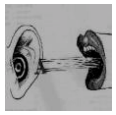
Lecture 18: Language Acquisition I

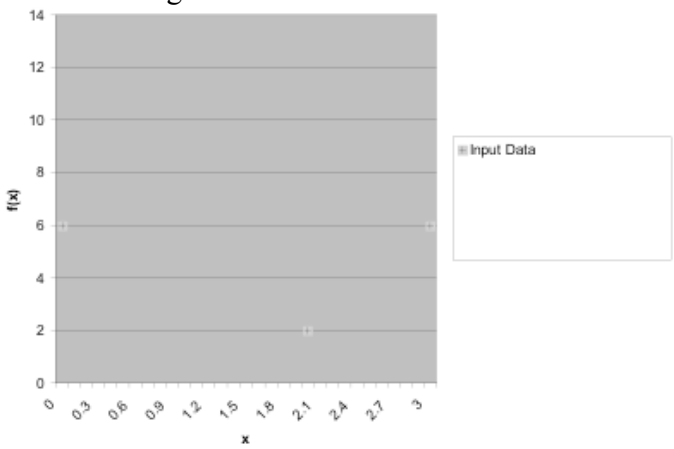


Professor Robert C. Berwick
berwick@csail.mit.edu

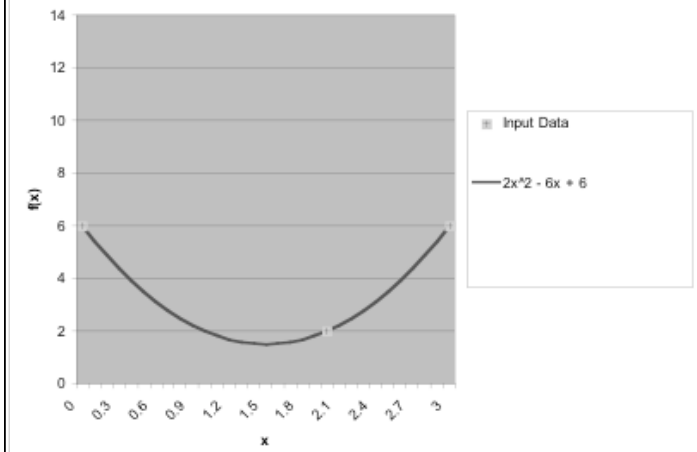
The Menu Bar

- Administrivia: lab 5-6 due this Weds!
- Language acquisition – some background; Big Bang Theory
- Language acquisition – the "Gold standard" & basic results *or* "the (Evil) Babysitter is Here" (apologies to Dar Williams)
  - Informal version
  - Formal version
- Can we meet the Gold standard? What about probabilistic accounts?

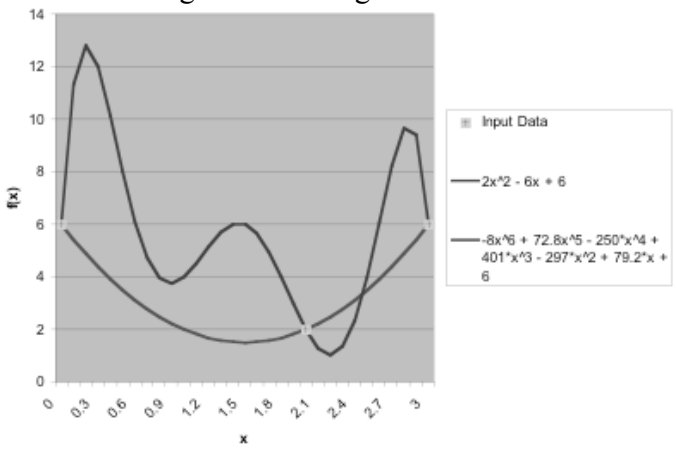
Learning: Observe some values of a function



