NLP and Linguistics

Introduction to Natural Language Processing
Computer Science 585—Fall 2009
University of Massachusetts Amherst

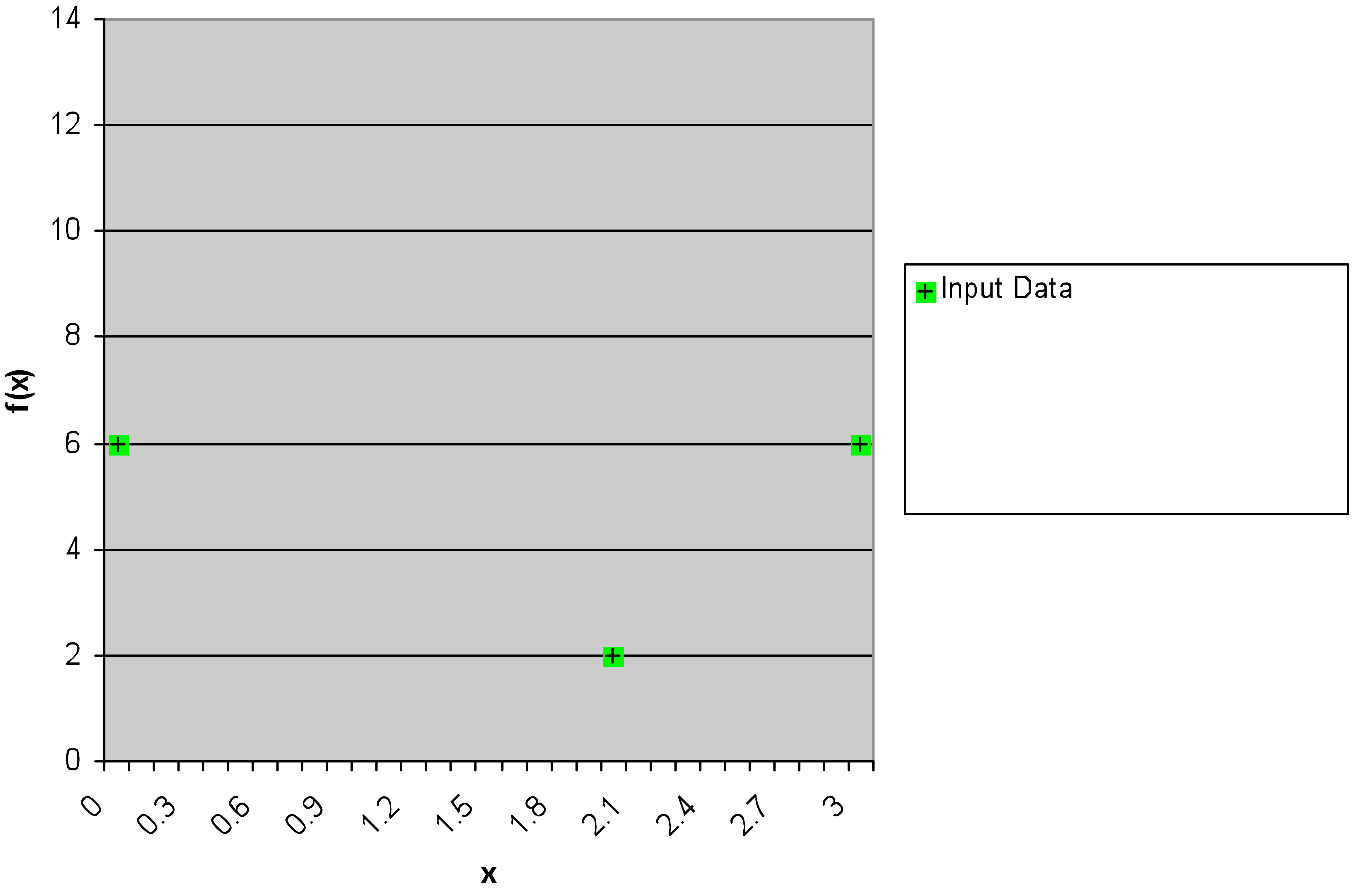
David Smith
With slides from Jason Eisner and Chris Manning

Engineering vs. Science?

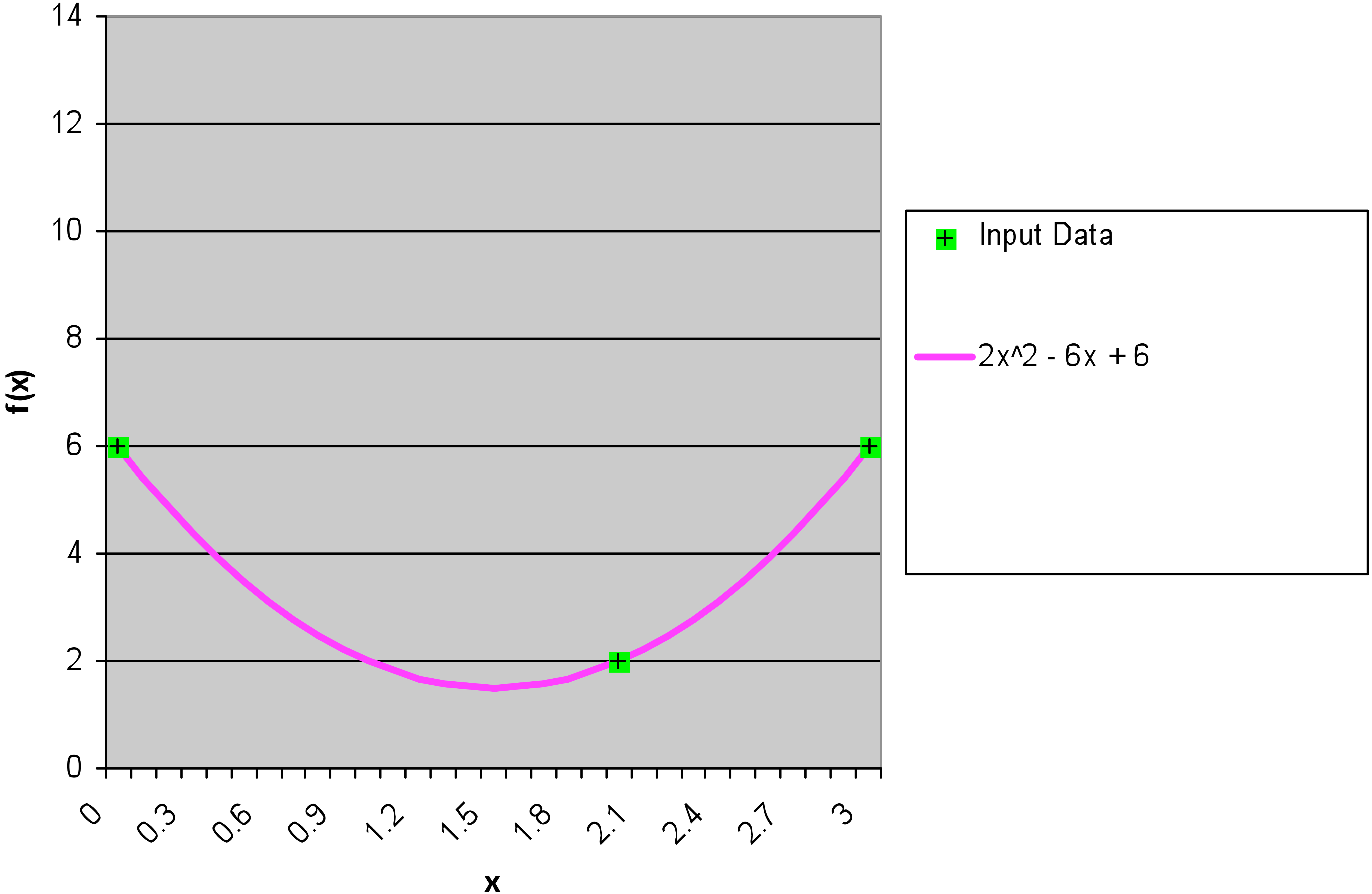
- One story
  - NLP took formal language theory and generative linguistics (same source?),
  - Built small AI systems for a while,
  - Then added statistics/machine learning.
- What now?
  - Shouldn't AI tell us about natural intelligence?
  - Are all NLP models lousy linguistics?

