Looking Beyond: What Indirect Evidence Can Tell Us About Universal Grammar

Lisa S. Pearl

Computation of Language Laboratory
UC Irvine

Workshop on Language, Cognition, and Computation &
Workshop on Language, Variation, and Change
University of Chicago
March 4, 2011

An induction problem by any other name…

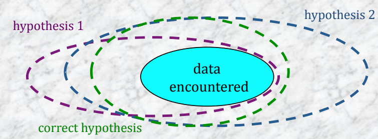
One of the most controversial claims in linguistics is that children face an induction problem:

"Poverty of the Stimulus" (Chomsky 1980, Crain 1991, Lightfoot 1989, Valian 2009)

"Logical Problem of Language Acquisition" (Baker 1981, Hornstein & Lightfoot 1981)

"Plato's Problem" (Chomsky 1988, Dresher 2003)

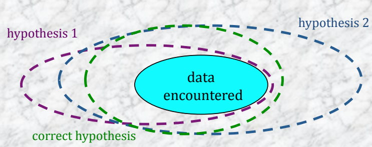
Basic claim:
The data encountered are compatible with multiple hypotheses.



The induction problem

Extended claim:
Given this, the data are insufficient for identifying the correct hypothesis as quickly as children do (Legate & Yang 2002) – or at all.

Big question: How do children do it, then?



One answer: Children come prepared

- Children are not unbiased learners.
- But if children come equipped with helpful learning biases, then what is the nature of these necessary biases?
  - Are they innate or derived from the input somehow?
  - Are they domain-specific or domain-general?
  - Are they about what's being learned or about how to learn?

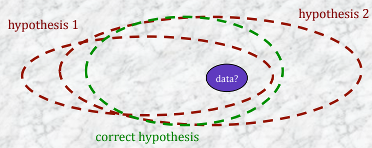
The Universal Grammar (UG) hypothesis (Chomsky 1965, Chomsky 1975):
These biases are innate and domain-specific.

Induction problems, UG, and informative data

Traditional Idea

Induction problems → Universal Grammar (UG)

Traditional assumption:
Only directly related data are informative data. These data are often rare, and that's why induction problems occur.



The direct evidence assumption

If you want to learn linguistic knowledge L, you learn it by observing examples of L in your input (and possibly by also being sensitive to indirect negative evidence about what examples are missing from the input.)

Learning complex yes/no questions

Direct evidence L:
"Is the boy who is in the corner t_{is} happy?"

Possible indirect negative evidence:
*"Is the boy who t_{is} in the corner is happy?"

The direct evidence assumption

If you want to learn linguistic knowledge L, you learn it by observing examples of L in your input (and possibly by also being sensitive to indirect negative evidence about what examples are missing from the input.)

Learning the representation of English anaphoric one

Situation

Direct evidence L:
"I see a red bottle… but there isn't another one around."

Possible indirect negative evidence:
*"I like the student of linguistics and he likes the one of computer science."

The direct evidence assumption

If you want to learn linguistic knowledge L, you learn it by observing examples of L in your input (and possibly by also being sensitive to indirect negative evidence about what examples are missing from the input.)

Learning syntactic islands

Direct evidence L:
"What did the teacher think t_{what} inspired the students?"
"Who did the teacher think the letter from the soldier inspired t_{who}?"
"Who t_{who} thought the letter from the soldier inspired the students?"

Possible indirect negative evidence:
*"Who did the teacher think [[the letter from t_{who}] inspired the students]?"

island

The direct evidence assumption

If you want to learn linguistic knowledge L, you learn it by observing examples of L in your input (and possibly by also being sensitive to indirect negative evidence about what examples are missing from the input.)

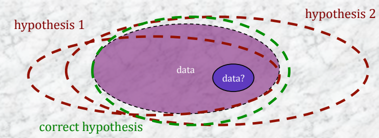
Learning about the distribution of noun phrases (what Case Theory explains):

Direct evidence L:
"**John** seems to be clever."
"**John** tries to be clever."
"It seems **John** is clever."

Possible indirect negative evidence:
*"It tries **John** is clever."

