Psych 156A/ Ling 150: Psychology of Language Learning

Lecture 4

Words in Fluent Speech

Announcements

Homework 1 is due today by the end of class today

Homework 2 available online, due 2/10/09 (after the midterm)

Computational Problem

Divide spoken speech into individual words

túðəkǽsəlbijándðəgáblınsíti

Computational Problem

Divide spoken speech into individual words

túðəkǽsəlbijándðəgáblınsíti

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

Word Segmentation

"One task faced by all language learners is the segmentation of fluent speech into words. This process is particularly difficult because word boundaries in fluent speech are marked inconsistently by discrete acoustic events such as pauses…it is not clear what information is used by infants to discover word boundaries…there is no invariant cue to word boundaries present in all languages."

- Saffran, Aslin, & Newport (1996)

Statistical Information Available

Maybe infants are sensitive to the statistical patterns contained in sequences of sounds.

"Over a corpus of speech there are measurable statistical regularities that distinguish recurring sound sequences that comprise words from the more accidental sound sequences that occur across word boundaries." - Saffran, Aslin, & Newport (1996)

to the castle beyond the goblin city

Statistical Information Available

Maybe infants are sensitive to the statistical patterns contained in sequences of sounds.

"Over a corpus of speech there are measurable statistical regularities that distinguish recurring sound sequences that comprise words from the more accidental sound sequences that occur across word boundaries." - Saffran, Aslin, & Newport (1996)

Statistical regularity: *ca* + *stle* is a common sound sequence

to the castle beyond the goblin city

Statistical Information Available

Maybe infants are sensitive to the statistical patterns contained in sequences of sounds.

"Over a corpus of speech there are measurable statistical regularities that distinguish recurring sound sequences that comprise words from the more accidental sound sequences that occur across word boundaries." - Saffran, Aslin, & Newport (1996)

No regularity: *stle* + *be* is an accidental sound sequence

to the castle beyond the goblin city

word boundary

Transitional Probability

"Within a language, the transitional probability from one sound to the next will generally be highest when the two sounds follow one another in a word, whereas transitional probabilities spanning a word boundary will be relatively low."
- Saffran, Aslin, & Newport (1996)

Transitional Probability = Conditional Probability

TrProb(AB) = Prob(B | A)

Transitional probability of sequence AB is the conditional probability of B, given that A has been encountered.

TrProb("gob" "lin") = Prob("lin" | "gob")

Read as "the probability of 'lin', given that 'gob' has just been encountered

Transitional Probability

"Within a language, the transitional probability from one sound to the next will generally be highest when the two sounds follow one another in a word, whereas transitional probabilities spanning a word boundary will be relatively low."
- Saffran, Aslin, & Newport (1996)

Transitional Probability = Conditional Probability

TrProb("gob" "lin") = Prob("lin" | "gob")

Example of how to calculate TrProb:
gob…
…ble, …bler, …bledygook, …let, …lin, …stopper
(6 options for what could follow "gob")

TrProb("gob" "lin") = Prob("lin" | "gob") = 1/6

Transitional Probability

"Within a language, the transitional probability from one sound to the next will generally be highest when the two sounds follow one another in a word, whereas transitional probabilities spanning a word boundary will be relatively low."
- Saffran, Aslin, & Newport (1996)

Idea: Prob("stle" | "ca") = high
Why? "ca" is usually followed by "stle"

to the castle beyond the goblin city

Transitional Probability

"Within a language, the transitional probability from one sound to the next will generally be highest when the two sounds follow one another in a word, whereas transitional probabilities spanning a word boundary will be relatively low."
- Saffran, Aslin, & Newport (1996)

Idea: Prob("be" | "stle") = lower
Why? "stle" is not usually followed by "be"

to the castle beyond the goblin city
word boundary

Transitional Probability

"Within a language, the transitional probability from one sound to the next will generally be highest when the two sounds follow one another in a word, whereas transitional probabilities spanning a word boundary will be relatively low."
- Saffran, Aslin, & Newport (1996)

Prob("yond" | "be") = higher

Why? "be" is commonly followed by "yond", among other options

to the castle beyond the goblin city

Transitional Probability

"Within a language, the transitional probability from one sound to the next will generally be highest when the two sounds follow one another in a word, whereas transitional probabilities spanning a word boundary will be relatively low."
- Saffran, Aslin, & Newport (1996)

Prob("be" | "stle") < Prob("stle" | "ca")
Prob("be" | "stle") < Prob("yond" | "be")

to the castle beyond the goblin city

TrProb learner posits word boundary here,
at the minimum of the TrProbs

Important: doesn't matter what the probability actually is, so long as it's a minimum when compared to the probabilities surrounding it

8-month-old statistical learning

Saffran, Aslin, & Newport 1996

Familiarization-Preference Procedure (Jusczyk & Aslin 1995)

Habituation:

Infants exposed to auditory material that serves as potential learning experience

Test stimuli (tested immediately after familiarization):

(familiar) Items contained within auditory material

(novel) Items not contained within auditory material, but which are nonetheless highly similar to that material

8-month-old statistical learning

Saffran, Aslin, & Newport 1996

Familiarization-Preference Procedure (Jusczyk & Aslin 1995)

Measure of infants' response:

Infants control duration of each test trial by their sustained visual fixation on a blinking light.

Idea: If infants have extracted information (based on transitional probabilities), then they will have different looking times for the different test stimuli.

