## Psych 156A/ Ling 150: Psychology of Language Learning

Lecture 9
Words in Fluent Speech II

## Announcements

Homework 3 due today

Homework 2 returned (Avg: 21.6 out of 27)

Quiz 3 returned (Avg: 8.6 out of 10)

Comments about how to do well in this class

## Computational Problem

Divide spoken speech into words

húwzəfréjdəvðəbÍgbæ‘dwə‘lf

## Computational Problem

Divide spoken speech into words

húwzəfréjdəvðəbÍgbæ‘dwə‘lf

↓

húwz əfréjd əv ðə bÍg bæ‘d wə‘lf
who‘s afraid of the big bad wolf

## Saffran, Aslin, & Newport (1996)

Experimental evidence suggests that 8 month old infants can track statistical information such as the transitional probability between syllables. This can help them solve the task of word segmentation.

Evidence comes from testing children in an artificial language paradigm, with very short exposure time.

## Computational Modeling Data
## (Digital Children)

## How good is transitional probability on real data?

Gambell & Yang (2006): Computational model goal

Real data, Psychologically plausible learning algorithm

Realistic data is important to use since the experimental study of Saffran, Aslin, & Newport (1996) used artificial language data

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. (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 word segmentation performance?

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

ðəbÍgbæ‘dwə‘lf
↓
ðə bÍg bæ‘d wə‘lf
the big bad wolf

## How do we measure word segmentation performance?

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

ðəbÍgbæ‘dwə‘lf
↓
ðə bÍg bæ‘d wə‘lf
the big bad wolf

Recall calculation:
Should have identified 4 words: the, big, bad, wolf
Identified 4 real words: the, big, bad, wolf
Recall Score: 4/4 = 1.0

## How do we measure word segmentation performance?

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

ðəbÍgbæ‘dwə‘lf

↓

ðə bÍg bæ‘d wə‘lf

the big bad wolf

Precision calculation:
Identified 4 words: the, big, bad, wolf
Identified 4 real words: the, big, bad, wolf
Precision Score: 4/4 = 1.0

## How do we measure word segmentation performance?

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

ðəbÍgbæ‘dwə‘lf

↓

Error

ðəbÍg bæ‘d wə‘lf

thebig bad wolf

## How do we measure word segmentation performance?

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

ðəbÍgbæ‘dwə‘lf

↓

Error

ðəbÍg bæ‘d wə‘lf

thebig bad wolf

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

## How do we measure word segmentation performance?

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

ðəbÍgbæ‘dwə‘lf

Error

| | | |
|---|---|---|
| ðəbÍg | bæ‘d | wə‘lf |
| thebig | bad | wolf |

Precision calculation:
- Identified 3 words: thebig, bad, wolf
- Identified 2 real words: big, bad

Precision Score: 2/3 = 0.666…

## How do we measure word segmentation performance?

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

Want good scores on both of these measures

## 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 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).

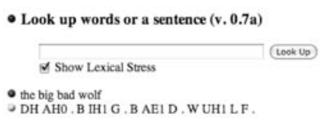
Converted the transcriptions to pronunciations using a pronunciation dictionary called the CMU Pronouncing Dictionary.

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.

ðə bɪ́g bæ‘d wə‘lf

DH AH0 . B IH1 G . B AE1 D . W UH1 L F .

## 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 bɪ́g bæ‘d wə‘lf

DH AH0 . B IH1 G . B AE1 D . W UH1 L F .

## 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 bæ‘d wə‘lf

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%

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

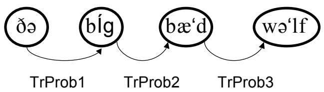
About 60% of the words posited by the transitional probability learner are not actually words (41.6% precision) and almost 80% of the actual words are not extracted (23.3 % recall).

## 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 surrounding the sequences]...” - Gambell & Yang (2006)

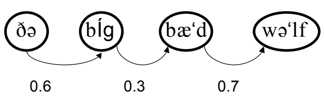
## Why such poor performance?