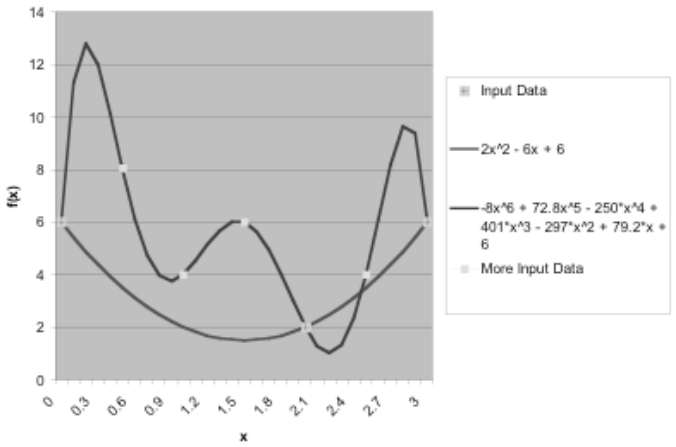
Guess the whole function



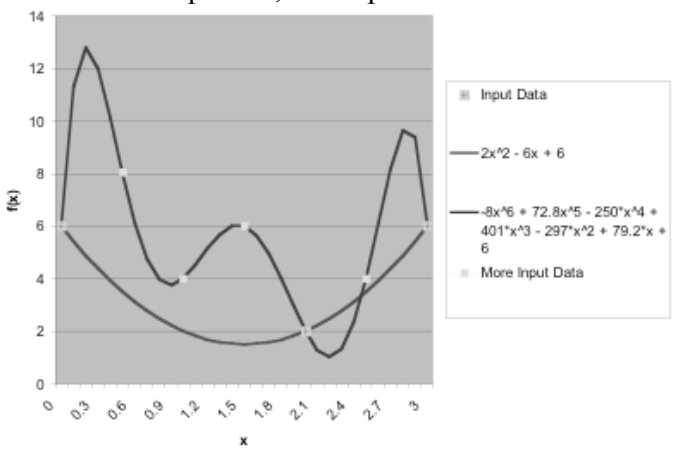
Another guess: Just as good?



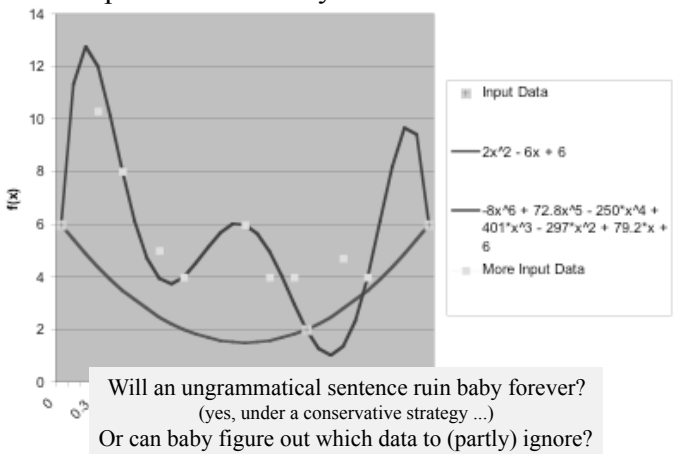
More data needed to decide



Perfect fit to perfect, incomplete data



Imperfect fit to noisy data

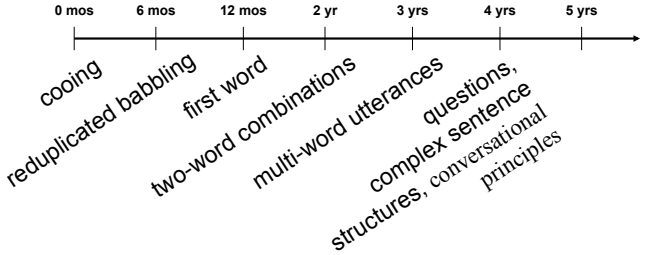


Will an ungrammatical sentence ruin baby forever?
(yes, under a conservative strategy ...)
Or can baby figure out which data to (partly) ignore?

Poverty of the Stimulus

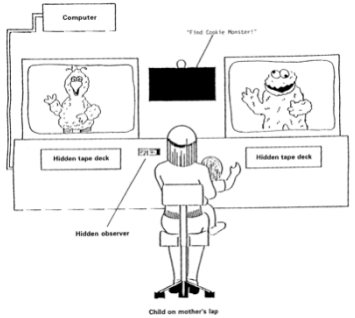
- Never enough input data to completely determine the polynomial …
  - Always have infinitely many possibilities
- … unless you know the order of the polynomial ahead of time.
  - 2 points determine a line
  - 3 points determine a quadratic
  - etc.
- In language learning, is it enough to know that the target language is generated by a CFG?
  - Without knowing the size of the CFG?
  - What else must be known?

