Psych 156A/ Ling 150: Acquisition of Language II

6/7/2012
Final Exam Review

Final Exam

Final Exam: 6/14/2012
1:30 – 3:30pm

HH178 (this room) OR SBSG G241

We will be holding office hours next week at our normal times

Part of Speech Learning

Two ideas:
- Semantic Bootstrapping Hypothesis
  - PoS matches (roughly) real world semantics
    - nouns → objects, states
    - verbs → actions
    - adjectives → properties
  - But only roughly…
    - a kick (verb-like, but a noun)
    - function words (a, the, of, but…)

Part of Speech Learning

Another idea:
- Frequent Frames

the _____ is	you _____ it
a _____ is	they _____ her
that _____ was	can _____ him

Proposed in Mintz (2003), simulated in Wang & Mintz (2008)

Language Structure

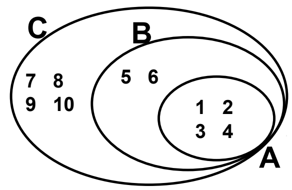
Phrases

Grammaticality judgments

Ambiguous/Unambiguous data

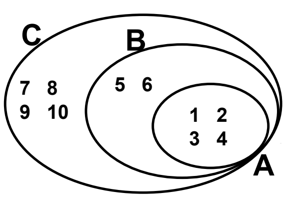
Principles & Parameters

Testing Hypotheses



1, 3, 2, 6, 4, 3

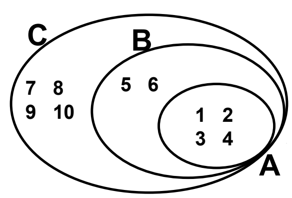
Bayesian Learning



D = 1, 3, 2, 6, 4, 3

P(A|D) = P(D|A) P(A) / P(D)

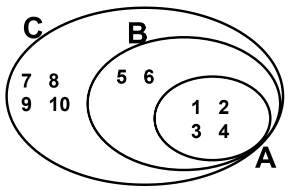
Bayesian Learning



D = 1, 3, 2, 6, 4, 3

$$P(D|A) = P(1|A) * P(3|A) * P(2|A) * P(6|A) * P(4|A) * P(3|A)$$
$$= \frac{1}{4} * \frac{1}{4} * \frac{1}{4} * 0 * \frac{1}{4} * \frac{1}{4}$$
$$= 0$$

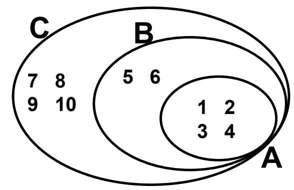
Bayesian Learning



D = 1, 3, 2, 6, 4, 3

P(D|B) = P(1|B) * P(3|B) * P(2|B) * P(6|B) * P(4|B) * P(3|B)
= 1/6 * 1/6 * 1/6 * 1/6 * 1/6 * 1/6 * 1/6
= 1/(6^5) = 1/7776 = .0001286

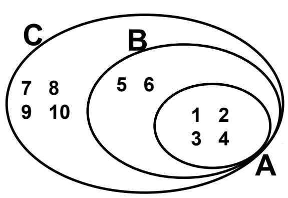
Bayesian Learning



D = 1, 3, 2, 6, 4, 3

P(D|C) = P(1|C) * P(3|C) * P(2|C) * P(6|C) * P(4|C) * P(3|C)
= 1/10 * 1/10 * 1/10 * 1/10 * 1/10 * 1/10 * 1/10
= 1/(10^5) = 1/100000 = .00001