"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 surrounding the sequences]..." - Gambell & Yang (2006)



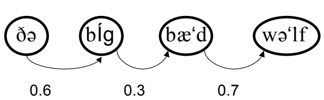

## Why such poor performance?

"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 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 surrounding the sequences]..." - Gambell & Yang (2006)

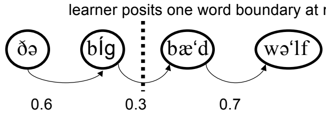



$0.6 > 0.3$, $0.3 < 0.7$

## 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 surrounding the sequences]..." - Gambell & Yang (2006)

learner posits one word boundary at minimum TrProb

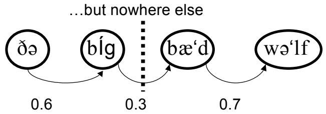



0.6 > 0.3, 0.3 < 0.7

## 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 surrounding the sequences]..." - Gambell & Yang (2006)

…but nowhere else

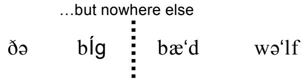



0.6 > 0.3, 0.3 < 0.7

## 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 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 surrounding the sequences]..." - Gambell & Yang (2006)

…but nowhere else

ðəbíg ⋮ bæ‘dwə‘lf

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

## 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, [a 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 word segmentation problem.

Unique Stress Constraint (USC)
A word can bear at most one primary stress.

| no stress | stress | stress | stress |
| --- | --- | --- | --- |
| ðə | bíg | bæ‘d | wə‘lf |

## 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 word segmentation problem.

Unique Stress Constraint (USC)
A word can bear at most one primary stress.

ðə (bɪ́g) ⋮ (bæ‘d) ⋮ (wə‘lf)

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 word segmentation problem.

Unique Stress Constraint (USC)
A word can bear at most one primary stress.

(húw) zə (fréjd) əv ðə (bɪ́g) ⋮ (bæ‘d) ⋮ (wə‘lf)

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 word segmentation problem.

Unique Stress Constraint (USC)
A word can bear at most one primary stress.

(húw) zə (fréjd) əv ðə (bɪ́g) ⋮ (bæ‘d) ⋮ (wə‘lf)

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 word segmentation problem.

Unique Stress Constraint (USC)
A word can bear at most one primary stress.

? ?
húw zə fréjd əv ðə bÍg bæ‘d wə‘lf

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

## USC + Transitional Probabilities

Precision: 73.5%

Recall: 71.2%

A learner relying only 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.

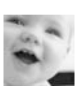
Only about 25% of the words posited by the transitional probability learner are not actually words (73.5% precision) and about 30% of the actual words are not extracted (71.2 % recall).

## Another Strategy

Algebraic Learning (Gambell & Yang (2003))

Subtraction process of figuring out unknown words.

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



bÍg = big (familiar word)

bÍggáblIn
bÍg
―――――
gáblIn = (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)

## Using Algebraic Learning + USC

| StrongSyl | WeakSyl1 | WeakSyl2 | StrongSyl |
|---|---|---|---|
| ma | ny | can | come |

"Many can come…"

## Using Algebraic Learning + USC

Familiar word: "many"

| StrongSyl | WeakSyl1 | WeakSyl2 | StrongSyl |
|---|---|---|---|
| ma | ny | can | come |

"Many can come…"

## Using Algebraic Learning + USC

Familiar word: “come”

StrongSyl WeakSyl1 WeakSyl2 StrongSyl
ma ny can come

“Many can come…”

## Using Algebraic Learning + USC

This must be a word:
add it to memory

StrongSyl WeakSyl1 WeakSyl2 StrongSyl
ma ny can come

“Many can come…”

## Algebraic Learning + USC

Precision: 95.9%

Recall: 93.4%

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.

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

Learning from transitional probabilities alone doesn't work so well on realistic data.

Models of children who have additional knowledge about the stress patterns of words in their language have a much better chance of succeeding at word segmentation if they learn via transitional probabilities.

However, models of children who use algebraic learning as well as have additional knowledge about language-specific stress patterns perform even better at word segmentation.

## Questions?