A broader set of informative data

Indirect evidence: other kinds of data that may also be relevant, thereby broadening the set of informative data



Recent computational models have been exploring this:

- Complex yes/no questions (Perfors, Tenenbaum, & Regier 2006, 2011)
- Anaphoric one (Regier & Gahl 2004, Pearl & Lidz 2009, Foraker et al. 2009)

Mapping out UG & the acquisition process

Big questions:

- When induction problems exist, what does it take to solve them?
  - What indirect evidence is available? How might a child leverage this evidence?
  - What learning biases can get the job done, given the available data? Are they necessarily innate and domain-specific (UG)?
- How can the necessary learning biases inform us about how the acquisition process works?

Three studies of indirect evidence at UC Irvine: "Testing the Universal Grammar Hypothesis"

Today

Learning the representation of English anaphoric one
*"I like the student of linguistics and he likes the one of computer science."

Learning syntactic islands
*"Who did the teacher think [[the letter from t_{who}] inspired the students]?"

Learning about the distribution of noun phrases (what Case Theory explains):
*"It tries **John** is clever."

Road Map

- Adult and child knowledge states for anaphoric *one*
- The learning problem, given the available data
- Previous proposals for how to solve this problem
- A broader view of informative data
- Representing the information in the data
- An online probabilistic learning framework
- Results & implications

Anaphoric *One*

Look - a red bottle!

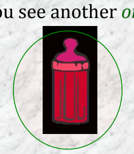
Do you see another *one*?

Anaphoric *One*

Look - a red bottle!

red bottle

Do you see another *one*?



Process: First determine the antecedent of *one* (what string *one* is referring to). →"red bottle"

Anaphoric *One*

Look - a red bottle!

red bottle

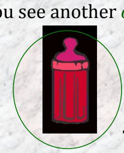
Do you see another *one*?

Process: Because the antecedent ("red bottle") includes the modifier "red", the property RED is important for the referent of *one* to have.
→ referent of *one* = RED BOTTLE

Anaphoric *One*

Look - a red bottle!

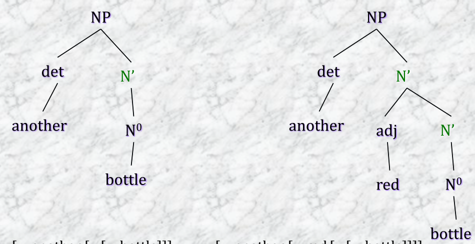
Do you see another *one*?



Two steps:
(1) Identify syntactic antecedent (based on syntactic category of *one*)
(2) Identify semantic referent (based on syntactic antecedent)

Anaphoric *One*: Syntactic Category

Standard linguistic theory claims that *one* in these kind of utterances is a syntactic category smaller than an entire noun phrase, but larger than just a noun (N^0). This category is sometimes called N'. This category includes strings like "bottle" and "red bottle".

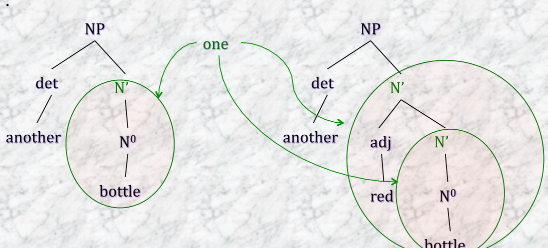


$[_{NP}$ another $[_{N'}$ $[_{N0}$ bottle]]] $[_{NP}$ another $[_{N'}$ red $[_{N'}$ $[_{N0}$ bottle]]]]

Anaphoric *One*: Syntactic Category

Standard linguistic theory claims that *one* in these kind of utterances is a syntactic category smaller than an entire noun phrase, but larger than just a noun (N^0). This category is sometimes called N'. This category includes strings like "bottle" and "red bottle".

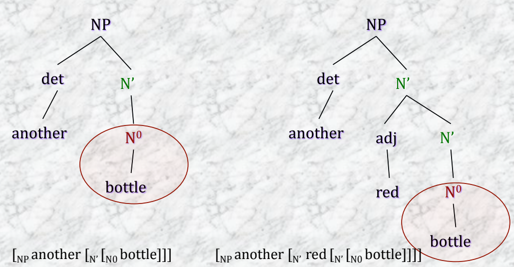
.