Learning in the Limit
Gold's Theorem

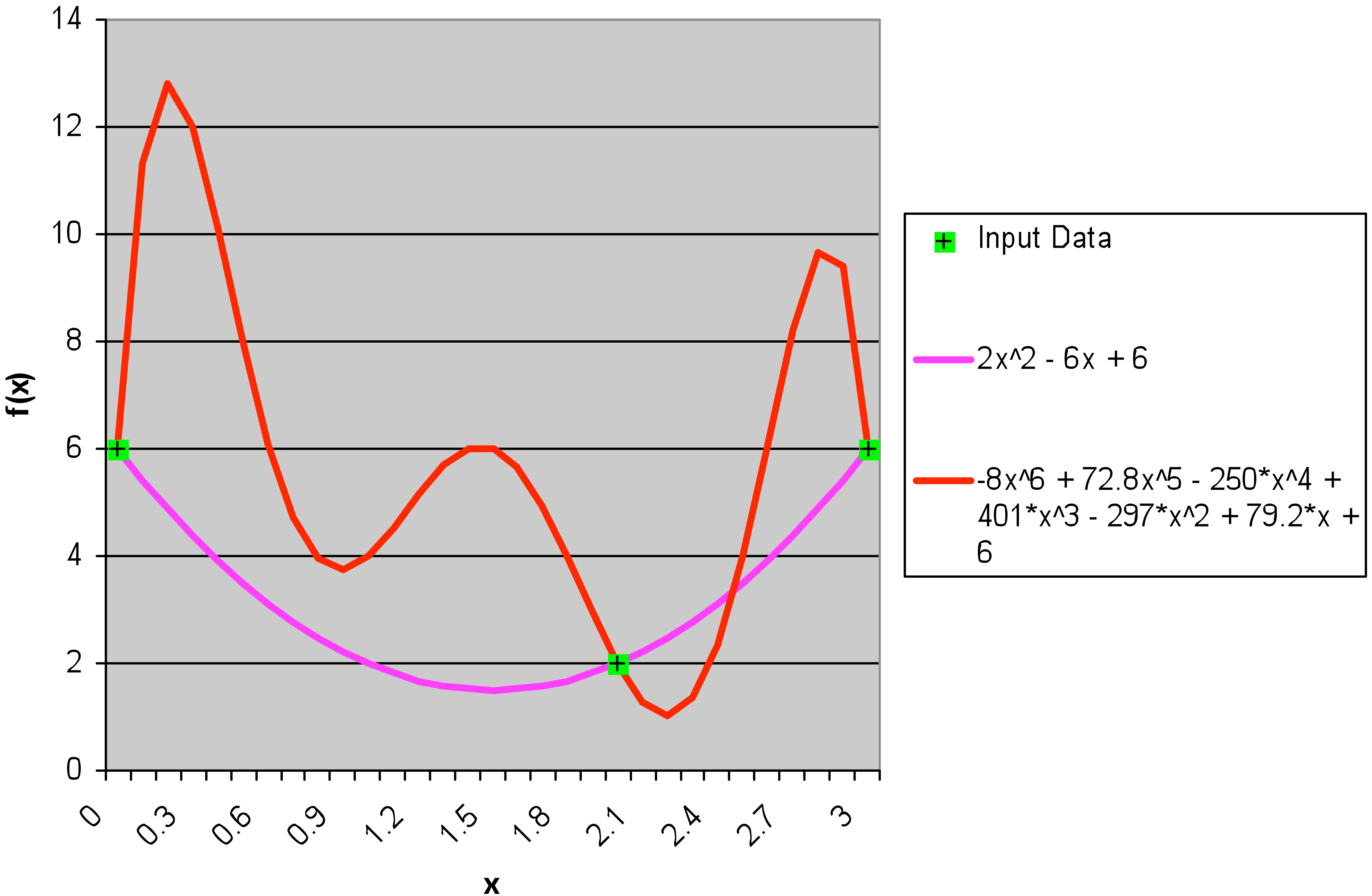
Observe some values of a function



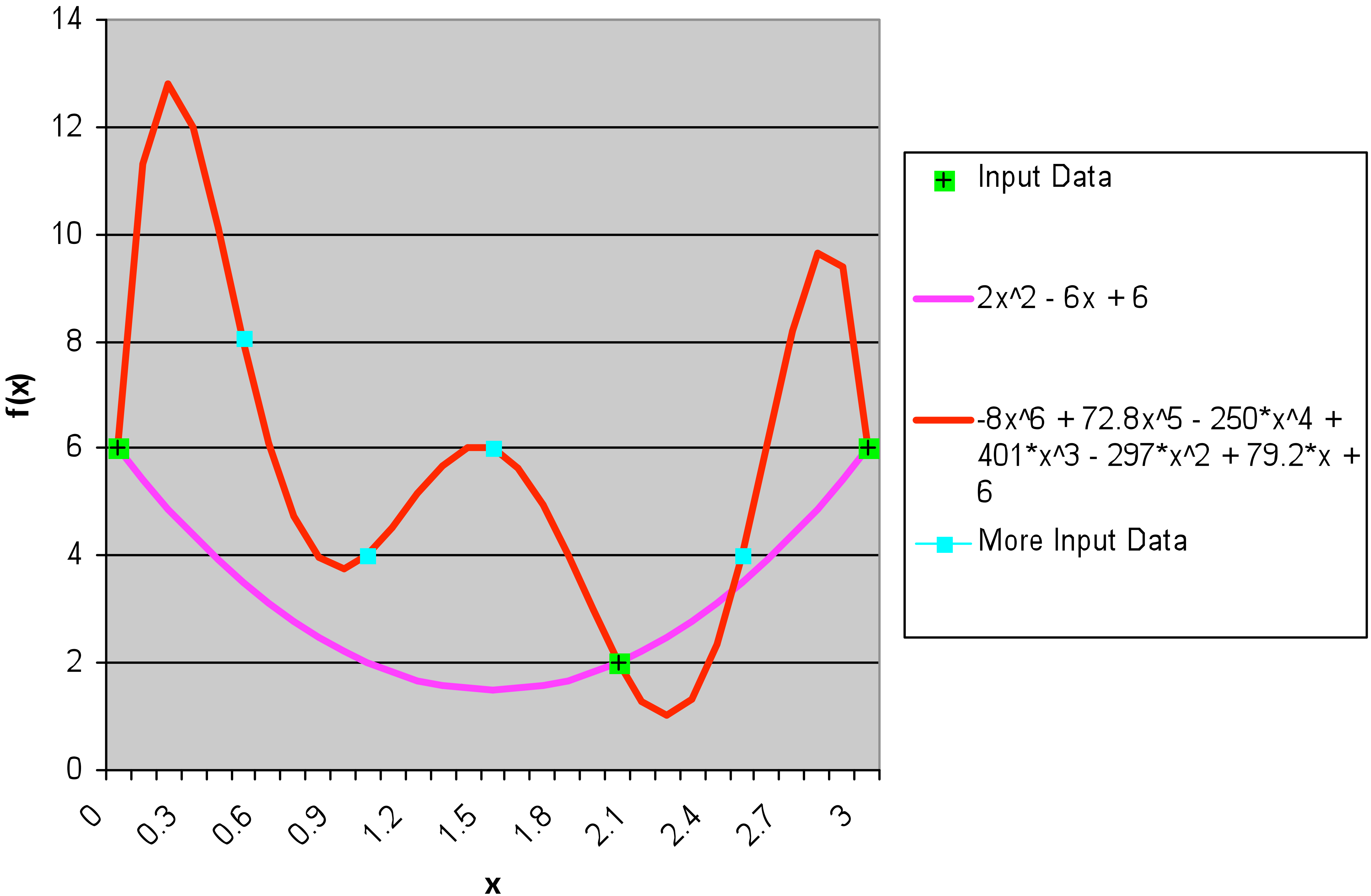
Guess the whole function



Another guess: Just as good?



More data needed to decide



Poverty of the Stimulus

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?

Language learning:

Language learning:

- Children listen to language [unsupervised]

Language learning:

- Children listen to language [unsupervised]
- Children are corrected?? [supervised]

Language learning:

- Children listen to language [unsupervised]
- Children are corrected?? [supervised]
- Children observe language in context

<u>Language</u> learning:

- Children listen to language [unsupervised]
- Children are corrected?? [supervised]
- Children observe language in context
- Children observe frequencies of language

Language learning:

- Children listen to language [unsupervised]
- Children are corrected?? [supervised]
- Children observe language in context
- Children observe frequencies of language

Language learning:

- Children listen to language [unsupervised]
- Children are corrected?? [supervised]
- Children observe language in context
- Children observe frequencies of language

Remember: Language = set of strings

Poverty of the Stimulus (1957)

- Children listen to language
- Children are corrected??
- Children observe language in context
- Children observe frequencies of language

Poverty of the Stimulus (1957)

Chomsky: 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 language

Gold’s Theorem (1967)

a simple negative result along these lines:
kids (or computers) can’t learn much
without supervision, inborn knowledge, or statistics

- Children listen to language
- Children are corrected??
- Children observe language in context
- Children observe frequencies of language

The Idealized Situation

The Idealized Situation

- Mom talks

The Idealized Situation

- Mom talks
- Baby listens

The Idealized Situation

- Mom talks
- Baby listens

The Idealized Situation

- Mom talks
- Baby listens

- 1. Mom outputs a sentence

The Idealized Situation

- Mom talks
- Baby listens

- 1. Mom outputs a sentence
- 2. Baby hypothesizes what the language is
  (given all sentences so far)

The Idealized Situation

- Mom talks
- Baby listens

- 1. Mom outputs a sentence
- 2. Baby hypothesizes what the language is (given all sentences so far)
- 3. Goto step 1

The Idealized Situation

- Mom talks
- Baby listens

- 1. Mom outputs a sentence
- 2. Baby hypothesizes what the language is
  (given all sentences so far)
- 3. Goto step 1

The Idealized Situation

- Mom talks
- Baby listens

- 1. Mom outputs a sentence
- 2. Baby hypothesizes what the language is
  (given all sentences so far)
- 3. Goto step 1

- Guarantee: Mom's language *is* in the set of hypotheses that Baby is choosing among

The Idealized Situation

- Mom talks
- Baby listens

- 1. Mom outputs a sentence
- 2. Baby hypothesizes what the language is
  (given all sentences so far)
- 3. Goto step 1

- Guarantee: Mom's language *is* in the set of hypotheses that Baby is choosing among
- Guarantee: Any sentence of Mom's language is *eventually* uttered by Mom (even if infinitely many)

The Idealized Situation

- Mom talks
- Baby listens

- 1. Mom outputs a sentence
- 2. Baby hypothesizes what the language is
  (given all sentences so far)
- 3. Goto step 1

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

Can Baby learn under these conditions?

Can Baby learn under these conditions?

- Learning 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.

Can Baby learn under these conditions?

- Learning 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 class C of languages is **learnable in the limit** if one could construct a perfect C-Baby that can learn any language $L \in C$ in the limit from a Mom who speaks L.

Can Baby learn under these conditions?

- Learning 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 class C of languages is **learnable in the limit** if one could construct a perfect C-Baby that can learn any language $L \in C$ in the limit from a Mom who speaks L.

Can Baby learn under these conditions?

- Learning 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 class C of languages is **learnable in the limit** if one could construct a perfect C-Baby that can learn any language L $\in$ C in the limit from a Mom who speaks L.

- Baby knows the class C of possibilities, but not L.

Can Baby learn under these conditions?

- Learning 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 class C of languages is **learnable in the limit** if one could construct a perfect C-Baby that can learn any language L $\in$ C in the limit from a Mom who speaks L.

- Baby knows the class C of possibilities, but not L.
- Is there a perfect finite-state Baby?

Can Baby learn under these conditions?

- Learning 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 class C of languages is **learnable in the limit** if one could construct a perfect C-Baby that can learn any language $L \in C$ in the limit from a Mom who speaks L.

- Baby knows the class C of possibilities, but not L.
- Is there a perfect finite-state Baby?
- Is there a perfect context-free Baby?

Languages vs. Grammars

- Does Baby have to get the right grammar?
- (E.g., does VP have to be called VP?)

- Assumption: Finite vocabulary.

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}

Mom
Baby

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}

Mom aa
Baby

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}

Mom	aa
Baby	L3

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}

Mom	aa	ab
Baby	L3	

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}

Mom	aa	ab
Baby	L3	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}

Mom	aa	ab	ac
Baby	L3	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}

Mom	aa	ab	ac
Baby	L3	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}

Mom	aa	ab	ac	ab
Baby	L3	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}

Mom	aa	ab	ac	ab
Baby	L3	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}

Mom	aa	ab	ac	ab	aa
Baby	L3	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}

Mom	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

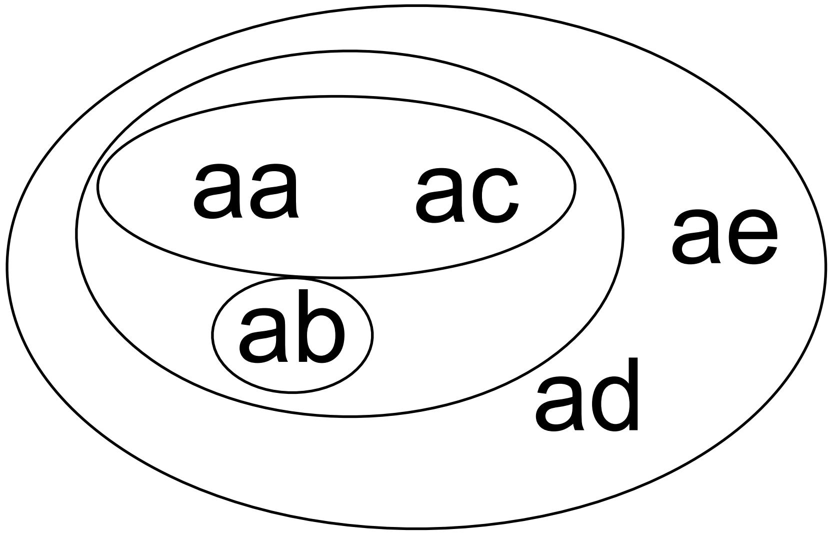
- Works for finite languages? Let's try it …
  - Language 1: {aa,ab,ac}
  - Language 2: {aa,ab,ac,ad,ae}
  - Language 3: {aa,ac}
  - Language 4: {ab}