Children are amazing learners



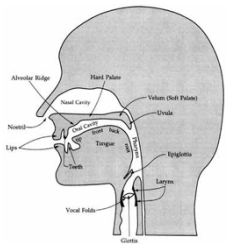
Word order: agent and patient

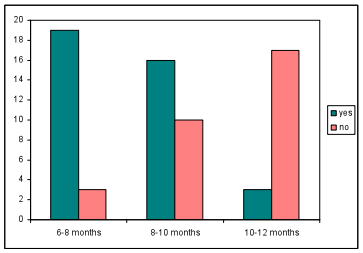
- Hirsch-Pasek and Golinkoff (1996)
- 1;4-1;7
- mostly still in the one-word stage
- Where is CM tickling BB?



Phonology: Non-native contrasts

- Werker and Tees (1984)
- Thompson: velar vs. uvular, /`ki/-/`qi/.
- Hindi: retroflex vs. dental, /t.a/-/ta/





The real situation

two same way
" little boy's mitten
two good ones .
twelve fourteen o'clock
post office mail
oh # your rope tricks
little red spots
little fish house
oh poor little fellow
your washing maxne
some child outside .
two seals # one strong man .
one bareback rider .
little soda water
same just like what ?
some water juice ?
green stamp basket
some pretty work ?
two street lights
one more what
dirty water germs
red street light .
three dozen milk
yes # and one dozen eggs .
your dozen what
one dozen eggs
two little birdies ?
your automatic rifle .
two dirty fingers # but I don't think they're broken .
dirty dirty fingers

Early syntax

- agent + action ‘Daddy sit’
- action + object ‘drive car’
- agent + object ‘Mommy sock’
- action + location ‘sit chair’
- entity + location ‘toy floor’
- possessor + possessed ‘my teddy’
- entity + attribute ‘crayon big’
- demonstrative + entity ‘this telephone’

Big Bang Theory: From Single Words To Complex Utterances

FATHER: Nomi are you climbing up the books?
NAOMI: up.
NAOMI: climbing.
NAOMI: books.
1;11.3

MOTHER: what are you doing?
NAOMI: I climbing up.
MOTHER: you’re climbing up?
2;0.18

FATHER: what’s the boy doing to the dog?
NAOMI: squeezing his neck.
NAOMI: and the dog climbed up the tree.
NAOMI: now they’re both safe.
NAOMI: but he can climb trees.
4;9.3

Sachs corpus (CHILDES)

Is this what children pick up from what they hear?

what did you throw it into?
they’re throwing this in here.
they’re throwing a ball.
don’t throw it Naomi.

well you really shouldn’t throw things Naomi you know.
remember how we told you you shouldn’t throw things.

- Assume: Children use rich situational context / cues to fill in the gaps
- They also have at their disposal embodied knowledge and statistical correlations (i.e. experience)

Is this really the situation?

Language learning:
What kind of evidence?

The guy is happy / is the guy happy
The guy who is happy is tired /

- Children listen to language [unsupervised, 'unlabeled' data]
- Children are corrected?? [supervised, 'labeled' data]
- Children observe language in context
- Children observe frequencies of sentences/bigrams/….

Remember: Language = set of strings (but we can generalize this)

Poverty of the Stimulus (circa 1955 on)

Chomsky (and the Cartesian tradition): Just like polynomials – never enough data unless you know something in advance. So kids must be born knowing what to expect in language.

- Children listen to language
- Children are corrected??
- Children observe language in context
- Children observe frequencies of sentences, bigrams, etc. used…

What *is* the input to children? Is there 'motherese'?