$[_{NP}$ another $[_{N'}$ $[_{N0}$ bottle]]] $[_{NP}$ another $[_{N'}$ red $[_{N'}$ $[_{N0}$ bottle]]]]

Anaphoric *One*: Syntactic Category

Importantly, *one* is not N^0. If it was, it could only have strings like "bottle" as its antecedent, and could never have strings like "red bottle" as its antecedent.



$[_{NP}$ another $[_{N'}$ $[_{N0}$ bottle]]] $[_{NP}$ another $[_{N'}$ red $[_{N'}$ $[_{N0}$ bottle]]]]

Anaphoric *One*: Interpretations based on Syntactic Category

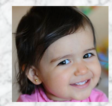
If *one* was N^0, we would have a different interpretation of

"Look – a red bottle! Do you see another one?"

Because *one*'s antecedent could only be "bottle", we would interpret the second part as "Do you see another *bottle*?" and the purple bottle would be a fine referent for *one*.

Since *one*'s antecedent is "red bottle", and "red bottle" cannot be N^0, *one* must not be N^0.

Anaphoric *One*: Children's Knowledge

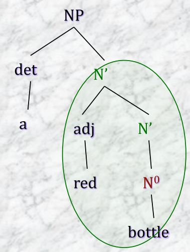


Lidz, Waxman, & Freedman (2003) [LWF] found that 18-month-olds have a preference for the red bottle in the same situation.

"Look – a red bottle! Do you see another one?"

LWF interpretation & conclusion:

Preference for the RED BOTTLE means the preferred syntactic antecedent is "red bottle".



LWF conclude that 18-month-old knowledge =

syntactic category of *one* = N'

syntactic antecedent when modifier is present includes modifier (e.g., red) = referent has modifier property

Anaphoric *One*: The induction problem

Acquisition: Children must learn the right syntactic category for *one*, and the right interpretation preference for *one* in situations with more than one option.

Anaphoric *One*: The induction problem

Acquisition: Children must learn the right syntactic category for *one*, and the right interpretation preference for *one* in situations with more than one option.

Problem: Most data children encounter are ambiguous.

Syntactically (SYN) ambiguous data:

"Look – a bottle! Oh, look – another one."

one's referent = BOTTLE

one's antecedent = $[_{N'}[_{N0}$ bottle$]]$ or $[_{N0}$ bottle$]$?

Anaphoric *One*: The induction problem

Acquisition: Children must learn the right syntactic category for *one*, and the right interpretation preference for *one* in situations with more than one option.

Problem: Most data children encounter are ambiguous.

Semantically and syntactically (SEM-SYN) ambiguous:

"Look – a red bottle! Oh, look – another one."

one's referent = RED BOTTLE or BOTTLE?

one's antecedent = $[_{N'}\text{ red}[_{N'}[_{N0}\text{ bottle}]]]$ or $[_{N'}[_{N0}\text{ bottle}]]$ or $[_{N0}\text{ bottle}]$?

Anaphoric *One*: The induction problem

Acquisition: Children must learn the right syntactic category for *one*, and the right interpretation preference for *one* in situations with more than one option.

Problem: Unambiguous data are rare (<0.25%, based on LWF's analysis)

Unambiguous (UNAMB) data:

"Look – a red bottle! Hmmm - there doesn't seem to be another one here, though."

one's referent = BOTTLE? If so, *one*'s antecedent = "bottle".

But it's strange to claim there's not another *bottle* here.

So, *one*'s referent must be RED BOTTLE, and *one*'s antecedent = $[_{N'}\text{ red}[_{N'}[_{N0}\text{ bottle}]]]$.

Previous proposals for learning about *one*

Baker (1978) [**Baker**] (also Hornstein & Lightfoot 1981, Lightfoot 1982, Hamburger & Crain 1984, Crain 1991): Only unambiguous data are informative. Because they're so rare, they can't be responsible for the acquisition of *one*.

How then?

Children have innate, domain-specific knowledge restricting the hypotheses about *one*: *one* cannot be syntactic category N^0.

What about when there are multiple N' antecedents?

$[_{N'}\text{ red}[_{N'}[_{N0}\text{ bottle}]]]$ or $[_{N'}[_{N0}\text{ bottle}]]$?

(No specific proposal for this.)

