Psych 156A/ Ling 150: Psychology of Language Learning

Lecture 5
Words in Fluent Speech II

Announcements

HW1 returned

Review question for words now posted

Reminder: be working on HW2

Computational Problem

Divide spoken speech into individual words

túðəkǽsəlbijándðəgáblınsíti

tú	ðə	kǽsəl	bijánd	ðə	gáblın	síti
to	the	castle	beyond	the	goblin	city

Recap: 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)

Computational model: a program that simulates the mental processes occurring in a child. This requires knowing what the input and output are, and then testing the algorithms that can take the given input and transform it into the desired output.

For word segmentation, the input is a sequence of syllables and the desired output is words (groups of syllables).

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*)

ðəbígbǽdwə́lf
↓
ðə bíg bǽd wə́lf
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*)

ðəbígbǽdwə́lf
↓
ðə bíg bǽd wə́lf
the big bad wolf

Recall calculation:
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 word segmentation performance?

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

ðəbígbǽdwə́lf
↓
ðə bíg bǽd wə́lf
the big bad wolf

Precision calculation:
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 word segmentation performance?

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

ðəbígbǽdwə́lf
↓
Error ðəbíg bǽd wə́lf
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*)

ðəbígbǽdwə́lf
↓
Error ðəbíg bǽd wə́lf
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 word segmentation performance?

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

ðəbígbǽdwə́lf
↓
Error ðəbíg bǽd wə́lf
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 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 in order to be sure that word segmentation is really successful

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).

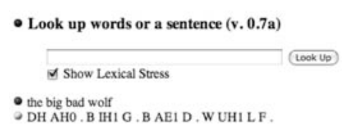
Converted the transcriptions to pronunciations using a pronunciation dictionary called the CMU Pronouncing Dictionary.

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 .
ð ə	b í g	b ǽ d	w ə́ 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.

ð ə   b í g   b ǽ d   w ʊ́ l f

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

“There is a word boundary AB and CD if

TrProb(A --> B) > TrProb(B-->C) < TrProb(C --> D).”

Transitional probability minimum

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 word segmentation

ð ə   b í g   b ǽ d   w ʊ́ l f

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)

ðə bíg bǽd wə́lf

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

ðə bíg bǽd wə́lf

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

ðə bíg bǽd wə́lf

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

learner posits one word boundary at minimum TrProb

ðə bíg | bǽd wə́lf

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

…but nowhere else

ðə bíg ⋮ bǽd wə́lf

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

…but nowhere else

ðə bíg ⋮ bǽd wə́lf

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)

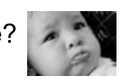
…but nowhere else

ðəbíg ⋮ bǽdwə́lf

thebig ⋮ badwolf

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	bǽd	wə́lf
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 word segmentation problem.

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

ðə bíg | bǽd | wə́lf
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 word segmentation problem.

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

húwz ə fréjd əv ðə bíg | bǽd | wə́lf
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 word segmentation problem.

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

húwz ə fréjd əv ðə bíg | bǽd | wə́lf
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 word segmentation problem.

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

? ?
húwz ə fréjd əv ðə bíg bǽd wólf
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%

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.

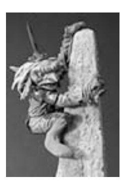
Only about 25% of the words posited by the transitional probability learner are not actually words (73.5% precision) and about 30% of the actual words are not extracted (71.2% recall).

Another Strategy

Algebraic Learning (Gambell & Yang (2003))

Subtraction process of figuring out unknown words.

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



bíg = big (familiar word)

bíggáblın
bíg

gáblın = (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
go	blins	will	see
gá	blɪnz	wɪl	sí

"Goblins will see…"

Using Algebraic Learning + USC

Familiar word: "goblins"

StrongSyl	WeakSyl1	WeakSyl2	StrongSyl
go	blins	will	see
gá	blɪnz	wɪl	sí

"Goblins will see…"

Using Algebraic Learning + USC

"see" is stressed - should be only stressed syllable in word.
Also, "see" is a familiar word

StrongSyl	WeakSyl1	WeakSyl2	StrongSyl
go	blins	will	see
gá	blɪnz	wɪl	sí

"Goblins will see…"

Using Algebraic Learning + USC

"wɪl" must be a word:
add it to memory

StrongSyl	WeakSyl1	WeakSyl2	StrongSyl
go	blins	will	see
gá	blɪnz	wɪl	sí

"Goblins will see…"

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 - 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

Learning from transitional probabilities alone doesn't work so well on realistic data, even though experimental research suggests 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 word segmentation if they learn via transitional probabilities.

However, models of children that use algebraic learning and have additional knowledge about the stress patterns of words perform even better at word segmentation than any of the models learning from the transitional probability between syllables.

Questions?