- Gleitman and Landau (1977): in fact, mothers use mostly grammatical sentences when they speak to children
- Children are not corrected for syntactically ill-formed utterances (why do you think?)
- "Mother I'd rather do it myself"

"Mother, I'd rather do it myself"

- In controlled studies (from Brown, 1967 on): children are corrected for *semantic* mal-formed sentences, but not really all that often for *syntactic* ones (why?)
- When they *are* corrected for (usually minor morphological) overgeneralizations, e.g., *goed-went*, they (surprise) don't usually listen to their parents…

Grammar may not matter

- Brown & Hanlon (1970):
- Adults understood **42%** of the grammatical sentences.
- Adults understood **47%** of the ungrammatical ones.
- Adults expressed approval after **45%** of the grammatical sentences.
- Adults expressed approval after **45%** of the ungrammatical sentences.

But it is not

- Sometimes adults say "no" when a sentence is correct
- Unless kids have extremely clear ability to distinguish types of "no" they will be confused
- So, direct reliance on provision of corrective feedback may not be such a great strategy

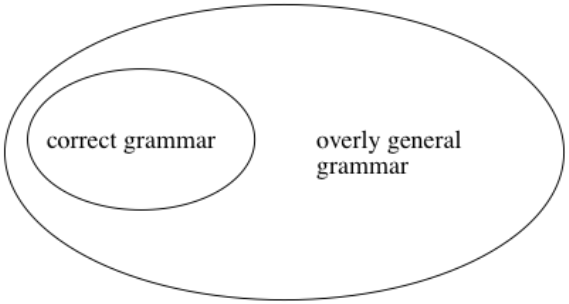
The classic example

- Child: Nobody don't like me.
- Mother: No, say "Nobody likes me."
- Child: Nobody don't like me.
- Mother: No, say "Nobody likes me."
- Child: Nobody don't like me.
- Mother: No, say "Nobody likes me."
- Child: Nobody don't like me.
- [dialogue repeated five more times]
- Mother: Now listen carefully, say "Nobody likes me."
- Child: Oh! Nobody don't likeS me.

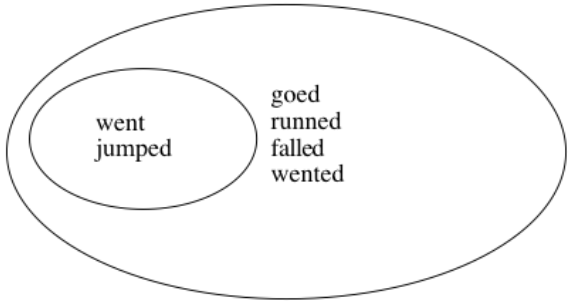
Recovery from Overgeneralization

- u-shaped curve: went - goed - went
- child must stop saying:
  - "goed"
  - "unsqueeze"
  - "deliver the library the book"

An overly general grammar



For example



Note the overgeneralization problem with positive-only evidence

- If only positive evidence, and the learner *guesses* that the language includes the wider circle, nothing in the input will tell them they are wrong

Negative Evidence

- Some constructions seem impossible to learn without *negative evidence*

John gave a painting to the museum
John gave the museum a painting
John donated a painting to the museum
* John donated the museum a painting

Feedback must be *consistent*

	Feedback	No Feedback
Error	Hit	Miss
Correct	False alarm	Correct rejection

High signal detection maximizes p(hit)/p(FA)

Hits close to 1.0 and FA close to 0.0

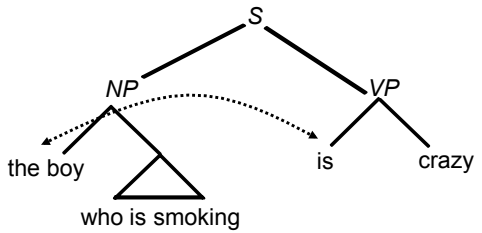
The issue of prior knowledge

kids (or computers) can't learn much
without supervision or *a priori* knowledge

- Children listen to language
- Children are corrected?? - NOT! (is this true?)
- Children observe language in context
- But even then, children don't seem to generalize properly…
- Example: *The child is crying → Is the child crying*
- Rule H1: move first aux vs.
- Rule H2: move structurally most prominent aux

Graphically

- *The boy who is smoking is crazy*
- If H1 → *Is the boy who smoking is crazy*
- If H2 → *Is the boy who is smoking crazy*



(Tentative) conclusion

- Children know *a priori* that rules apply to structure, not left-to-right order ("structure-dependence)
- This rules out many hypothetically plausible grammars
- They (seem!) to do this w/o explicit positive evidence – or do they?