Mom	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}

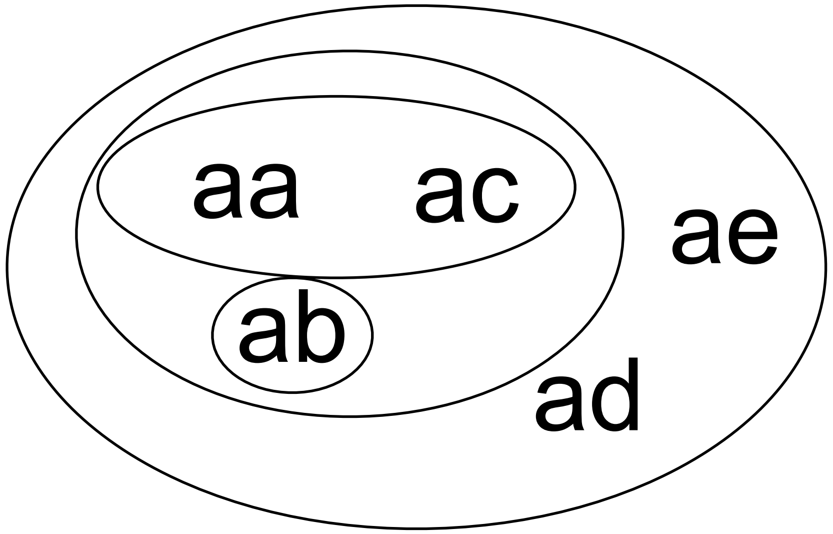


Mom
Baby

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}

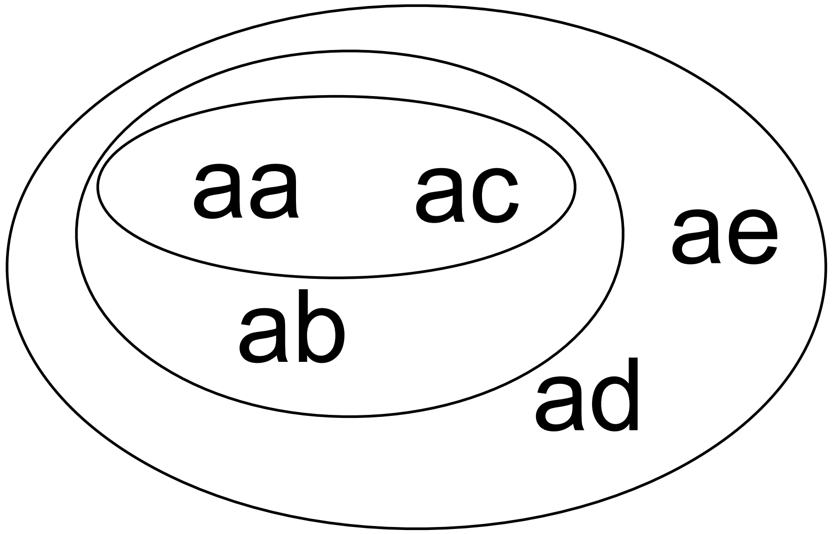


Mom aa
Baby

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}

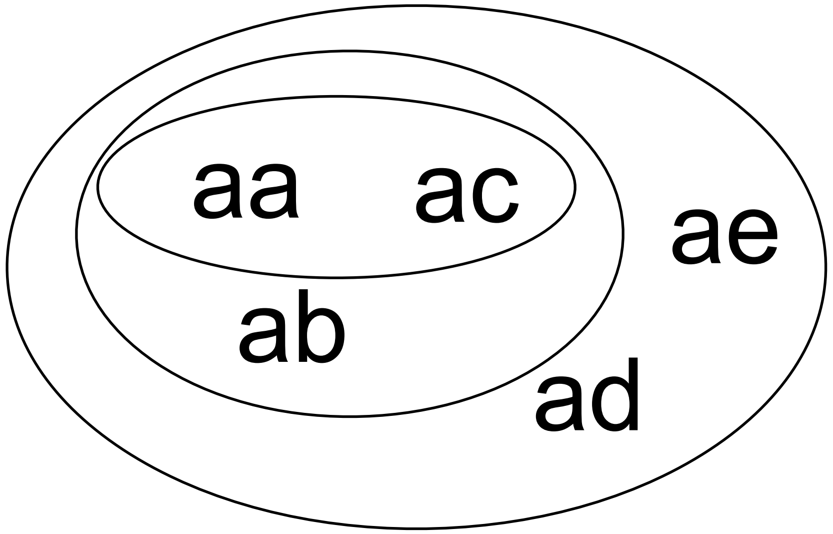


Mom aa
Baby L3

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}

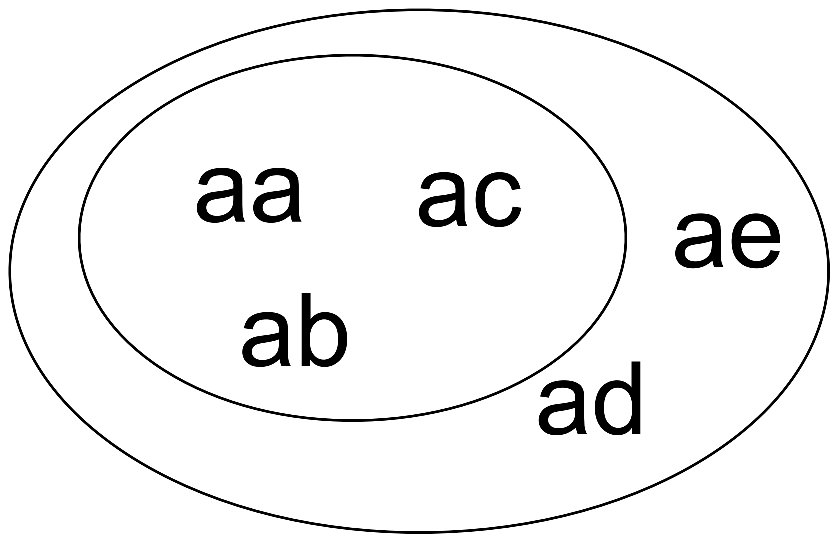


Mom	aa	ab
Baby	L3	

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}

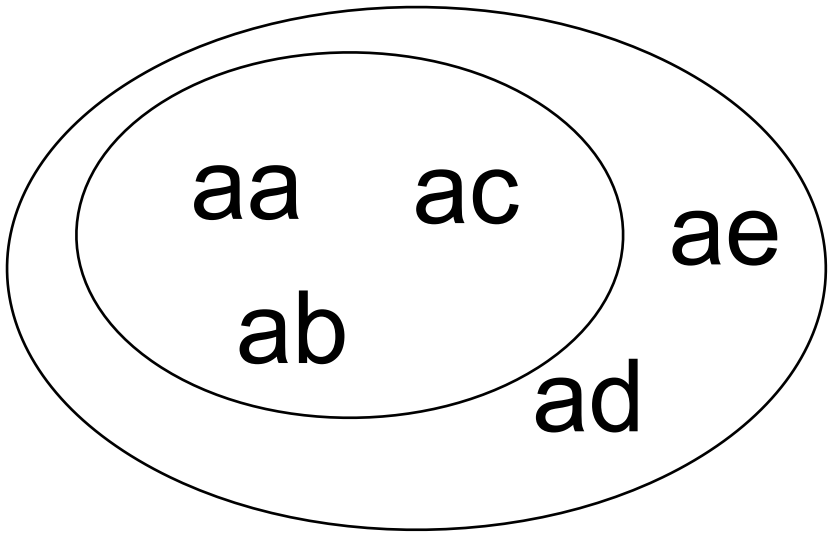


Mom	aa	ab
Baby	L3	L1

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}

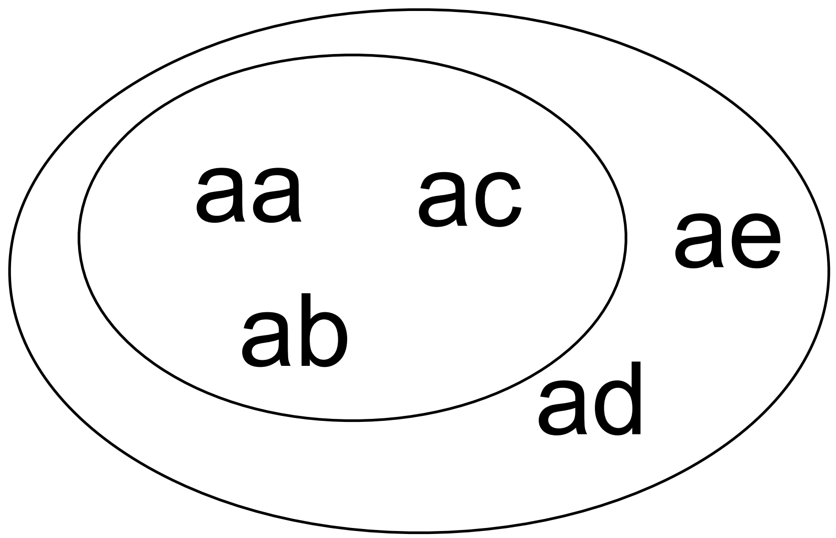


Mom	aa	ab	ac
Baby	L3	L1	

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}

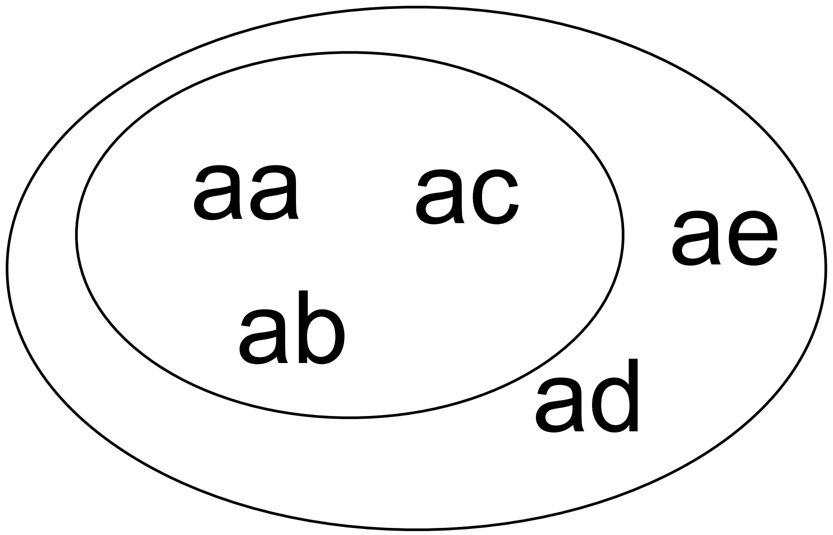


Mom	aa	ab	ac
Baby	L3	L1	L1

Conservative Strategy

- Baby's hypothesis should always be smallest language consistent with the data

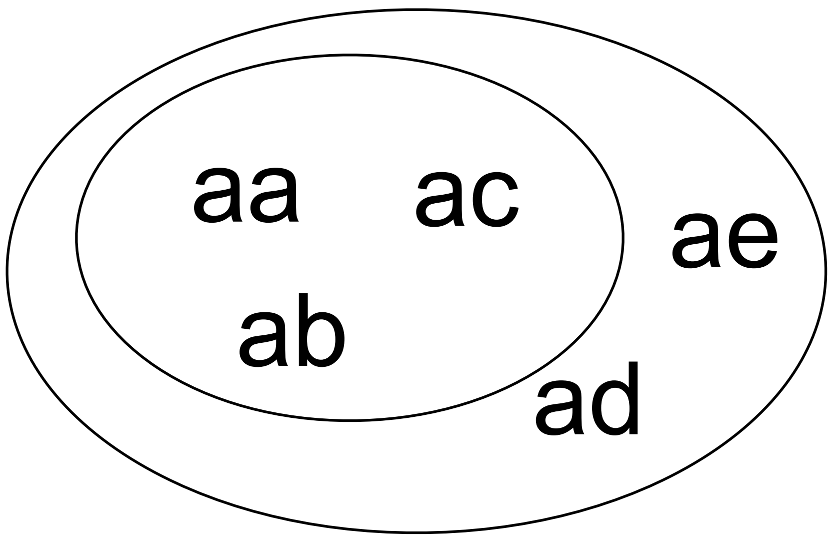
- Works for finite languages? Let's try it …
  - Language 1: {aa,ab,ac}
  - Language 2: {aa,ab,ac,ad,ae}
  - Language 3: {aa,ac}
  - Language 4: {ab}