Bayesian Learning

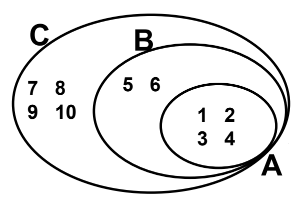


D = 1, 3, 2, 6, 4, 3

P(D|A) = 0 P(D|B) = .0001286 P(D|C) = .00001
P(A) = 1/3 P(B) = 1/3 P(C) = 1/3

P(D) = 0*1/3 + .0001286*1/3 + .00001*1/3
= .0000462

Bayesian Learning

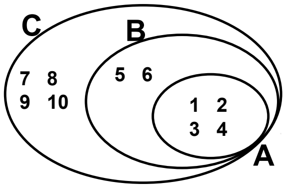


D = 1, 3, 2, 6, 4, 3

P(D|A) = 0 P(D|B) = .0001286 P(D|C) = .00001
P(A) = 1/3 P(B) = 1/3 P(C) = 1/3
P(D) = .0000462

P(A|D) = 0 * 1/3 / .0000462 = 0

Bayesian Learning

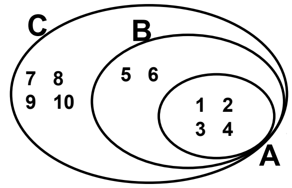


D = 1, 3, 2, 6, 4, 3

P(D|A) = 0 P(D|B) = .0001286 P(D|C) = .00001
P(A) = 1/3 P(B) = 1/3 P(C) = 1/3
P(D) = .0000462

P(B|D) = .0001286 * 1/3 / .0000462 = .9278

Bayesian Learning

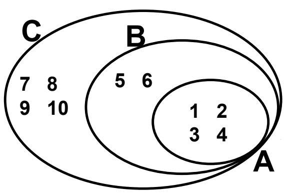


D = 1, 3, 2, 6, 4, 3

P(D|A) = 0 P(D|B) = .0001286 P(D|C) = .00001
P(A) = 1/3 P(B) = 1/3 P(C) = 1/3
P(D) = .0000462

P(B|D) = .0001286 * 1/3 / .0000462 = .9278
P(C|D) = .00001 * 1/3 / .0000462 = .07215

Bayesian Learning

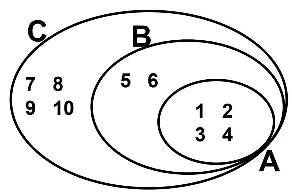


D = 7
But… you already saw 2

Calculate P(H|2) for each hypothesis
P(A|2) = 15/31 P(B|2) = 10/31 P(C|2) = 6/31

Use these posteriors as the new prior for the new datapoint

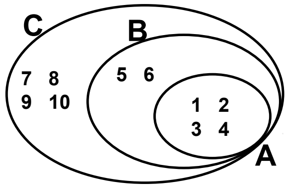
Bayesian Learning



D = 7
But… you already saw 2

P(A|2) = P(2|A) * P(A) / P(D)
= (¼ * 1/3) / (31/180) = 15/31
P(B|2) = P(2|B) * P(B) / P(D)
= (1/6 * 1/3) / (31/180) = 10/31
P(C|2) = P(2|C) * P(C) / P(D)
= (1/10 * 1/3) / (31/180) = 6/31
P(D) = (1/4 * 1/3) + (1/6 * 1/3) + (1/10 * 1/3) = 31/180

Bayesian Learning

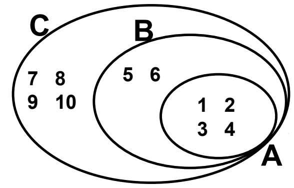


D = 7
But… you already saw 2

P(A|2) = 15/31 P(B|2) = 10/31 P(C|2) = 6/31

P(A|7,2) = P(7|A) * P(A|2) / P(D)
= 0 * 15/31 / P(D)
= 0

Bayesian Learning

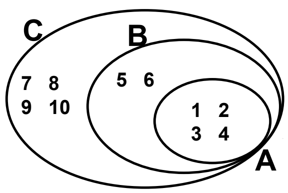


D = 7
But… you already saw 2

P(A|2) = 15/31 P(B|2) = 10/31 P(C|2) = 6/31

P(B|7,2) = P(7|B) * P(B|2) / P(D)
= 0 * 10/31 / P(D)
= 0

Bayesian Learning



D = 7
But… you already saw 2

P(A|2) = 15/31 P(B|2) = 10/31 P(C|2) = 6/31

P(C|7,2) = P(7|C) * P(C|2) / P(D)
= 1/10 * 6/31 / P(D)
= 1/10 * 6/31 / (1/10 * 6/31 + 0 + 0)
= 1

Parameters

Review Questions: Structure

Question #10:

Suppose we have a parameter Q, we don't know what structures match that parameter though. We think maybe A, B, C & D connect to Q, but aren't sure. Q can only take two values, x1 and x2

a) A, B, and C tend to show x1 while D shows z1, which structures are connected to parameter Q?