Previous proposals for learning about *one*

Regier & Gahl 2004 [**R&G**]: Sem-Syn ambiguous data can be leveraged, in addition to using unambiguous data.

"Look – a red bottle! Oh, look – another one!" 

How?

Use innate domain-general statistical learning abilities to track how often *one*'s referent has the mentioned property (e.g. *red*). If the referent often has the property (RED BOTTLE), this is a suspicious coincidence unless the antecedent really does include the modifier ("red bottle") and *one*'s category is N'.

$[_{N'}\text{ red}[_{N'}[_{N0}\text{ bottle}]]]$

Previous proposals for learning about *one*

Pearl & Lidz 2009 [**P&L**]: Syn ambiguous data must not be leveraged, even if Sem-Syn and unambiguous data are used.

"Look – a bottle! Oh, look – another one!"

Why?

These data cause an "equal-opportunity" (EO) probabilistic learner to think *one*'s category is N^0.

$[_{N0}$ bottle$]$

How?

P&L propose a domain-specific learning bias to ignore just these ambiguous data, though they speculate how this bias could be derived from an innate domain-general preference for learning when there is local uncertainty.

Previous proposals for learning about *one*

Foraker et al. 2009 [**F&al**]: Leverage the syntactic distribution of *one* with innate domain-general statistical learning, by using subtle domain-specific semantic distinctions that indicate syntactic category.

"ball with stripes"	"side of the road"
"one with dots"	*"one of the river"
[modifiers]	[complements = conceptually evoked by head noun]
[head noun = N']	[head noun = N^0]

How?

Indirect negative evidence (never seeing *one* with a complement, even though other nouns take complements) indicates one is not N^0.

A new proposal: Broadening the data set

Pearl & Mis, submitted [**P&M**]: Other pronouns in the language can also be used anaphorically: *him, her, it, …*

Look at the cute penguin. I want to hug him/her/it.

$[_{NP}$ the $[_{N'}$ cute $[_{N'}$ $[_{N0}$ penguin$]]]]$ → $[_{NP}$ him/her/it$]$

Look! A cute penguin. I want one.

$[_{NP}$ a $[_{N'}$ cute $[_{N'}$ $[_{N0}$ penguin$]]]]$ → $[_{NP}$ one$]$

Note: The issue of *one*'s category only occurs when *one* is used in a syntactic environment that indicates it is smaller than an NP (**<NP**).

A new proposal: Broadening the data set

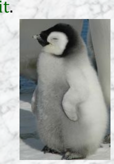
Pearl & Mis, submitted [**P&M**]: Track how often the referent of the anaphoric element *(one, him, her, it*, etc.) has the property mentioned in the potential antecedent, using innate domain-general statistical learning abilities.

Important: This applies, even when the syntactic category is known.

Look at the cute penguin. I want to hug him/her/it.

Look! A cute penguin. I want one.

Is the referent cute? Yes!
So it's important that the antecedent include the modifier "cute".



A new proposal: Broadening the data set

Pearl & Mis, submitted **[P&M]**: Track how often the referent of the anaphoric element *(one, him, her, it,* etc.) has the property mentioned in the potential antecedent, using innate domain-general statistical learning abilities.

Important: This applies, even when the syntactic category is known.

Look at the cute penguin. I want to hug him/her/it.

Look! A cute penguin. I want one.

Data points like those above will always include the modifier in the antecedent, since the category of the pronoun is NP and so the antecedent is the entire NP. These data are unambiguous: the referent must have the mentioned property.

Data set comparisons: Learners using syntactic and semantic information

Unamb <NP
"Look – a red bottle! Hmmm - there doesn't seem to be another *one* here, though."

Learners: **Baker, R&G, P&L's EO, P&M**

Sem-Syn Amb
"Look – a red bottle! Oh, look – another *one*!"

Learners: **R&G, P&L's EO, P&M**

Syn Amb
"Look – a bottle! Oh, look – another *one*!"

Learners: **P&L's EO, P&M**

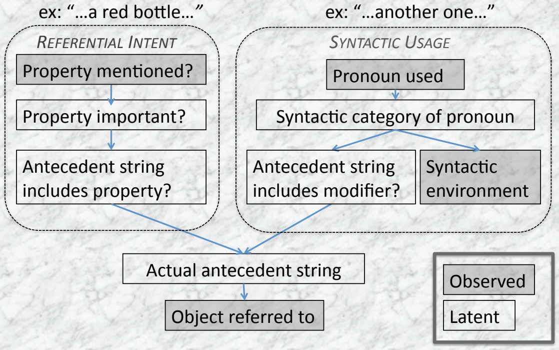
Unamb NP
"Look – a red bottle! I want *one/it*."

