ́lf
who's	a	fraid	of	the	big	bad	wolf

Can use this in tandem with transitional probabilities when there are weak (unstressed) syllables between stressed syllables.

Additional learning bias

Gambell & Yang (2006) idea

Children are sensitive to the properties of their native language like stress patterns very early on. Maybe they can use those sensitivities to help them solve the segmentation problem.

Hypothesis: Unique Stress Constraint (USC)

Children think a word can bear at most one primary stress.

? ?

húz ə fréd əv ðə bíg bæd wʌ́lf

who's a fraid of the big bad wolf

There's a word boundary at one of these two.

USC + Transitional Probabilities

Precision: 73.5%

Recall: 71.2%

F-score: 72.3%

A learner relying on transitional probability but who also has knowledge of the Unique Stress Constraint does a much better job at segmenting words such as those in child-directed English.

Only about 25% of the words posited by the transitional probability learner are not actually words (73.5% precision) and about 30% of the actual words are not extracted (71.2% recall).

Another strategy

Using words you recognize to help you figure out words you don't recognize (a more formal version of the "familiar words" strategy)

Another strategy: Algebraic learning

Algebraic learning (Gambell & Yang 2003)

Subtraction process of figuring out unknown words.

"Look, honey - it's a big goblin!"

bíggáblɪn

bíg: big (familiar word)

bíggáblɪn

bíg

gáblɪn = (new word)

Evidence of algebraic learning in children

"Behave yourself!"
"I was have!"
(be-have = be + have)

"Was there an adult there?"
"No, there were two dults."
(a-dult = a + dult)

"Did she have the hiccups?"
"Yeah, she was hiccing-up."
(hicc-up = hicc + up)

Experimental evidence of algebraic learning

Experimental studies show young infants can use familiar words to segment novel words from their language

- Bortfeld, Morgan, Golinkoff, & Rathbun 2005:
  6-month-old English infants use their own name or *Mommy/Mama*
- Shi, Werker, & Cutler 2006
  11-month-old English infants use English articles like *her, its,* and *the*
- Shi, Cutler, Werker, & Cruickshank 2006
  11-month-old English infants (but not 8-month-old English infants) use the English article *the*

Experimental evidence of algebraic learning

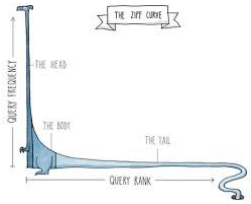
Experimental studies show young infants can use familiar words to segment novel words from their language

- Hallé, Durand, Bardies, & de Boysson 2008
  11-month-old French infants use French articles like *le, les,* and *la*

- Mersad & Nazzi 2012
  8-month-old French infants can use words like *mamã* to segment words in an artificial language

Computational support for algebraic learning

Kurumada, Meylan, & Frank (2013) discovered that the Zipfian nature of natural language data is much more beneficial to a segmentation strategy that looks for coherent chunks (like an algebraic learning strategy would).



Using algebraic learning + USC

WeakSyl	StrongSyl	StrongSyl	StrongSyl
the	big	bad	wolf
ðə	bíg	bæd	wʌ́lf

"the big bad wolf"

Using algebraic learning + USC

Familiar word: "the" (algebraic learning)

WeakSyl	StrongSyl	StrongSyl	StrongSyl
the	big	bad	wolf
ðə	bíg	bæd	wʌ́lf

"the big bad wolf"

Using algebraic learning + USC

USC says these must be separate words

WeakSyl	StrongSyl	StrongSyl	StrongSyl
the	big	bad	wolf
ðə	bíg	bæd	wʌ́lf

"the big bad wolf"

Using algebraic learning + USC

Correct segmentation!

WeakSyl	StrongSyl	StrongSyl	StrongSyl
the	big	bad	wolf
ðə	bíg	bæd	wʌ́lf

"the big bad wolf"

Algebraic learning + USC

Precision: 95.9%

Recall: 93.4%

F-score: 94.6%

A learner relying on algebraic learning and who also has knowledge of the Unique Stress Constraint does a really great job at segmenting words such as those in child-directed English - even better than one relying on the transitional probability between syllables.

Only about 5% of the words posited by the transitional probability learner are not actually words (95.9% precision) and about 7% of the actual words are not extracted (93.4% recall).

Gambell & Yang 2006 summary

Using a simple learning strategy involving transitional probabilities doesn't work so well on realistic data, even though experimental research suggests that infants are capable of tracking and learning from this information.