Parameters

Review Questions: Structure

Question #10:

Suppose we have a parameter Q, we don't know what structures match that parameter though. We think maybe A, B, C & D connect to Q, but aren't sure. Q can only take two values, x1 and x2

b) If Q really does have value x1 which structures (A,B,C,D) are likely to also have value x1?

Parameters

Review Questions: Structure

Question #10:

Suppose we have a parameter Q, we don't know what structures match that parameter though. We think maybe A, B, C & D connect to Q, but aren't sure. Q can only take two values, x1 and x2

c) Children rarely see structure C, but often see A, B and D. If A & B show x1, and D shows z1, given your answer to (b) what value should the infant suppose for structure C?

Experiments

Dewar & Xu (2010)

- Examine overhypotheses (abstract generalizations based on limited data with apparent regularities)

Gerken (2006)

- How do children generalize?
- Children don't generalize from AAdi stimuli to AAB

Pearl & Mis (2011)

- Baker (1978) assumes only unambiguous data is informative
- Can learn anaphoric one using all ambiguous data if we include data from other pronouns too!

Experiments

Thompson & Newport (2007)

- Adults can learn phrases using transitional probability (TP)

Hudson, Kam & Newport (2005)

- Adults match inconsistent input with inconsistent output
- Children generalize to the most frequent input type

Hudsom, Kam & Newport (2009)

- Adults will generalize if one input is dominant
- But children in this case generalize one determiner and use it almost always

Psych 156A/ Ling 150: Acquisition of Language II

5/3/2012

Midterm Review

Marr's 3 Levels

Any problem can be decomposed into 3 levels:
- Computational level
  - What's the problem to be solved?
- Algorithmic level
  - What (abstract) set of rules solves the problem?
- Implementational level
  - How are those rules physically implemented?

Computational Level

Abstract Problem:
- How do we regulate traffic at an intersection?

Goal:
- Direct lanes of traffic to avoid congestion/accidents

Algorithmic Level

What kind of rules can we use?
- Let Lane go whenever X cars are waiting?
- Let Lane go every X minutes?
- Let 1 car at a time go through the intersection?
- Make one direction always yield to the other?

Implementational Level

How do we physically implement the rule?
- Set up a stop light
- Set up a blinking stop light
- Put up a stop sign
- Have someone direct traffic
- Put up nothing and have drivers implement the rules themselves!

Transitional Probability

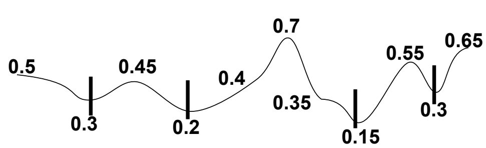
TP(AB) = P(AB|A) = # of times you saw AB / # of times you saw A

ka/ko/si
ko/li/ja
ja/ko
li/je/vo

TP(ko/si) = # of times ko/si / # of times ko

TP(ja/vo) = # of times ja/vo / # of times ja

TP Minima



TP can be though of like a tide

Every time the TP is at "low tide" we put a boundary

Precision & Recall

I wonder how well I can segment this sentence today

Iwonder how well Ican seg ment this sen tencetoday

Precision & Recall

I wonder how well I can segment this sentence today

Iwonder how well Ican seg ment this sen tencetoday

Precision:
# of correct / # guessed

3 correct / 9 guessed

Precision & Recall

I wonder how well I can segment this sentence today

Iwonder how well Ican seg ment this sen tencetoday

Recall:
# of correct / # true words

3 correct / 10 true

Stress-based Segmentation

how **WELL** can a **STRESS** based **LEARNER SEG**ment **THIS**?