Learners: **P&M**

Information in the data

previous context = ex: "...a red bottle..."

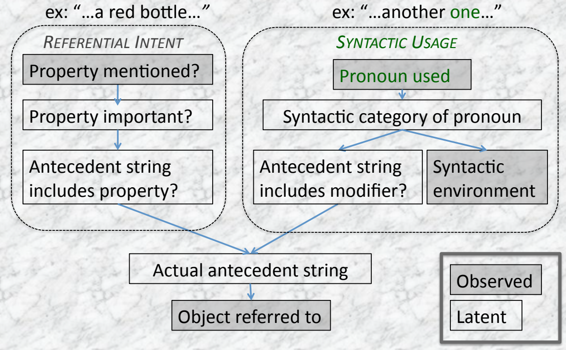
current usage = pronoun ex: "...another one..."

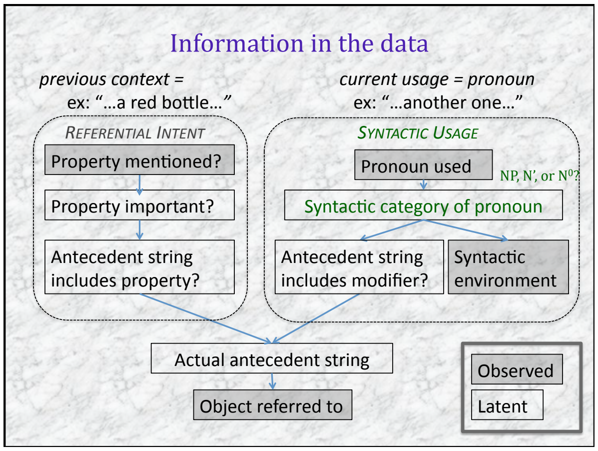


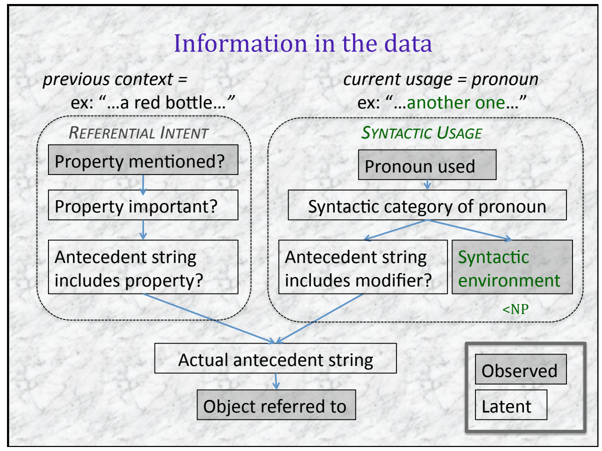
Information in the data

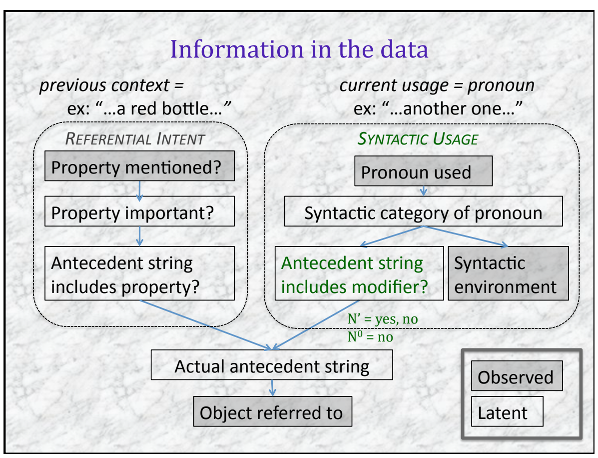
previous context = ex: "...a red bottle..."