Mom	aa	ab	ac	ab
Baby	L3	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}

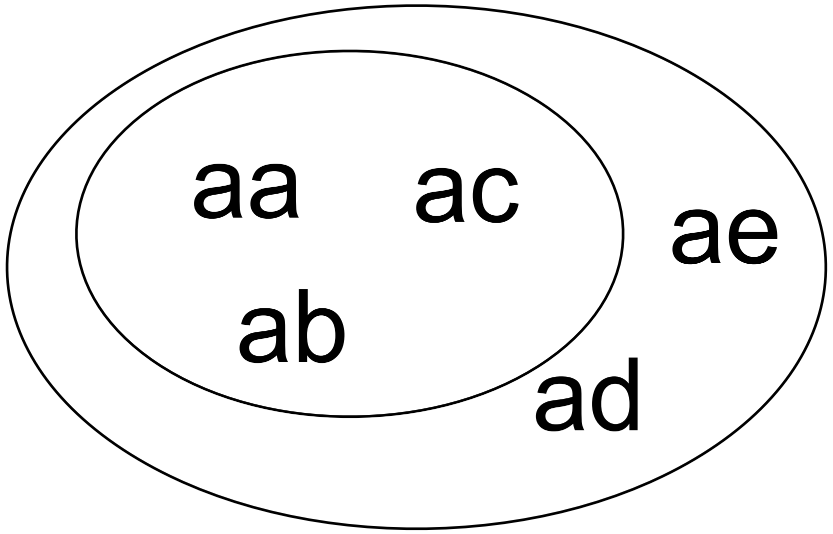


Mom	aa	ab	ac	ab
Baby	L3	L1	L1	L1

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}

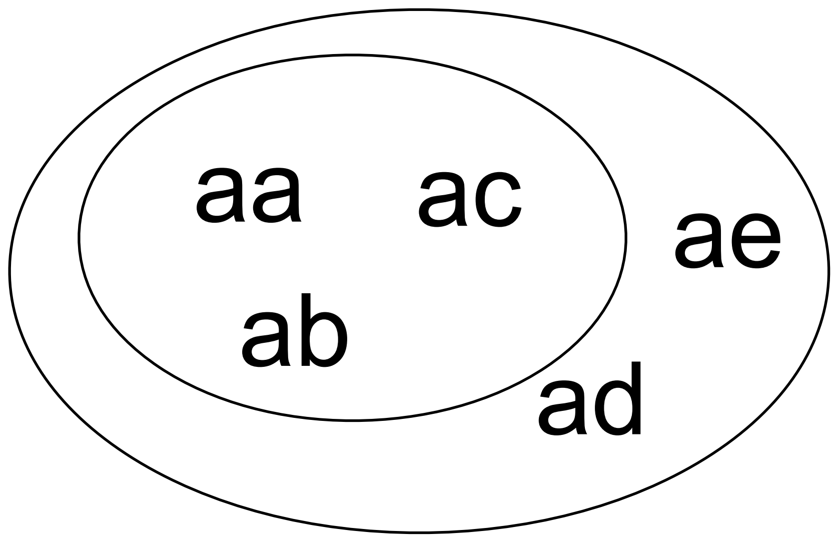


Mom	aa	ab	ac	ab	aa
Baby	L3	L1	L1	L1	

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}

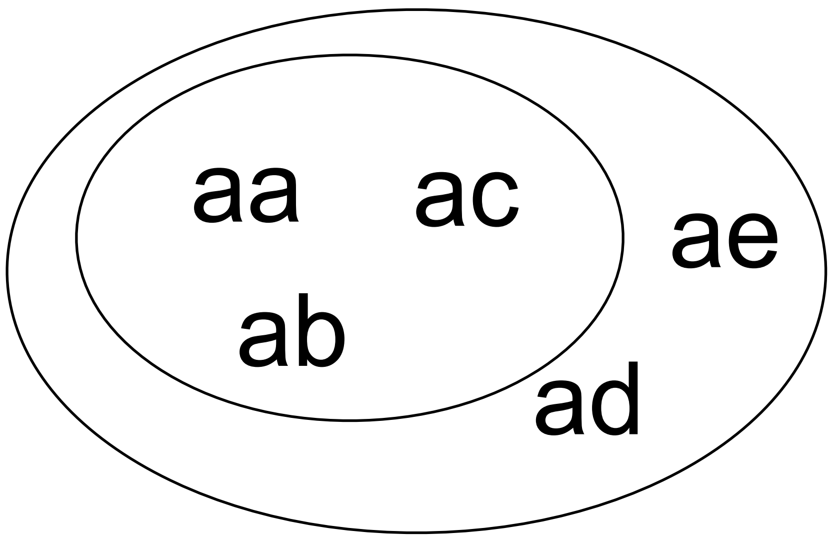


Mom	aa	ab	ac	ab	aa
Baby	L3	L1	L1	L1	L1

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}



Mom	aa	ab	ac	ab	aa	…
Baby	L3	L1	L1	L1	L1	

Evil Mom

- To find out whether Baby is perfect, we have to see whether it gets 100% even in the most adversarial conditions
- Assume Mom is trying to fool Baby
  - although she must speak only sentences from L
  - and she must eventually speak each such sentence
- Does Baby's strategy work?

An Unlearnable Class

- Class of languages:
  - Let L_n = set of all strings of length < n
  - What is L_0?
  - What is L_1?
  - What is L_∞?
    - If the true language is L_∞, can Mom really follow rules?
    - Must eventually speak every sentence of L_∞. Possible?
    - Yes: ε; a, b; aa, ab, ba, bb; aaa, aab, aba, abb, baa, …
  - Our class is C = $\{L_0, L_1, \ldots L_\infty\}$

An Unlearnable Class

An Unlearnable Class

- Let L_n = set of all strings of length < n
  - What is L_0?
  - What is L_1?
  - What is L_∞?

An Unlearnable Class

- Let L_n = set of all strings of length < n
  - What is L_0?
  - What is L_1?
  - What is L_∞?
- Our class is $C = \{L_0, L_1, \ldots L_\infty\}$

An Unlearnable Class

- Let L_n = set of all strings of length < n
  - What is L_0?
  - What is L_1?
  - What is L_∞?
- Our class is $C = \{L_0, L_1, \ldots L_\infty\}$
- A perfect C-baby will distinguish among all of these depending on the input.

An Unlearnable Class

- Let L_n = set of all strings of length < n
  - What is L_0?
  - What is L_1?
  - What is L_∞?
- Our class is C = $\{L_0, L_1, \ldots L_\infty\}$
- A perfect C-baby will distinguish among all of these depending on the input.
- But there is no perfect C-baby …

An Unlearnable Class

An Unlearnable Class

- Our class is $C = \{L_0, L_1, \ldots L_\infty\}$

An Unlearnable Class

- Our class is $C = \{L_0, L_1, \ldots L_\infty\}$
- Suppose Baby adopts conservative strategy, always picking smallest possible language in C.

An Unlearnable Class

- Our class is $C = \{L_0, L_1, \ldots L_\infty\}$
- Suppose Baby adopts conservative strategy, always picking smallest possible language in C.
- So if Mom's longest sentence so far has 75 words, baby's hypothesis is L_{76}.

An Unlearnable Class

- Our class is C = {L_0, L_1, … L_∞}
- Suppose Baby adopts conservative strategy, always picking smallest possible language in C.
- So if Mom’s longest sentence so far has 75 words, baby’s hypothesis is L_{76}.
- This won’t always work: What language can’t a conservative Baby learn?

An Unlearnable Class

An Unlearnable Class

- Our class is $C = \{L_0, L_1, \ldots L_\infty\}$

An Unlearnable Class

- Our class is C = $\{L_0, L_1, \ldots L_\infty\}$
- Could a non-conservative baby be a perfect C-Baby, and eventually converge to any of these?

An Unlearnable Class

- Our class is $C = \{L_0, L_1, \ldots L_\infty\}$
- Could a non-conservative baby be a perfect C-Baby, and eventually converge to any of these?
- **Claim:** *Any* perfect C-Baby must be “quasi-conservative”:

An Unlearnable Class

- Our class is $C = \{L_0, L_1, \ldots L_\infty\}$
- Could a non-conservative baby be a perfect C-Baby, and eventually converge to any of these?
- **Claim:** *Any* perfect C-Baby must be “quasi-conservative”:
  - If true language is L_{76}, and baby posits something else, baby must still eventually come back and guess L_{76} (since it’s perfect).

An Unlearnable Class

- Our class is C = {L_0, L_1, … L_∞}
- Could a non-conservative baby be a perfect C-Baby, and eventually converge to any of these?
- **Claim:** *Any* perfect C-Baby must be "quasi-conservative":
  - If true language is L_{76}, and baby posits something else, baby must still eventually come back and guess L_{76} (since it's perfect).
  - So if longest sentence so far is 75 words, and Mom keeps talking from L_{76}, then eventually baby must actually return to the conservative guess L_{76}.

An Unlearnable Class

- Our class is C = {L_0, L_1, … L_∞}
- Could a non-conservative baby be a perfect C-Baby, and eventually converge to any of these?
- **Claim:** *Any* perfect C-Baby must be “quasi-conservative”:
  - If true language is L_{76}, and baby posits something else, baby must still eventually come back and guess L_{76} (since it’s perfect).
  - So if longest sentence so far is 75 words, and Mom keeps talking from L_{76}, then eventually baby must actually return to the conservative guess L_{76}.
  - Agreed?

Mom's Revenge

If longest sentence so far is 75 words, and Mom keeps talking from L_{76}, then eventually a perfect C-baby must actually return to the conservative guess L_{76}.

- Suppose true language is L_∞.
- Evil Mom can prevent our supposedly perfect C-Baby from converging to it.
- If Baby ever guesses L_∞, say when the longest sentence is 75 words:
  - Then Evil Mom keeps talking from L_{76} until Baby capitulates and revises her guess to L_{76} – as any perfect C-Baby must.
  - So Baby has *not* stayed at L_∞ as required.
- Then Mom can go ahead with longer sentences. If Baby ever guesses L_∞ again, she plays the same trick again.

Mom's Revenge

If longest sentence so far is 75 words, and Mom keeps talking from L_{76}, then eventually a perfect C-baby must actually return to the conservative guess L_{76}.

- Suppose true language is L_∞.
- Evil Mom can prevent our supposedly perfect C-Baby from converging to it.
- If Baby ever guesses L_∞, say when the longest sentence is 75 words:
  - Then Evil Mom keeps talking from L_{76} until Baby capitulates and revises her guess to L_{76} – as any perfect C-Baby must.
  - So Baby has *not* stayed at L_∞ as required.
- Conclusion: There's no perfect Baby that is guaranteed to converge to L_0, L_1, … or L_∞ as appropriate. If it always succeeds on finite languages, Evil Mom can trick it on infinite language.

Implications

Implications

- We found that $C = \{L_0, L_1, \ldots L_\infty\}$ isn't learnable in the limit.

Implications

- We found that C = $\{L_0, L_1, \ldots L_\infty\}$ isn't learnable in the limit.

Implications

- We found that C = $\{L_0, L_1, \ldots L_\infty\}$ isn't learnable in the limit.
- How about class of finite-state languages?

Implications

- We found that C = {L_0, L_1, … L_∞} isn't learnable in the limit.

- How about class of finite-state languages?
  - Not unless you limit it further (e.g., # of states)

Implications

- We found that C = $\{L_0, L_1, \ldots L_\infty\}$ isn't learnable in the limit.

- How about class of finite-state languages?
  - Not unless you limit it further (e.g., # of states)
  - After all, it includes all languages in C, and more, so learner has harder choice

Implications

- We found that C = $\{L_0, L_1, \ldots L_\infty\}$ isn't learnable in the limit.