Models of children that have additional knowledge about the stress patterns of words seem to have a much better chance of succeeding at segmentation if they learn via a simple transitional-probability-based strategy.

However, models of children that use algebraic learning and have additional knowledge about the stress patterns of words perform even better at segmentation than any of the models using a simple transitional probability strategy.

Gambell & Yang 2006 critiques

Do infants have access to the Unique Stress Constraint (USC)?

- Children definitely use transitional probabilities & algebraic learning – but how precise is their knowledge of lexical stress?

Skoruppa, Pons, Bosch, Christophe, Cabrol, & Peperkamp 2012:
6-month-old Spanish and French infants don't appear to even recognize the difference between words with initial vs. final lexical stress unless the word forms are identical. (No generalization of lexical stress patterns for words.)

✗ píma vs. latú

✓ píma vs. pimá

Gambell & Yang 2006 critiques

Do infants have access to the Unique Stress Constraint (USC)?

However, Börschinger & Johnson (2014) demonstrated how a very sophisticated statistical learner (a learner with some idea about how languages are organized) can quickly learn that the Unique Stress Constraint exists at the same time it's learning how to segment words out of fluent speech in English.

Gambell & Yang 2006 critiques

Does dictionary stress really match actual stress patterns?

Gambell & Yang estimate: the *bíg bád wólf*

Typical speech: the big bad *wólf*

It's unclear how well this algorithm works with real stress patterns in fluent speech…

More sophisticated learning strategies

What if children are capable of tracking more sophisticated distributional information (that is, they're not just restricted to transitional probability minima)? In that case, how well do they do on realistic data, if all they're using is statistical learning (no stress information)?

Bayesian inference

What if children can use Bayesian inference?
Human cognitive behavior is consistent with this kind of reasoning.
(Tenenbaum & Griffiths 2001, Griffiths & Tenenbaum 2005,
Xu & Tenenbaum 2007)

Bayesian inference is a sophisticated kind of probabilistic reasoning that tries to find hypotheses that

(1) are consistent with the observed data
(2) conform to a child's prior expectations

Bayesian inference for word segmentation

What kind of hypotheses might a child have for segmentation?

Observed data:
"to the ca stle be yond the go blin ci ty"

Hypothesis = sequence of lexical items producing this observable data

Some sample hypotheses

Hypothesis 1:
"tothe castle beyond thegoblin city"
Items: *tothe, castle, beyond, thegoblin, city*

Hypothesis 2:
"to the castle beyond the goblin city"
Items: *to, the, castle, beyond, goblin, city*
Note: the *is used twice*

Bayesian model

Learner expectations about segmentation:

(1) Words tend to be shorter rather than longer
(2) Vocabulary tends to be small rather than large

Used by these research studies (among others):
Goldwater, Griffiths, & Johnson 2009
Pearl, Goldwater, & Steyvers 2011
Phillips & Pearl 2012, 2014a, 2014b, 2015a, 2015b

Bayesian model

Learner expectations about segmentation:

(1) Words tend to be shorter rather than longer
(2) Vocabulary tends to be small rather than large

How would a Bayesian learner with these kind of expectations decide between the two hypotheses from before?

Hypothesis 1:
"tothe castle beyond thegoblin city"
Items: *tothe, castle, beyond, thegoblin, city*

How long are words? Between 2 and 3 syllables, average = 2.2
How large is the vocabulary? 5 words

Bayesian model

Learner expectations about segmentation:

(1) Words tend to be shorter rather than longer
(2) Vocabulary tends to be small rather than large

How would a Bayesian learner with these kind of expectations decide between the two hypotheses from before?

Hypothesis 2:
"to the castle beyond the goblin city"
Items: *to, the, castle, beyond, goblin, city*

How long are words? Between 1 and 2 syllables, average = 1.7
How large is the vocabulary? 6 words

Bayesian model

Comparing hypotheses - which is most likely?

Hypothesis 1: longer words, but fewer words
How long are words? Avg = 2.2 syllables
How large is the vocabulary? 5 words

Hypothesis 2: shorter words, but more words
How long are words? Avg = 1.7 syllables
How large is the vocabulary? 6 words

A Bayesian learner makes a decision based on how important each of its expectations is (in this case, it's a balance of the two constraints as determined by the mathematical implementation of the Bayesian startegy: fewer words vs. shorter words).

Bayesian model

Comparing hypotheses - which is most likely?