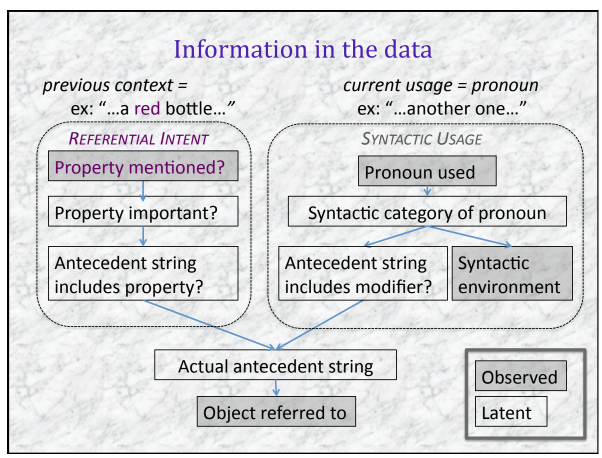
current usage = pronoun ex: "...another one..."

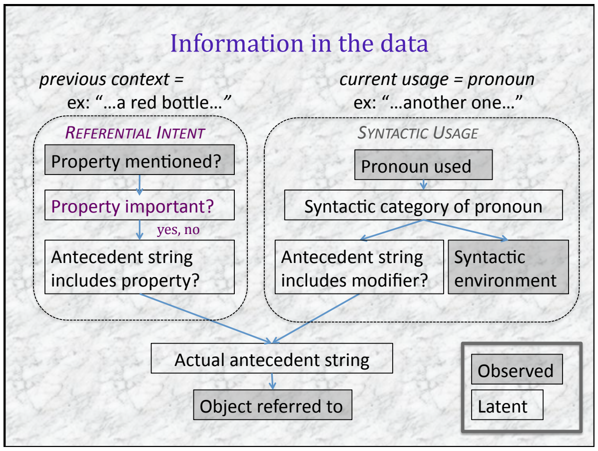


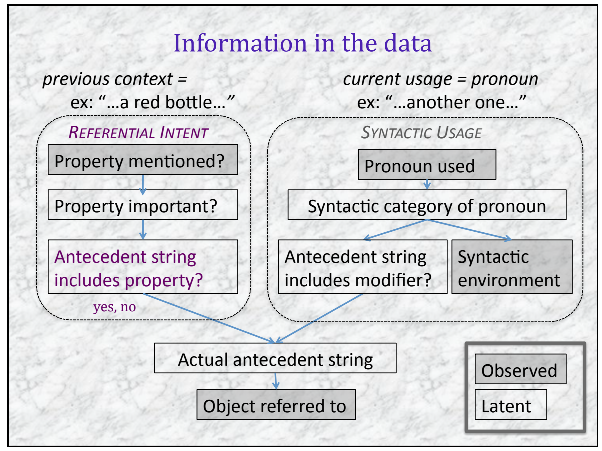

Information in the data
previous context =
ex: "…a red bottle…"
current usage = pronoun
ex: "…another one…"
Referential Intent
Property mentioned?
Property important?
Antecedent string includes property?
Syntactic Usage
Pronoun used
NP, N', or N^0?
Syntactic category of pronoun
Antecedent string includes modifier?
Syntactic environment
Actual antecedent string
Object referred to
Observed
Latent

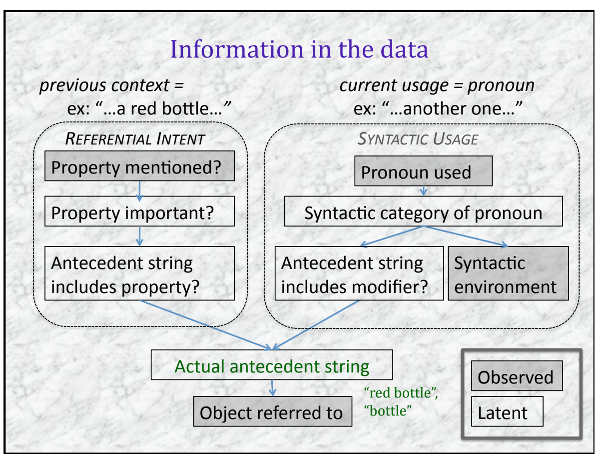

Information in the data
previous context =
ex: "…a red bottle…"
current usage = pronoun
ex: "…another one…"
Referential Intent
Property mentioned?
Property important?
Antecedent string includes property?
Syntactic Usage
Pronoun used
Syntactic category of pronoun
Antecedent string includes modifier?
Syntactic environment
<NP
Actual antecedent string
Object referred to
Observed
Latent