- How about class of finite-state languages?
  - Not unless you limit it further (e.g., # of states)
  - After all, it includes all languages in C, and more, so learner has harder choice

Implications

- We found that $C = \{L_0, L_1, \ldots L_\infty\}$ isn't learnable in the limit.
- How about class of finite-state languages?
  - Not unless you limit it further (e.g., # of states)
  - After all, it includes all languages in C, and more, so learner has harder choice
- How about class of context-free languages?

Implications

- We found that C = $\{L_0, L_1, \ldots L_\infty\}$ isn't learnable in the limit.

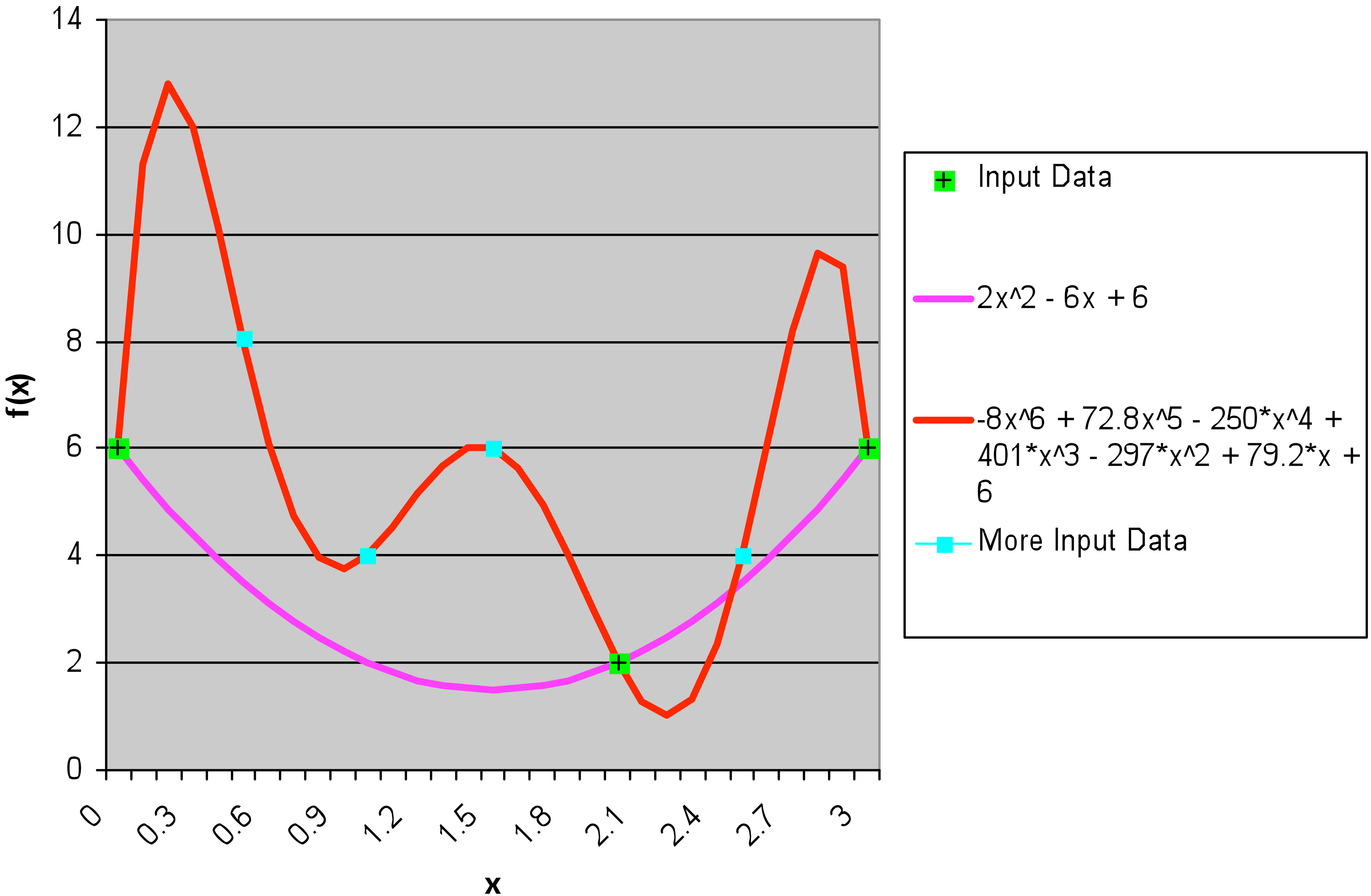
- How about class of finite-state languages?
  - Not unless you limit it further (e.g., # of states)
  - After all, it includes all languages in C, and more, so learner has harder choice

- How about class of context-free languages?
  - Not unless you limit it further (e.g., # of rules)

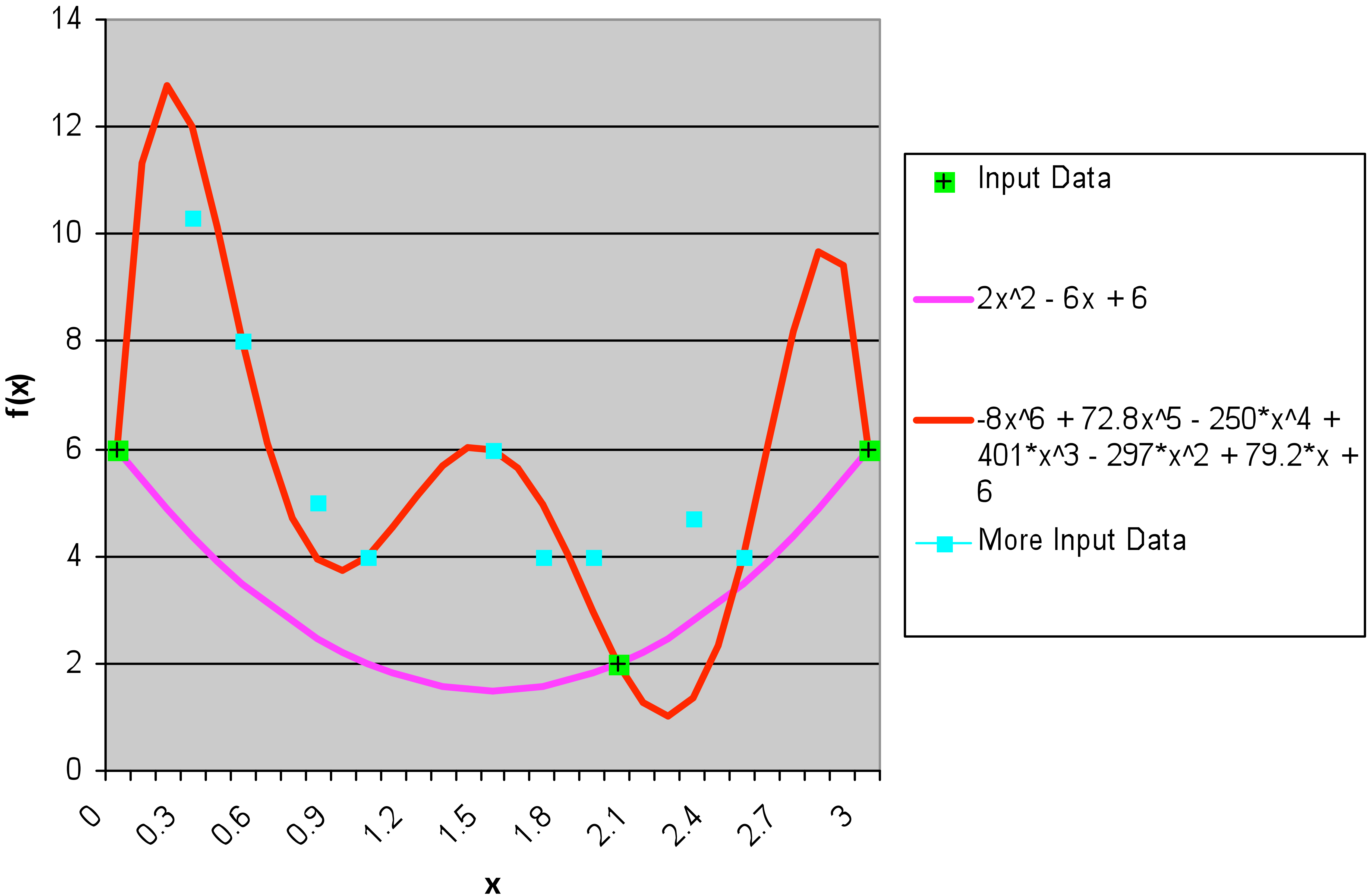
Punchline

- But class of *probabilistic* context-free languages *is* learnable in the limit!!
- If Mom has to output sentences randomly with the appropriate probabilities,
  - she's unable to be too evil
  - there are then perfect Babies that are guaranteed to converge to an appropriate probabilistic CFG
- I.e., from hearing a finite number of sentences, Baby can correctly converge on a grammar that predicts an infinite number of sentences.
  - Baby is generalizing! Just like real babies!