Do children get any positive evidence about these examples?

- A search of the English CHILDES database (3 million utterances) by Lewis and Elman (2001) found only one utterance in the input to Adam.
- MacWhinney also found one in the input in the Hall corpus.

No evidence, no production

- So there is no positive evidence of this type
- But there is also no production, so how can we know that children follow the constraint?

Experimental Evidence

- Crain and Nakayama (1987): *Ask Jabba if the boy who is watching Mickey is happy*
- Children (3-5) never moved the AUX of the relative, although they did other strange things
- But the procedure gives the children the relative clause as a frozen unit
  - "the boy who is watching Mickey"
  - There is a fundamental pragmatic fact about relative clause freezing that the structural analyses are ignoring, but …
- Still, let us grant that children have some sense of this constraint by age 4

Complex-NP Constraint

- Who did John believe __ kissed his buddy?
- * Who did John believe the man that kissed __ arrived
- * Who did pictures of ___ surprise you?
- * What did you see a happy ___ ?
- * What did you stand between the wall and ___ ?

Data from Seth (3-4)

- What am I cooking on a hot __ ? (stove)
- What are we gonna look for some __ ? (houses)
- What is this a funny __ , Dad?
- What are we gonna push number __ ? (9)
- Where did you pin this on my __ ? (robe)
- What are you shaking all the __ ? (batter and milk)
- What is this medicine for my __ ? (cold)

Binding conditions

- Devilliers, Roeper, and Vainikka 1990
  1. *When did the boy say he hurt himself?*
  2. *When did the boy say how he hurt himself?*
  3. *Who did the boy ask what to throw?*
- Young children can't understand #3
- Children will associate "when" with "hurt" in #1 more than #2, but this understanding grows with age
- Therefore binding seems to be learned

Wh-movement

Who$_i$ do you think Lord Emsworth will invite t_i?
Who$_i$ do you think that Lord Emsworth will invite t_i?

Who$_i$ do you think t_i will arrive first?
* Who$_i$ do you think that t_i will arrive first?

Is this error-free learning?

- There are definitely errors.
- There is definitely positive evidence.
- But it is true that errors seem to be relatively scarce.
- So, this is "low error" learning.
- We will return to this later

Conservatism

- Conservative child learners only use forms they have heard adults use
- This is to *avoid* over-generalizing and guessing *too large* a language, from which they can't recover

Conservativism (cont.)

- Who hit the little girl with the block today?
- Who did the little girl hit _ with the block today?
- Who did the boy play with _ behind his mother?
- Who did the boy read a story about _ ?
- Child never hears: Who do the boy read a story that described _ ?

Conservativism (cont.)

- *Who did John believe the man that kissed ___ arrived?
- Who did John believe __ kissed his buddy?
- *What did you stand between the wall and __?
- *What did you see a happy ___?

What about this kind of negative evidence?

- If you haven't heard a certain constraint/construction after a certain time, you'll never hear it (because it's not part of your target language/grammar)
- Example: suppose you hear, "I am happy" a thousand times
- But never "He am happy"
- Can we infer from (comparatively) low frequency to ungrammatical? Tricky!

Indirect negative evidence and frequency

- Note that low probability on its own cannot imply ungrammaticality: if infinite # of grammatical sentences, then there cannot be a lower bound on their probability
- Because: if all grammatical sentences at least probability ε, then there can be at most $1/\varepsilon$ grammatical sentences, which implies nautral language is finite

Indirect Negative Evidence

- Lasnik, Chomsky, Braine, Berwick, Siskind

$$\frac{\text{average frequency of V}}{\text{average frequency of V-ed}} = \frac{\text{frequency of } go}{\text{frequency of } goed}$$

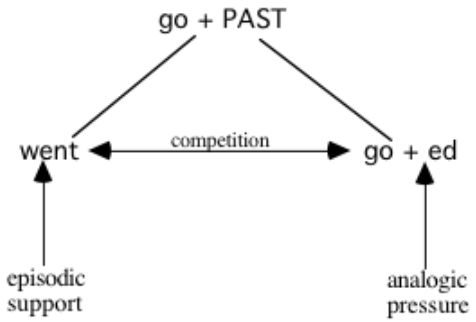
$$\frac{x}{y} = \frac{x}{y'}$$

- If $x/y < x'/y'$ by a large amount
- and if y is frequent, then y′ must be incorrect

Indirect Negative Evidence (cont.)

do	undo
tie	untie
zip	unzip
squeeze	(unsqueeze)

Competition - example



Indirect Negative Evidence (cont.)