If we assume Stress-INITIAL syllables:

How **WELL**cana **STRESS**based **LEARNER** **SEG**ment **THIS**?

Precision = 3/6 Recall = 3/9

Stress-based Segmentation

how **WELL** can a **STRESS** based **LEARNER SEG**ment **THIS**?

If we assume Stress-FINAL syllables:

How**WELL** cana**STRESS** based**LEARNER** **SEG** ment**THIS**?

Precision = 0/5 Recall = 0/9

Bayesian Learning

All (statistical) learning is a form of **INFERENCE**

We have data…

But which hypothesis is true?

P(H|D) ?

P(H | D) = P(D | H) * P(H) / P(D)

posterior likelihood prior prob. of data

Cross-Situational Learning

Use information across trials to identify a word/meaning mapping

Scene 1:	"dugme" Object 1	"lutka" Object 2	"prozor" Object 3
Scene 2:	"lutka" Object 1	"zid" Object 3	"prozor" Object 4

Cross-Situational Learning

Scene 1:	"dugme" Object 1	"lutka" Object 2	"prozor" Object 3
Scene 2:	"lutka" Object 1	"zid" Object 3	"prozor" Object 4

P(H|D) = P(D|H) * P(H) / P(D)

Posterior = likelihood * prior / prob. of data

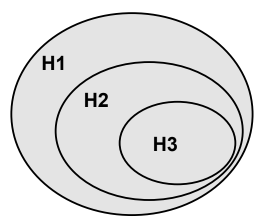
P(lutka == 1) = ¼ Prior (let's call this H1)

P(D | H1) = 1 Likelihood

P(D) = P(H1)*P(D|H1) + P(H2)*P(D|H2) + P(H3)*P(D|H3)…

P(H1 | D) = P(D | H1) * P(H1) / P(D)

Suspicious Coincedence



Three hypotheses:

- Superordinate: "mammal"
- Basic: "dog"
- Subordinate: "beagle"

Given a picture of a beagle:

P(data|H3) = 1/# of beagles

> P(data|H2) = 1/# of dogs

> P(data|H1) = 1/# of mammals

Contrastive Sounds

A pair of sounds are contrastive if:
Switching the sounds changes the **MEANING**

In English:

"food":	[f u d]	← Contrastive
"rude":	[r u d]	

In German:

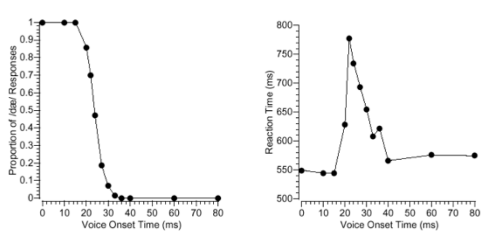
"street":	[s t R a s ə]	← Not contrastive
"street":	[s t r a s ə]	

Learning Sounds

Maintenance & Loss Theory:
- If you use a distinction in your language
  - Keep it
- If you don't use it
  - Ignore the distinction

Functional Reorganization:
- Create a filter between acoustics and phonemes
- If you hear a language sound
  - Impose filter to ignore non-native distinctions
- If you hear a non-language sound
  - Don't impose the filter

Sound Identification



Sound Discrimination

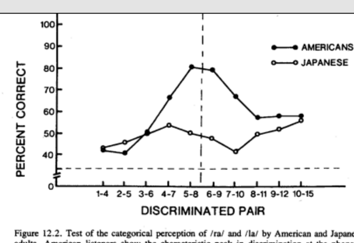


Figure 12.2. Test of the categorical perception of /ra/ and /la/ by American and Japanese adults. American listeners show the characteristic peak in discrimination at the phonetic boundary; Japanese listeners do not. (From Miyawaki et al., 1975.)