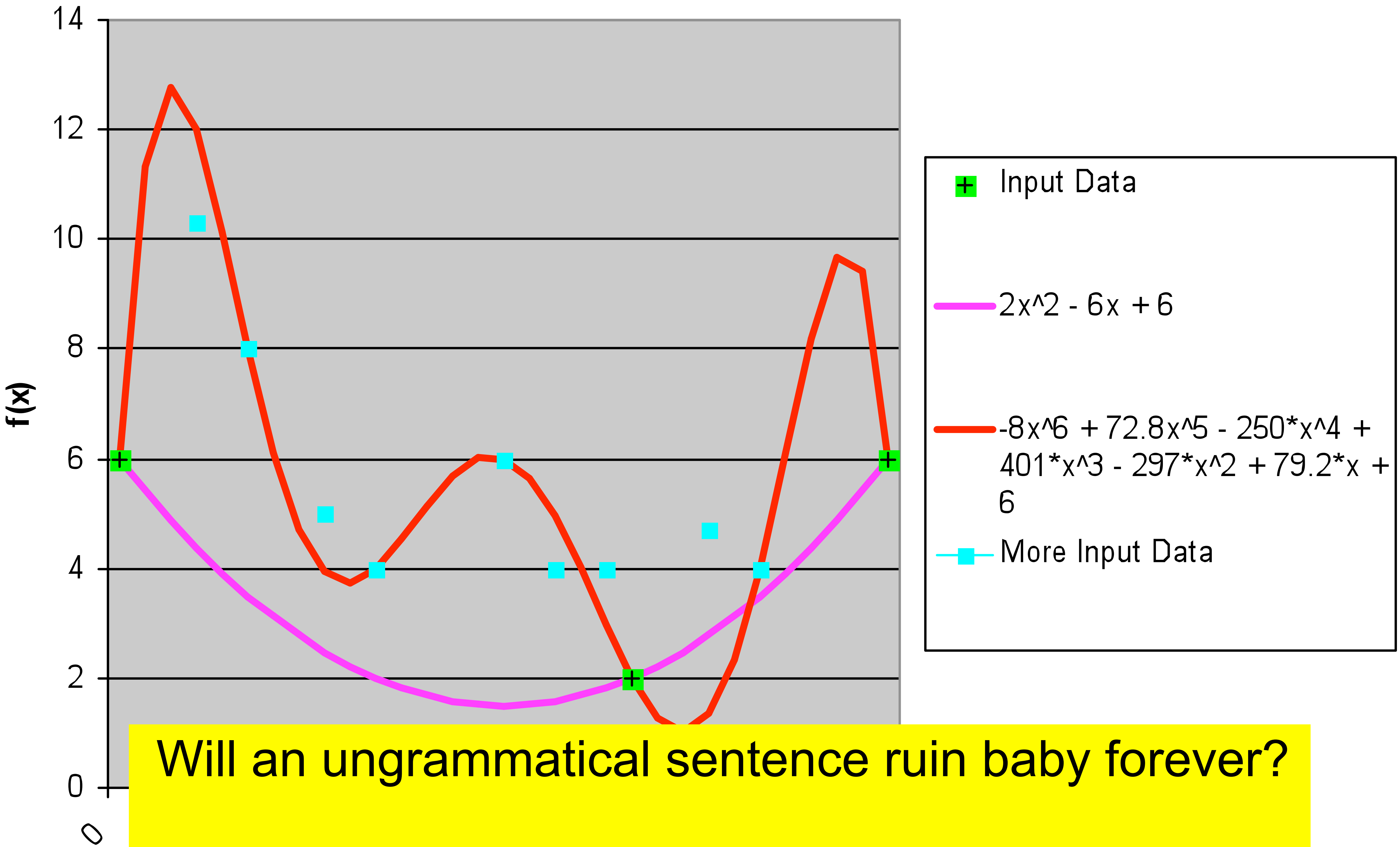
Perfect fit to perfect, incomplete data



Imperfect fit to noisy data

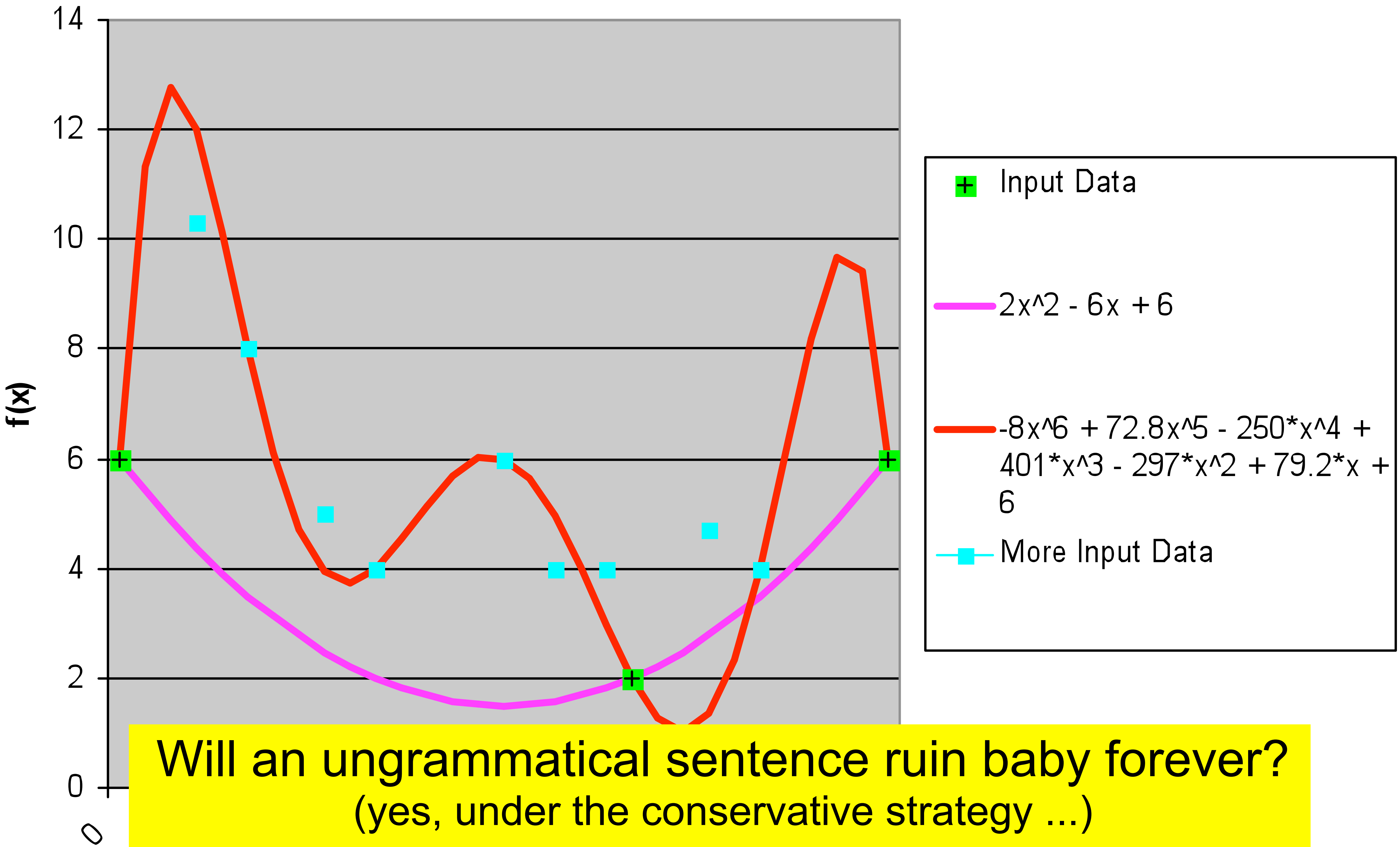


Imperfect fit to noisy data



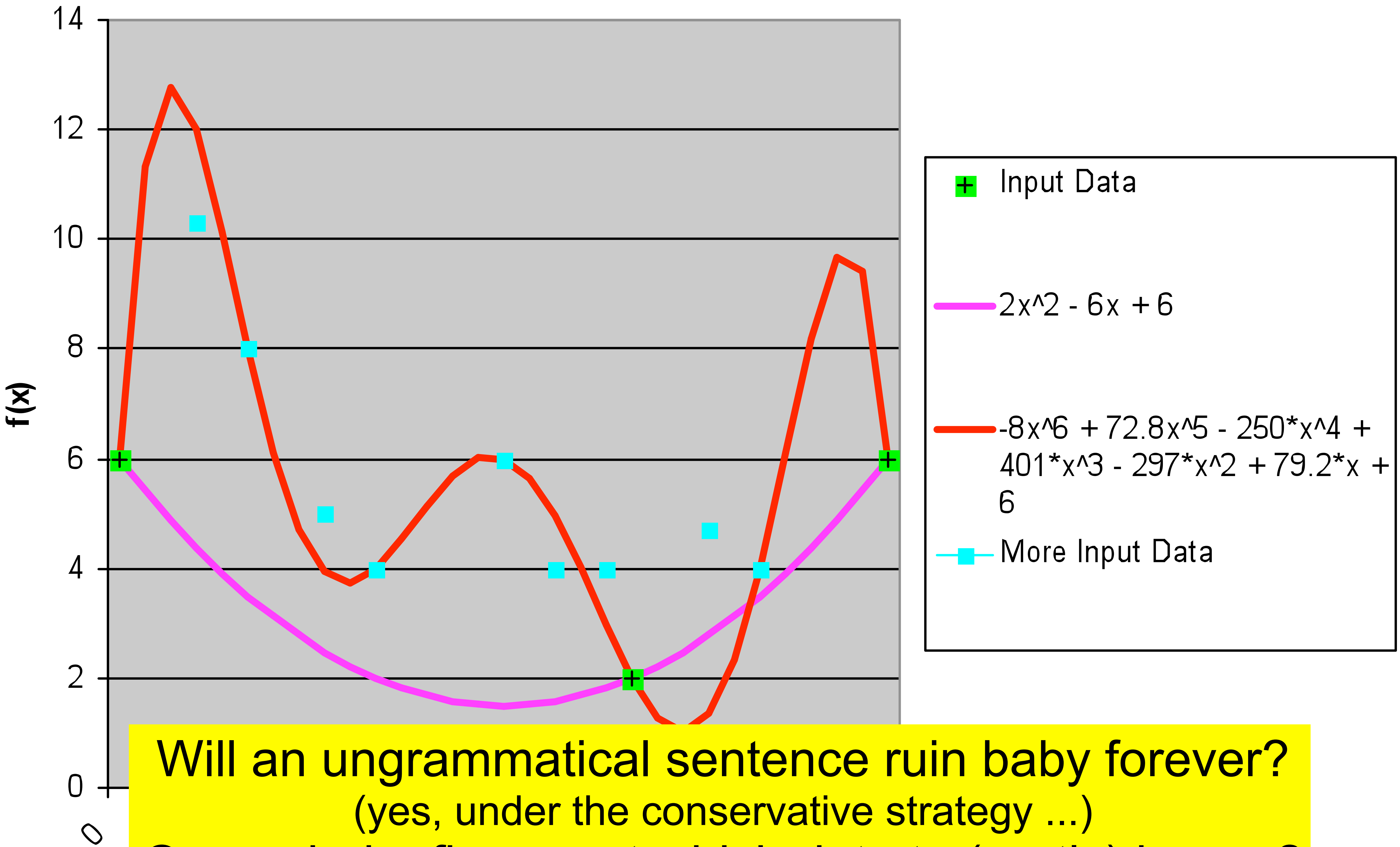
Will an ungrammatical sentence ruin baby forever?

Imperfect fit to noisy data



Will an ungrammatical sentence ruin baby forever?
(yes, under the conservative strategy ...)

Imperfect fit to noisy data

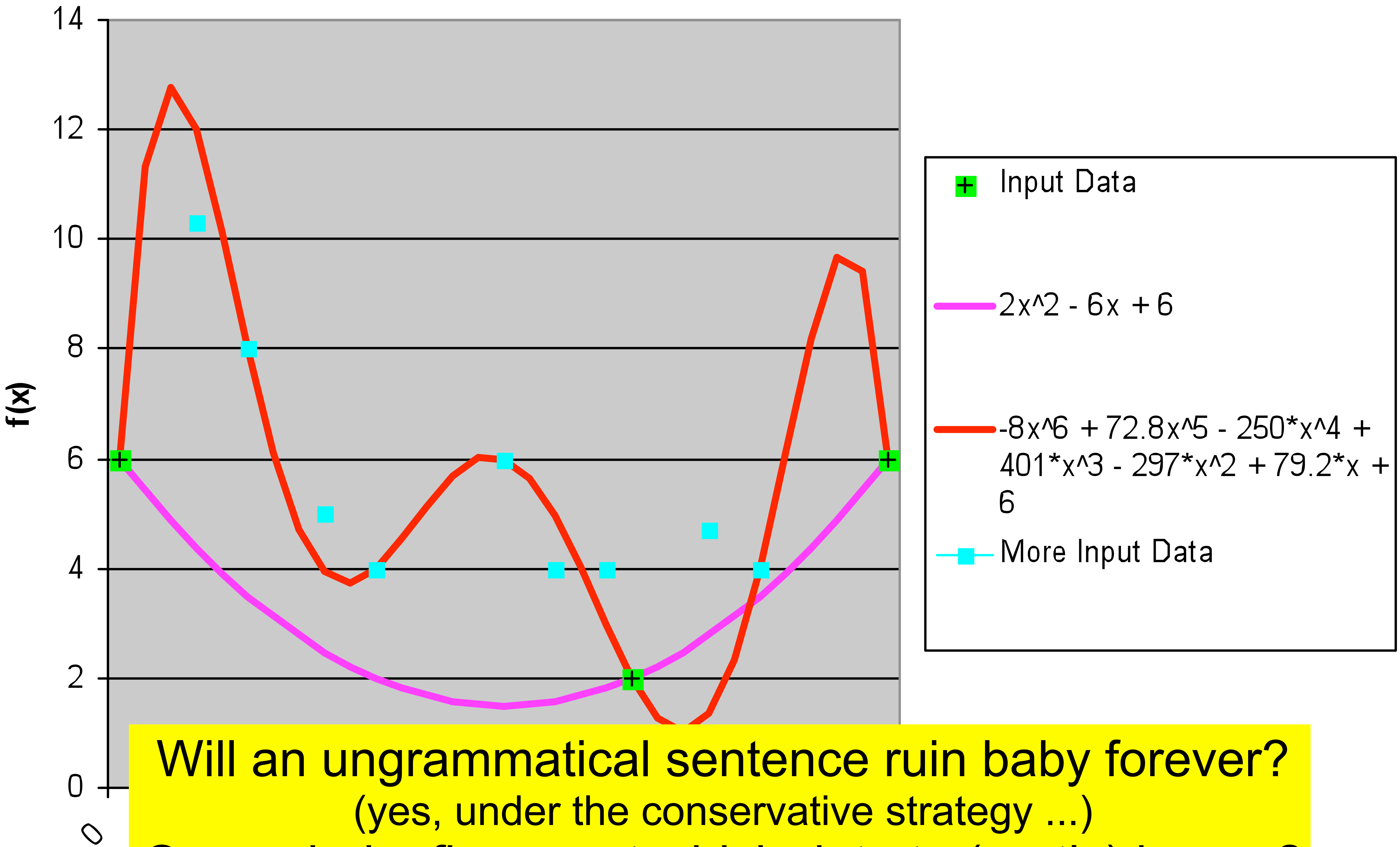


Will an ungrammatical sentence ruin baby forever?