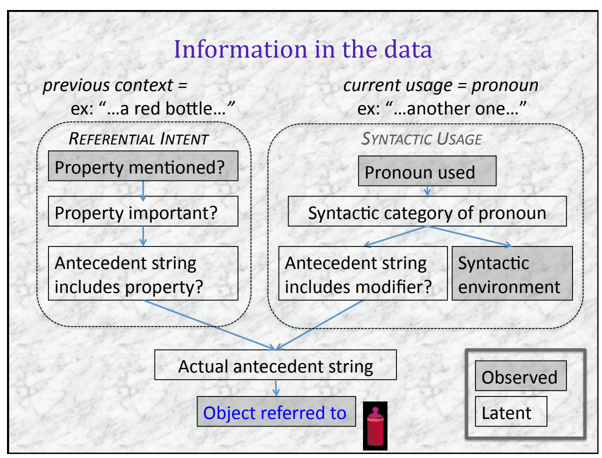

Information in the data
previous context =
ex: "…a red bottle…"
current usage = pronoun
ex: "…another one…"
Referential Intent
Property mentioned?
Property important?
Antecedent string includes property?
Syntactic Usage
Pronoun used
Syntactic category of pronoun
Antecedent string includes modifier?
Syntactic environment
N' = yes, no
N^0 = no
Actual antecedent string
Object referred to
Observed
Latent

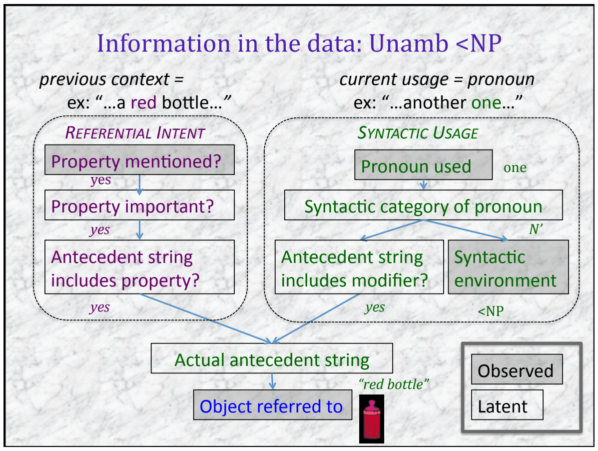

Information in the data
previous context =
ex: "…a red bottle…"
current usage = pronoun
ex: "…another one…"
Referential Intent
Property mentioned?
Property important?
Antecedent string includes property?
Syntactic Usage
Pronoun used
Syntactic category of pronoun
Antecedent string includes modifier?
Syntactic environment
Actual antecedent string
Object referred to
Observed
Latent

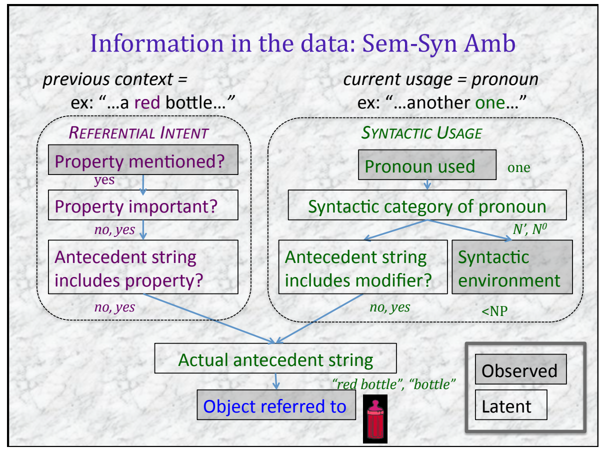

Information in the data
previous context =
ex: "…a red bottle…"
current usage = pronoun
ex: "…another one…"
Referential Intent
Property mentioned?
Property important?
yes, no
Antecedent string includes property?
Syntactic Usage
Pronoun used
Syntactic category of pronoun
Antecedent string includes modifier?
Syntactic environment
Actual antecedent string
Object referred to
Observed
Latent