- N in relative = N in complement
- N extracted N extracted

- *Bill thought the thieves were carrying the loot*
- *What did Bill think the thieves were carrying*

- *The police arrested the thieves who were carrying the loot.*

- * *What did the police arrest the thieves who were carrying*

How to formalize all this?

- We'll use a model first proposed by E.M. Gold (1965, 1967)
- Gold wanted to prove results given the most generous computational resources available to a learner
- So we assume only that it is possible for the learner to enumerate the family of possible languages ("recursively enumerable") without any other constraints

What else?
The "Evil Babysitter" is here…

- Babysitter talks: outputs sentences from a single 'target' Language L_T, s_1, s_2, …
- Baby listens

1. Babysitter outputs a sentence
2. Baby hypothesizes what the language is (given all sentences so far)
3. Go to step 1

- Guarantee: Babysitter's language *is* in the set of hypotheses that Baby is choosing among
- Guarantee: Any sentence of Babysitter's language is *eventually* uttered by Caretaker (even if infinitely many) "fair presentation"
- Assumption: Vocabulary (or alphabet) is finite.

What strategy should Baby adopt?

- C-baby: conservative baby – guess smallest language consistent with input so far
- L-baby: liberal baby – pick largest language consistent with input

(Obviously, other strategies are possible…)

Can baby learn under these conditions?

- Dfn: *Learnability in the limit*
  - There is some point at which Baby's hypothesis is correct and never changes again. Baby has converged!
  - Baby doesn't have to **know** that it's reached this point – it can keep an open mind about new evidence – but if its hypothesis is right, no such new evidence will ever come along.
- A family *C* of languages is *learnable in the limit* iff one could construct a perfect C-Baby that can learn any language $L \in C$ in the limit from a Babysitter who speaks *L*
- Baby knows the class *C* of possibilities, but not *L*
- Is there a perfect Baby for all finite-state languages?
- What about for all context-free languages?

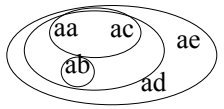
Conservative Strategy

- Baby's hypothesis should always be smallest language consistent with the data
- Works for finite languages? Let's try it …
  - Language 1: {aa,ab,ac}
  - Language 2: {aa,ab,ac,ad,ae}
  - Language 3: {aa,ac}
  - Language 4: {ab}

Babysitter	aa	ab	ac	ab	aa	…
Baby	L3	L1	L1	L1	L1	

Conservative Strategy

- Baby's hypothesis should always be smallest language consistent with the data
- Works for finite languages? Let's try it …
  - Language 1: {aa,ab,ac}
  - Language 2: {aa,ab,ac,ad,ae}
  - Language 3: {aa,ac}
  - Language 4: {ab}



Babysitter	aa	ab	ac	ab	aa	…
Baby	L3	L1	L1	L1	L1	

Evil Babysitter

- To find out whether Baby is perfect, we have to see whether it gets 100% correct even in the most adversarial conditions
- Assume Babysitter is trying to *fool* Baby
  - although she must speak only sentences from L_T
  - and she must eventually speak each such sentence
- Does C-Baby's strategy work on *every* possible 'fair sequence' for *every* possible langauge?

In finite # of languages case, Yes – why?

A learnable ("identifiable") family of Languages

- Family of languages:
  - Let L_n = set of all strings of length < n, over some fixed alphabet = {a, b}
  - What is L_0?
  - What is L_1?
  - What is L_n?
  - Let the family L= {L_0, L_1, …, L_n}
    - No matter what the L_i can Babysitter really follow rules?
    - Must eventually speak every sentence of L. Is this possible?
    - Yes: ∅; a, b; aa, ab, ba, bb; aaa, aab, aba, abb, baa,