(yes, under the conservative strategy ...)

Or can baby figure out which data to (partly) ignore?

Imperfect fit to noisy data



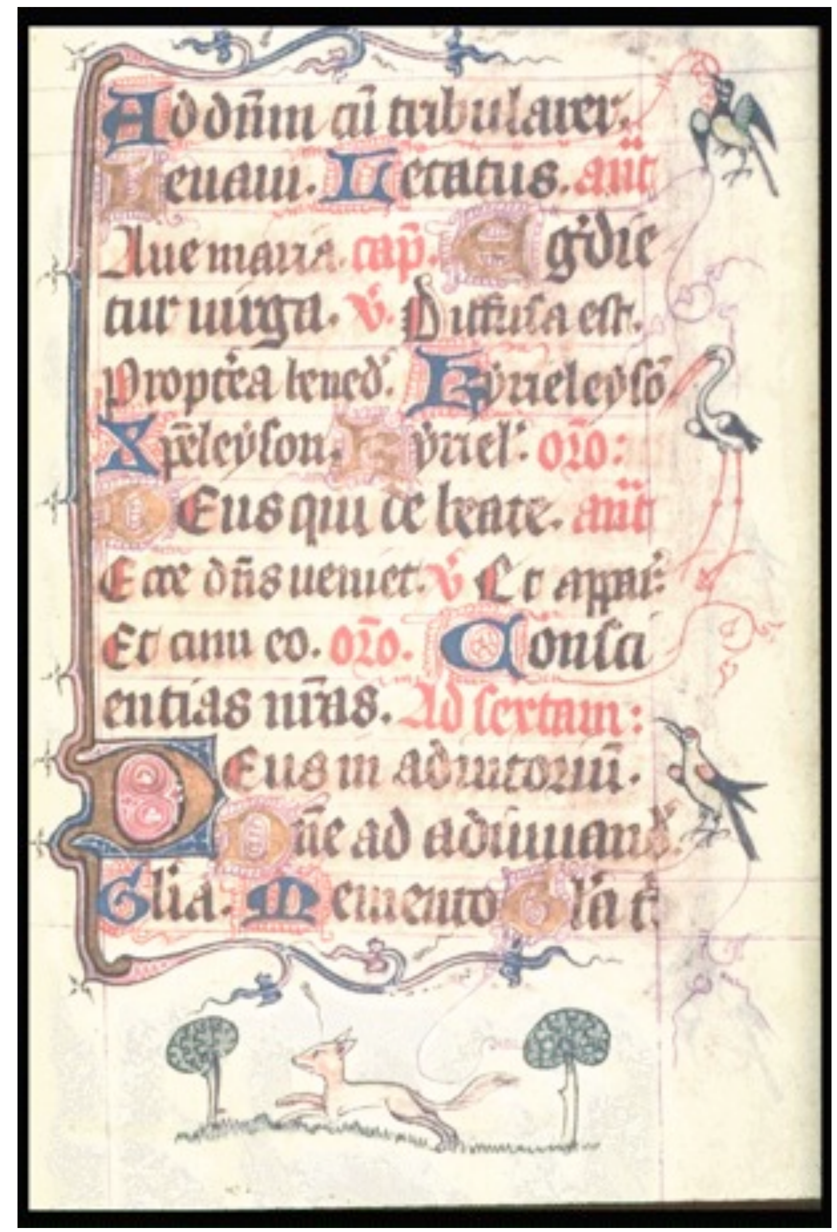
Will an ungrammatical sentence ruin baby forever?
(yes, under the conservative strategy ...)
Or can baby figure out which data to (partly) ignore?
Statistics can help again ... how?

Frequencies and Probabilities in Natural Languages

Chris Manning and others

Models for language

- Human languages are the prototypical example of a symbolic system
- From the beginning, logics and logical reasoning were invented for handling natural language understanding
- Logics and formal languages have a language-like form that draws from and meshes well with natural languages
- Where are the numbers?



29

Dominant answer in linguistic theory: Nowhere

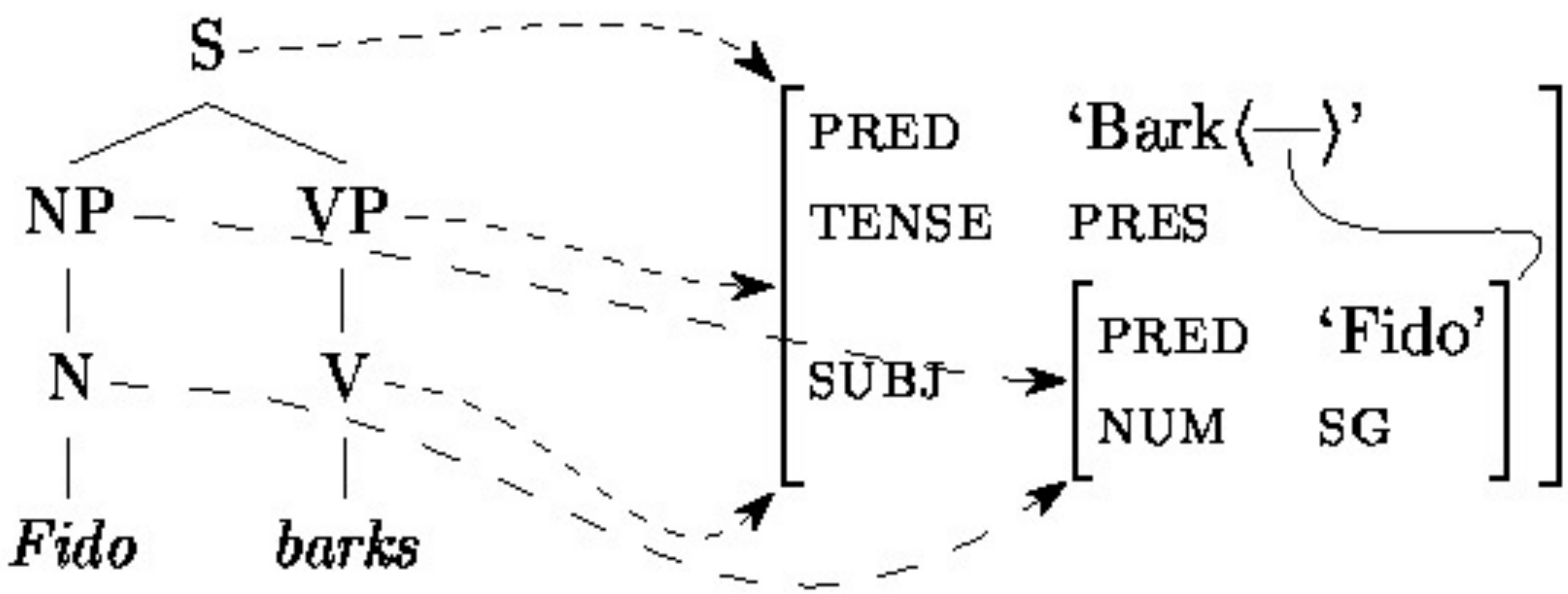
Chomsky again (1969: 57; also 1956, 1957, etc.):

- "It must be recognized that the notion 'probability of a sentence' is an entirely useless one, under any known interpretation of this term."
- Probabilistic models wrongly mix in world knowledge
  - New York vs. Dayton, Ohio
- They don't model grammaticality [also, Tesnière 1959]
  - Colorless green ideas sleep furiously
  - Furiously sleep ideas green colorless

30

Categorical linguistic theories (GB, Minimalism, LFG, HPSG, CG, …)

- Systems of variously rules, principles, and representations is used to describe an infinite set of grammatical sentences of the language
- Other sentences are deemed ungrammatical
- Word strings are given a (hidden) structure



The need for frequencies / probability distributions

The motivation comes from two sides:

- Categorical linguistic theories claim too much:
  - They place a hard categorical boundary of grammaticality, where really there is a fuzzy edge, determined by many conflicting constraints and issues of conventionality vs. human creativity
- Categorical linguistic theories explain too little:
  - They say nothing at all about the soft constraints which explain how people choose to say things
    - Something that language educators, computational NLP people – and historical linguists and sociolinguists dealing with real language – usually want to know about

1. The hard constraints of categorical grammars

- Sentences must satisfy all the rules of the grammar
  - One group specifies the arguments that different verbs take – lexical subcategorization information
    - Some verbs must take objects: *Kim devoured
      [* means ungrammatical]
    - Others do not: *Kim's lip quivered the straw
    - Others take various forms of sentential complements
- In NLP systems, ungrammatical sentences don't parse
- But the problem with this model was noticed early on:
  - "All grammars leak." (Sapir 1921: 38)

Example: verbal clausal subcategorization frames

- Some verbs take various types of sentential complements, given as subcategorization frames:
  - regard: __ NP[acc] as {NP, AdjP}
  - consider: __ NP[acc] {AdjP, NP, VP[inf]}
  - think: __ CP[that]; __ NP[acc] NP
- **Problem:** in context, language is used more flexibly than this model suggests
  - Most such subcategorization ‘facts’ are **wrong**

34

Standard subcategorization rules (Pollard and Sag 1994)

- We consider Kim to be an acceptable candidate
- We consider Kim an acceptable candidate
- We consider Kim quite acceptable
- We consider Kim among the most acceptable candidates
- *We consider Kim as an acceptable candidate
- *We consider Kim as quite acceptable
- *We consider Kim as among the most acceptable candidates
- ?*We consider Kim as being among the most acceptable candidates

35

Subcategorization facts from The New York Times

Consider as:

- The boys consider her as family and she participates in everything we do.
- Greenspan said, "I don't consider it as something that gives me great concern.
- "We consider that as part of the job," Keep said.
- Although the Raiders missed the playoffs for the second time in the past three seasons, he said he considers them as having championship potential.
- Culturally, the Croats consider themselves as belonging to the "civilized" West, ...

36

More subcategorization facts: regard

Pollard and Sag (1994):

- *We regard Kim to be an acceptable candidate
- We regard Kim as an acceptable candidate

The New York Times:

- As 70 to 80 percent of the cost of blood tests, like prescriptions, is paid for by the state, neither physicians nor patients regard expense to be a consideration.
- Conservatives argue that the Bible regards homosexuality to be a sin.

More subcategorization facts: turn out and end up

Pollard and Sag (1994):

- Kim turned out political
- *Kim turned out doing all the work

The New York Times:

- But it turned out having a greater impact than any of us dreamed.