Hypothesis 1: longer words, but fewer words
How long are words? Avg = 2.2 syllables
How large is the vocabulary? 5 words

Hypothesis 2: shorter words, but more words
How long are words? Avg = 1.7 syllables
How large is the vocabulary? 6 words

There will be some probability the Bayesian learner assigns to each hypothesis. The most probable hypothesis will be the one the learner chooses.

Bayesian model

Comparing hypotheses - which is most likely?

Hypothesis 1: longer words, but fewer words
How long are words? Avg = 2.2 syllables
How large is the vocabulary? 5 words

Probability: 0.33

Hypothesis 2: shorter words, but more words
How long are words? Avg = 1.7 syllables
How large is the vocabulary? 6 words

Probability: 0.67

There will be some probability the Bayesian learner assigns to each hypothesis. The most probable hypothesis will be the one the learner chooses.

Bayesian model

Comparing hypotheses - which is most likely?

Hypothesis 1: longer words, but fewer words
How long are words? Avg = 2.2 syllables
How large is the vocabulary? 5 words

Probability: 0.33

Hypothesis 2: shorter words, but more words
How long are words? Avg = 1.7 syllables
How large is the vocabulary? 6 words

Probability: 0.67

There will be some probability the Bayesian learner assigns to each hypothesis. The most probable hypothesis will be the one the learner chooses.

Realistic Bayesian learners

Phillips and Pearl 2012, 2015a tested their Bayesian learners on realistic input: 28,391 utterances of child-directed speech from the Brent corpus in CHILDES. (Average utterance length: 3.4 words and 4.2 syllables)

Best performance by a Bayesian learner:

F-score: 86.3%

This is much better than what we found for a learner that hypothesizes a word boundary at a transitional probability minimum (F-score = 29.9%). Statistical learning by itself isn't always so bad after all!

Realistic Bayesian learners

Phillips and Pearl 2014a, 2014b tested these same Bayesian learners on realistic input from seven different languages: English, German, Spanish, Italian, Farsi, Hungarian, and Japanese.

Best performance by a Bayesian learner, averaged across languages:

F-score: 69.8%

This is still much better than what we found for a learner that hypothesizes a word boundary at a transitional probability minimum (F-score = 29.9%). Statistical learning by itself isn't always so bad after all!

More realistic segmentation output

Important point: What a seven-month-old thinks are useful units to segment out of fluent speech may not match what we adults think of as words.

Example:
"See the kitty playing with the string."

Useful unit smaller than a word:
-ing = ongoing action
Oversegmentation (split words up):
playing = play ing

Useful unit larger than a word:
thekitty = maps to specific concrete object
Undersegmentation (squish words together):
the kitty = thekitty

More realistic segmentation output

Important point: What a seven-month-old thinks are useful units to segment out of fluent speech may not match what we adults think of as words.

When we count these "useful units" as reasonable segmentation output for a seven-month-old, both Bayesian learners and algebraic learners that incorporate some statistical learning do well cross-linguistically (Phillips & Pearl 2014b, Phillips & Pearl under rev).

Bayesian learner average F-score: 77.5%
Algebraic learner (Lignos 2012) F-score: 71.6%

This is again much better than what we found for a learner that hypothesizes a word boundary at a transitional probability minimum (F-score = 29.9%).
Statistical learning by itself isn't always so bad after all - especially if we recognize that different kinds of output may be useful to a young infant.

Statistical learning for segmentation

Saffran et al. (1996) found that human infants are capable of tracking transitional probability between syllables and using that information to accomplish word segmentation in an artificial language.

Gambell & Yang (2006) found that this same statistical learning strategy (positing word boundaries at transitional probability minima) failed on realistic child-directed speech data.

Statistical learning for segmentation

More recent studies (Goldwater et al. 2009, Pearl et al. 2011, Phillips & Pearl 2012, 2014a, 2014b, 2015a, 2015b) found that more sophisticated statistical learning -- Bayesian inference -- did much better on realistic child-directed speech data, suggesting that children may be able to use statistical learning to help them with segmentation - even before they use other strategies like lexical stress.

Notably, both Bayesian inference and algebraic learning strategies can work for learning to segment a variety of languages, especially if we recognize that an infant's segmentation may not perfectly match an adult's segmentation (Phillips & Pearl 2014a, Phillips & Pearl in rev).

Questions?

You should be able to do up through question 6 on HW2 and all of the speech segmentation review questions.