Announcements

Pick up your HW1 if you haven’t yet

Review questions available for word meaning

Be working on HW2 (due 5/15/12)
  - Note: Remember that working in a group can be very beneficial.

Midterm review in class on 5/3/12

Midterm exam during class on 5/8/12

What does “gavagai” mean?
What does “gavagai” mean?

- Rabbit?
- Mammal?
- gray rabbit?
- Animal?
- Carrot eater?
- vegetarian?
- Ears?
- Long ears?
- Is it gray?
- Fluffy?
- What a cutie!

Thumping
Hopping
Scurrying
Stay!
Look!

Same problem the child faces

Meal!
Rabbit only until eaten!
Cheeks and left ear!
That’s not a dog!

A little more context…

“Look! There’s a goblin!”

Goblin = ????

The Mapping Problem

Even if something is explicitly labeled in the input (“Look! There’s a goblin”), how does the child know what specifically that word refers to? (Is it the head? The feet? The staff? The combination of eyes and hands? Attached goblin parts?)

Quine (1960): An infinite number of hypotheses about word meaning are possible given the input the child has. That is, the input underspecifies the word’s meaning.
So how do children figure it out? Obviously, they do....

“I love my daxes.”

Dax = that specific toy, teddy bear, stuffed animal, toy, object, …?

One solution: fast mapping
Children begin by making an initial fast mapping between a new word they hear and its likely meaning. They guess, and then modify the guess as more input comes in.

Experimental evidence of fast mapping
One solution: fast mapping

Children begin by making an initial fast mapping between a new word they hear and its likely meaning. They guess, and then modify the guess as more input comes in.

Experimental evidence of fast mapping

ball  bear  kitty  [unknown]

"Can I have the zib?"

20 months

A slight problem…

"…not all opportunities for word learning are as uncluttered as the experimental settings in which fast-mapping has been demonstrated. In everyday contexts, there are typically many words, many potential referents, limited cues as to which words go with which referents, and rapid attentional shifts among the many entities in the scene." - Smith & Yu (2008)

Cross-situational Learning

New approach: infants accrue statistical evidence across multiple trials that are individually ambiguous but can be disambiguated when the information from the trials is aggregated.

Fig. 1. Associations among words and referents across two individually ambiguous scenes. If a young listener validates co-occurrences temporally across these two trials, she can find the proper mapping of "ball" to BALL.
How does learning work?

Bayesian inference is one way.

In Bayesian inference, the belief in a particular hypothesis (H) (or the probability of that hypothesis), given the data observed (D) can be calculated the following way:

$$P(H | D) = \frac{P(D | H) \times P(H)}{P(D)}$$

Posterior probability of hypothesis H, given that data D have been observed
How does learning work?

Bayesian inference is one way.

In Bayesian inference, the belief in a particular hypothesis ($H$) (or the probability of that hypothesis), given the data observed ($D$) can be calculated the following way:

$$P(H | D) = \frac{P(D | H) * P(H)}{P(D)}$$

- **Posterior probability**
- **Likelihood**
- **Prior**

Probability of observing the data, no matter what hypothesis is true:

**Calculate by summing over all hypotheses**

Cross-situational Learning

Let's apply Bayesian inference to this scenario.

Fig. 1. Associations among words and objects across two individually contingent scenes. If a young learner calculates co-occurrence frequencies across these two trials, she can find the proper mapping of “ball” to BALL.
Cross-situational Learning
Let's apply Bayesian inference to this scenario.

Let's apply Bayesian inference to this scenario.

Observable data
Hypothesis 1 (H1): "ball" =
Hypothesis 2 (H2): "ball" =

If this is the only data available,
P(D | H1) = would this be observed if H1 were true? Yes. Therefore p(D | H1) = 1.0.

P(D | H2) = would this be observed if H2 were true? Yes. Therefore p(D | H2) = 1.0.
Let's apply Bayesian inference to this scenario.

If this is the only data available, P(D | H1) = would this be observed if H1 were true? Yes. Therefore P(D | H1) = 1.0.

Hypothesis 1 (H1): "ball" =
Hypothesis 2 (H2): "ball" =
Hypothesis 3 (H3): "ball" =

This feels intuitively right, since "ball" could refer to either object, given this data point.

Since there are three hypotheses in the hypothesis space at this point
P(H1) = 1/3 = 0.33
P(H2) = 1/3 = 0.33
P(H3) = 1/3 = 0.33
Let's apply Bayesian inference to this scenario.