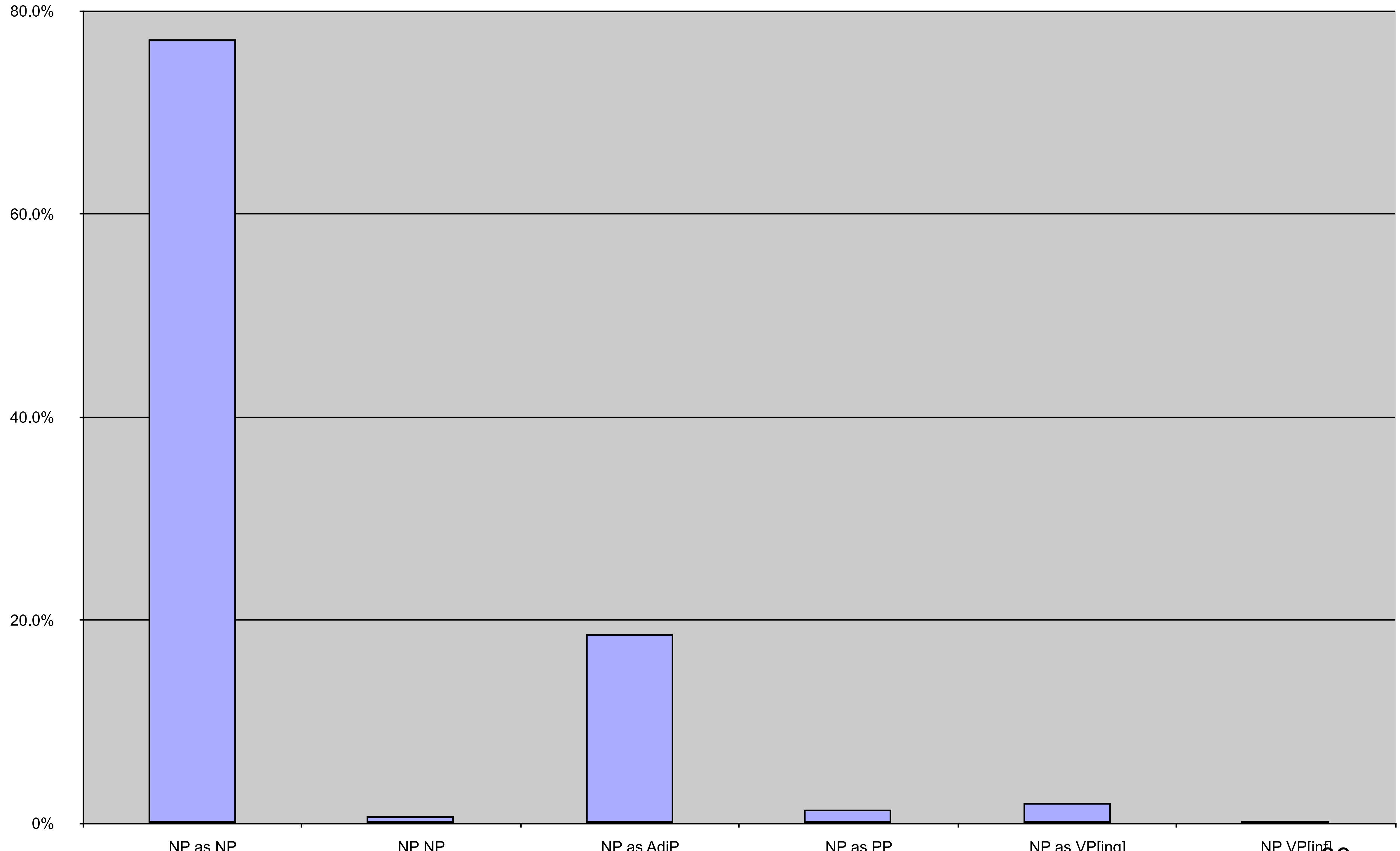
Pollard and Sag (1994):

- Kim ended up political
- *Kim ended up sent more and more leaflets

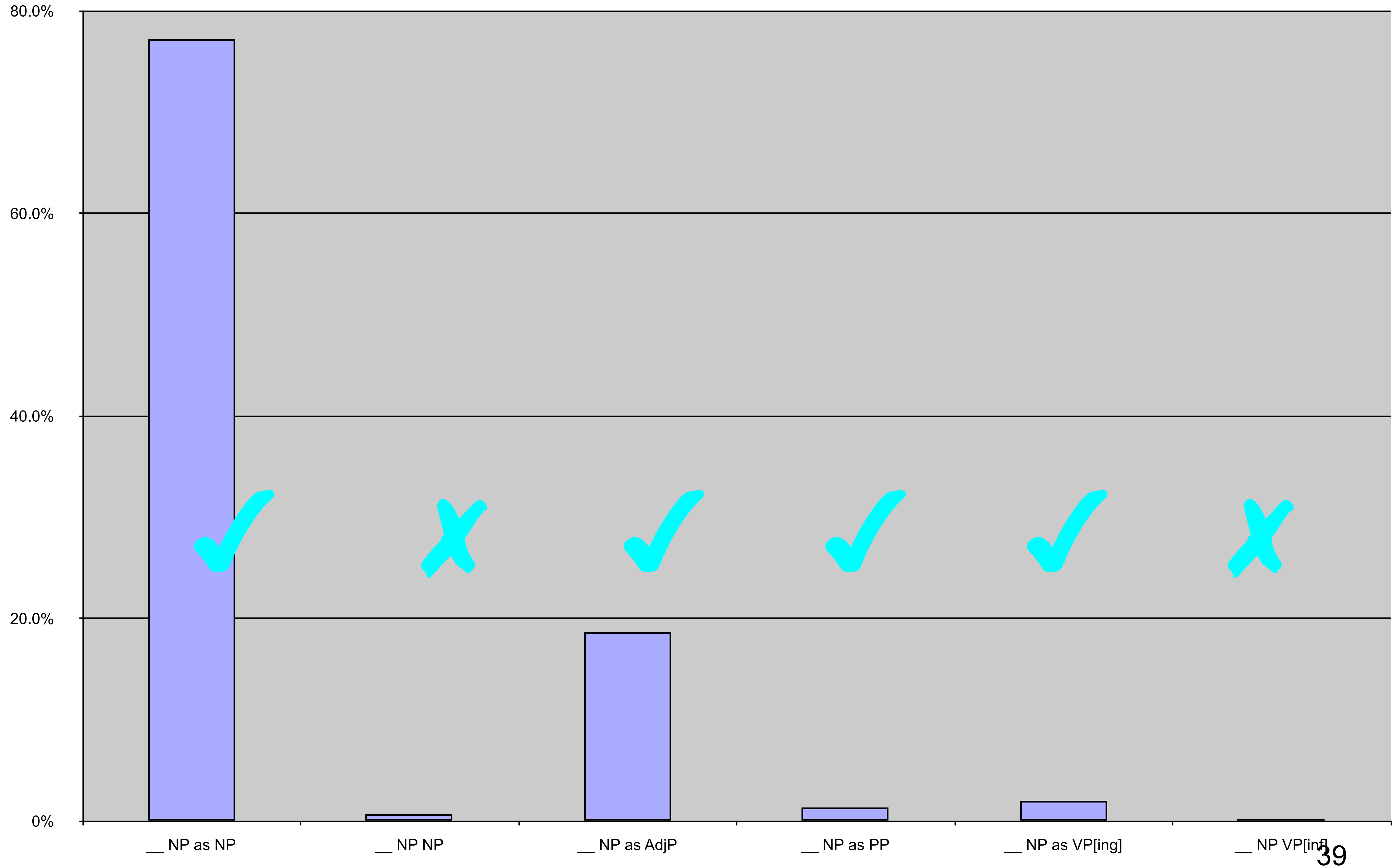
The New York Times:

- On the big night, Horatio ended up flattened on the ground like a fried egg with the yolk broken.

Probability mass functions: subcategorization of regard



Probability mass functions: subcategorization of regard



39

Leakage leads to change

- People continually stretch the 'rules' of grammar to meet new communicative needs, to better align grammar and meaning, etc.
- As a result language slowly changes
  - **while:** used to be only a noun (That takes a while); now mainly used as a subordinate clause introducer (While you were out)
  - **e-mail:** started as a mass noun like mail (most junk e-mail is annoying); it's moving to be a count noun (filling the role of e-letter): I just got an interesting email about that.

40

Blurring of categories: “Marginal prepositions”

- An example of blurring in syntactic category during linguistic change is so-called ‘marginal prepositions’ in English, which are moving from being participles to prepositions
- Some still clearly maintain a verbal existence, like following, concerning, considering; for some it is marginal, like according, excepting; for others their verbal character is completely lost, such as during [cf. endure], pending, notwithstanding.

Verb (VBG) ➡ Preposition IN

As verbal participle, understood subject agrees with noun:

- They moved slowly, toward the main gate, following the wall
- Repeat the instructions following the asterisk

A temporal use with a controlling noun becomes common:

- This continued most of the week following that ill-starred trip to church

Prep. uses (meaning is after, no controlling noun) appear

- He bled profusely following circumcision
- Following a telephone call, a little earlier, Winter had said …

42

Mapping the recent change of following: participle ➡ prep.

- Fowler (1926): “there is a continual change going on by which certain participles or adjectives acquire the character of prepositions or adverbs, no longer needing the prop of a noun to cling to … [we see] a development caught in the act”
- Fowler (1926) -- no mention of following in particular
- Fowler [Gowers] (1948): “Following is not a preposition. It is the participle of the verb follow and must have a noun to agree with”
- Fowler [Gowers] (1954): generally condemns temporal usage, but says it can be justified in certain circumstances

2. Explaining more: What do people say?

- What people do say has two parts:
  - Contingent facts about the world
    - People in Amherst have talked a lot about snow falling, not stocks falling, lately
  - The way speakers choose to express ideas using the resources of their language
    - People don’t often put that clauses pre-verbally:
      - That we will have to revise this program is almost certain
- The latter is properly part of people’s Knowledge of Language. Part of linguistics.

What do people say?

- Simply delimiting a set of grammatical sentences provides only a very weak description of a language, and of the ways people choose to express ideas in it
- Probability densities over sentences and sentence structures can give a much richer view of language structure and use
- In particular, we find that the same soft generalizations and tendencies of one language often appear as (apparently) categorical constraints in other languages
- A syntactic theory should be able to uniformly capture these constraints, rather than only recognizing them when they are categorical

45

Example: Bresnan, Dingare & Manning

- Project modeling English diathesis alternations (active/passive, locative inversion, etc.)
- In some languages passives are categorically restricted by person considerations:
  - In Lummi (Salishan, Washington state), 1/2 person must be the subject if other argument is 3rd person. There is variation if both arguments are 3rd person. (Jelinek and Demers 1983) [cf. also Navajo, etc.]
    - *That example was provided by me
    - *He likes me
    - ✓I am liked by him

Bresnan, Dingare & Manning

- In English, there is no such categorical constraint, but we can still see it at work as a soft constraint.
- Collected data from verbs with an agent and patient argument (canonical transitives) from treebanked portions of the Switchboard corpus of conversational American English, analyzing for person and act/pass

	Active	Passive
1/2 Ag, 1/2 Pt	158	0 (0.0%)
1/2 Ag, 3 Pt	5120	1 (0.0%)
3 Ag, 1/2 Pt	552	16 (2.8%)
3 Ag, 3 Pt	3307	46 (1.4%)

47

Bresnan, Dingare & Manning

- While person is only a small part of the picture in determining the choice of active/passive in English (information structure, genre, etc. is more important), there is nonetheless a highly significant (X^2 $p < 0.0001$) effect of person on active/passive choice
- The exact same hard constraint of Lummi appears as a soft constraint in English
- This behavior is predicted by the universal hierarchies within our stochastic OT model (which extends existing OT approaches to valence – Aissen 1999, Lødrup 1999)
- Conversely linguistic model predicts that no "anti-English" [which is just the opposite] exists

48

Conclusions

- There are many phenomena in language that cry out for non-categorical and probabilistic modeling and explanation
- Probabilistic models can be applied on top of one's favorite sophisticated linguistic representations!
- Frequency evidence can enrich linguistic theory by revealing soft constraints at work in language use

49