Artificial Language

Saffran, Aslin, & Newport 1996

4 made-up words with 3 syllables each

Condition A:

tupiro, golabu, bidaku, padoti

Condition B:

dapiku, tilado, burobi, pagotu

Artificial Language

Saffran, Aslin, & Newport 1996

Infants were familiarized with a sequence of these words generated by speech synthesizer for 2 minutes. Speaker's voice was female and intonation was monotone. There were no acoustic indicators of word boundaries.

Sample speech: 

tu pi ro go la bu bi da ku pa do ti go la bu tu pi ro pa do ti…

Artificial Language

Saffran, Aslin, & Newport 1996

The only cues to word boundaries were the transitional probabilities between syllables.

Within words, transitional probability of syllables = 1.0

Across word boundaries, transitional probability of syllables = 0.33

tu pi ro go la bu bi da ku pa do ti go la bu tu pi ro pa do ti…

Artificial Language

Saffran, Aslin, & Newport 1996

The only cues to word boundaries were the transitional probabilities between syllables.

Within words, transitional probability of syllables = 1.0

Across word boundaries, transitional probability of syllables = 0.33

TrProb("tu" "pi") = 1.0

tu pi ro go la bu bi da ku pa do ti go la bu tu pi ro pa do ti…

Artificial Language

Saffran, Aslin, & Newport 1996

The only cues to word boundaries were the transitional probabilities between syllables.

Within words, transitional probability of syllables = 1.0

Across word boundaries, transitional probability of syllables = 0.33

TrProb("tu" "pi") = 1.0 = TrProb("go" "la"), TrProb("pa" "do")

tu pi ro go la bu bi da ku pa do ti go la bu tu pi ro pa do ti…

Artificial Language

Saffran, Aslin, & Newport 1996

The only cues to word boundaries were the transitional probabilities between syllables.

Within words, transitional probability of syllables = 1.0

Across word boundaries, transitional probability of syllables = 0.33

TrProb("ro" "go") < 1.0 (0.3333…)

tu pi ro go la bu bi da ku pa do ti go la bu tu pi ro pa do ti…

Artificial Language

Saffran, Aslin, & Newport 1996

The only cues to word boundaries were the transitional probabilities between syllables.

Within words, transitional probability of syllables = 1.0

Across word boundaries, transitional probability of syllables = 0.33

TrProb("ro" "go"), TrProb("ro" "pa") = 0.3333… <
1.0 = TrPrb("pi" ro"), TrProb ("go" "la"), TrProb("pa" "do")

tu pi ro go la bu bi da ku pa do ti go la bu tu pi ro pa do ti…

word boundary

word boundary

Testing Infant Sensitivity

Saffran, Aslin, & Newport 1996

Expt 1, test trial:

Each infant presented with repetitions of 1 of 4 words

2 were "real" words

(ex: *tupiro, golabu*)

2 were "fake" words whose syllables were jumbled up

(ex: *ropitu, bulago*)

tu pi ro go la bu bi da ku pa do ti go la bu tu pi ro pa do ti…

Testing Infant Sensitivity

Saffran, Aslin, & Newport 1996

Expt 1, test trial:

Each infant presented with repetitions of 1 of 4 words

2 were "real" words

(ex: *tupiro, golabu*)

2 were "fake" words whose syllables were jumbled up

(ex: *ropitu, bulago*)

tu pi ro go la bu bi da ku pa do ti go la bu tu pi ro pa do ti…

Testing Infant Sensitivity

Saffran, Aslin, & Newport 1996

Expt 1, results:

Infants listened longer to novel items (non-words)

(7.97 seconds for real words, 8.85 seconds for non-words)

Implication: Infants noticed the difference between real words and non-words from the artificial language after only 2 minutes of listening time!

Testing Infant Sensitivity

Saffran, Aslin, & Newport 1996

Expt 1, results:

Infants listened longer to novel items (non-words)

(7.97 seconds for real words, 8.85 seconds for non-words)

Implication: Infants noticed the difference between real words and non-words from the artificial language after only 2 minutes of listening time!

But why?

Could be that they just noticed a familiar sequence of sounds ("tupiro" familiar while "ropitu" never appeared), and didn't notice the differences in transitional probabilities.

Testing Infant Sensitivity

Saffran, Aslin, & Newport 1996

Expt 2, test trial:

Each infant presented with repetitions of 1 of 4 words

2 were "real" words

(ex: *tupiro, golabu*)

2 were "part" words whose syllables came from two different words in order

(ex: *pirogo, bubida*)

tu pi ro go la bu bi da ku pa do ti go la bu tu pi ro pa do ti…

Testing Infant Sensitivity

Saffran, Aslin, & Newport 1996

Expt 2, test trial:

Each infant presented with repetitions of 1 of 4 words

2 were "real" words

(ex: *tupiro, golabu*)

2 were "part" words whose syllables came from two different words in order

(ex: *pirogo, bubida*)

tu pi ro go la bu bi da ku pa do ti go la bu tu pi ro pa do ti…

Testing Infant Sensitivity

Saffran, Aslin, & Newport 1996

Expt 2, test trial:

Each infant presented with repetitions of 1 of 4 words

2 were "real" words

(ex: *tupiro, golabu*)

2 were "part" words whose syllables came from two different words in order

(ex: *pirogo, bubida*)

tu pi ro go la bu bi da ku pa do ti go la bu tu pi ro pa do ti…

Testing Infant Sensitivity

Saffran, Aslin, & Newport 1996

Expt 2, results:

Infants listened longer to novel items (part-words)

(6.77 seconds for real words, 7.60 seconds for part-words)

Implication: Infants noticed the difference between real words and part-words from the artificial language after only 2 minutes of listening time! They are sensitive to the transitional probability information.

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.

Questions?