If this is the only data available, the formula is:

\[
P(D | h) P(h) = P(D | H_1) * P(H_1) = 1.0 * 0.33 = 0.33
\]

\[
P(D | H_2) * P(H_2) = 0.0 * 0.33 = 0.0
\]

\[
P(D | H_3) * P(H_3) = 0.0 * 0.33 = 0.0
\]

so

\[
\sum_{h} P(D | h) P(h) = 0.33 + 0.0 + 0.0 = 0.33
\]

This feels intuitively right, since “ball” could only refer to the ball, when these two scenes are reconciled with each other.

---

**Smith & Yu (2008)**


Smith & Yu (2008) ask: Can 12- and 14-month-old infants do this? (Relevant age for beginning word-learning.)
Smith & Yu (2008): Experiment

Training: 30 slides with 2 objects named with two words (total time: 4 min)

Example training slides

Smith & Yu (2008): Experiment

Testing: 12 trials with one word repeated 4 times and 2 objects (correct one and distractor) present

Which one does the infant think is manu? That should be the one the infant prefers to look at.

Smith & Yu (2008): Experiment

Results: Infants preferentially look at target over distracter, and 14-month-olds looked longer than 12-month-olds. This means they were able to tabulate distributional information across situations.

Implication: 12 and 14-month-old infants can do cross-situational learning

Something to think about...

The real world isn't necessarily as simple as these experimental setups - often times, there will be many potential referents.

Fig. 3. (b) a double word learning environment for the word manu. All the double word learning environments for this task in real world industrial distracting experiments.
Something else to think about…

Having more referents may not be a bad thing.

Why not?

It's easier for the correct associations to emerge from spurious associations when there are more object-referent pairing opportunities. Let's see an example of this.

Why more may not always be harder…

Suppose there are six objects total, the amount used in the Smith & Yu (2008) experiment.

First, let's consider their condition, where two objects are shown at a time. Let's say we get three slides/scenes of data.

“manu”
“colat”
“bosa”
“gasser”
“kaki”
“regli”

Why more may not always be harder…

Suppose there are six objects total, the amount used in the Smith & Yu (2008) experiment.

Now, let's consider a more complex condition, where four objects are shown at a time. Let's say we get three slides/scenes of data.

“manu”
“colat”
“bosa”
“gasser”
“manu”
“colat”
“bosa”
“gasser”
“manu”
“gasser”
“kaki”
“regli”

“manu”
“colat”
“bosa”
“gasser”
“kaki”
“regli”

No - both hypotheses are equally compatible with these data.
Why more may not always be harder…

Suppose there are six objects total, the amount used in the Smith & Yu (2008) experiment.

Can we tell whether “manu” refers to \( \text{bosa} \) or \( \text{colat} \) or \( \text{gasser} \) or \( \text{regli} \)?

Well, the first slide isn’t helpful in distinguishing between these four hypotheses…

Why more may not always be harder…

Suppose there are six objects total, the amount used in the Smith & Yu (2008) experiment.

Can we tell whether “manu” refers to \( \text{bosa} \) or \( \text{colat} \) or \( \text{gasser} \) or \( \text{regli} \)?

The second slide suggests “manu” can’t be \( \text{kaki} \) - otherwise, that object would appear in the second slide.

Why more may not always be harder…

Suppose there are six objects total, the amount used in the Smith & Yu (2008) experiment.

Can we tell whether “manu” refers to \( \text{bosa} \) or \( \text{colat} \) or \( \text{gasser} \) or \( \text{regli} \)?

The third slide suggests “manu” can’t be \( \text{kaki} \) or \( \text{regli} \) - otherwise, those objects would appear in the third slide.

Therefore, “manu” is \( \text{bosa} \).

This shows us that having more things appear (and be named) at once actually offers more opportunities for the correct associations to emerge.
Recap: Word-Meaning Mapping

Cross-situational learning, which relies on distributional information across situations, can help children learn which words refer to which things in the world.

One way to implement the reasoning process behind cross-situation learning is Bayesian inference.

Experimental evidence suggests that infants are capable of this kind of reasoning in controlled experimental setups.

Questions?

Use the remaining time to work on HW2 and the review questions for word meaning. You should be able to do up through question 5 on HW2 and up through question 4 on the review questions.