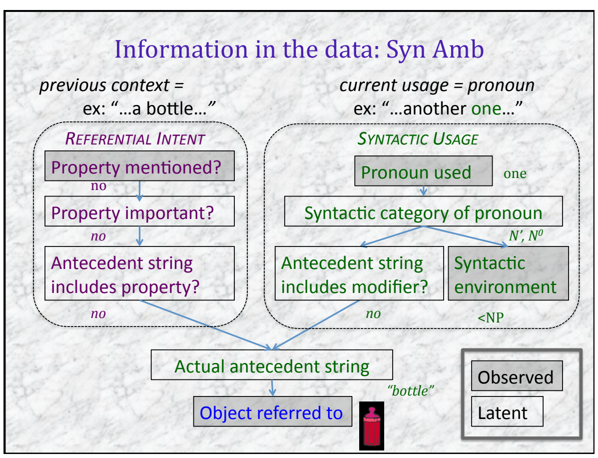

Information in the data
previous context =
ex: "…a red bottle…"
current usage = pronoun
ex: "…another one…"
Referential Intent
Property mentioned?
Property important?
Antecedent string includes property?
yes, no
Syntactic Usage
Pronoun used
Syntactic category of pronoun
Antecedent string includes modifier?
Syntactic environment
Actual antecedent string
Object referred to
Observed
Latent

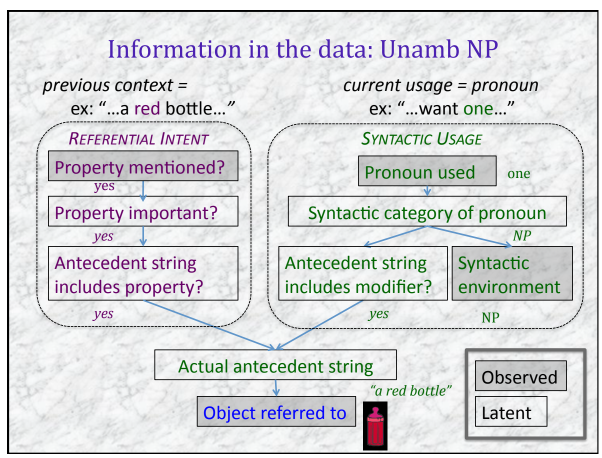

Information in the data
previous context =
ex: "…a red bottle…"
current usage = pronoun
ex: "…another one…"
Referential Intent
Property mentioned?
Property important?
Antecedent string includes property?
Syntactic Usage
Pronoun used
Syntactic category of pronoun
Antecedent string includes modifier?
Syntactic environment
Actual antecedent string
"red bottle",
"bottle"
Object referred to
Observed
Latent


Information in the data
previous context =
ex: "…a red bottle…"
current usage = pronoun
ex: "…another one…"
Referential Intent
Property mentioned?
Property important?
Antecedent string includes property?
Syntactic Usage
Pronoun used
Syntactic category of pronoun
Antecedent string includes modifier?
Syntactic environment
Actual antecedent string
Object referred to
Observed
Latent


Information in the data: Unamb <NP
previous context =
ex: "…a red bottle…"
current usage = pronoun
ex: "…another one…"
Referential Intent
Property mentioned?
yes
Property important?
yes
Antecedent string includes property?
yes
Syntactic Usage
Pronoun used
one
Syntactic category of pronoun
N'
Antecedent string includes modifier?
Syntactic environment
yes
<NP
Actual antecedent string
"red bottle"
Object referred to
Observed
Latent


Information in the data: Sem-Syn Amb
previous context =
ex: "…a red bottle…"
current usage = pronoun
ex: "…another one…"
Referential Intent
Property mentioned?
yes
Property important?
no, yes
Antecedent string includes property?
no, yes
Syntactic Usage
Pronoun used
one
Syntactic category of pronoun
N', N0
Antecedent string includes modifier?
Syntactic environment
no, yes
<NP
Actual antecedent string
"red bottle", "bottle"
Object referred to
Observed
Latent


Information in the data: Syn Amb
previous context =
ex: "…a bottle…"
current usage = pronoun
ex: "…another one…"
Referential Intent
Property mentioned?
no
Property important?
no
Antecedent string includes property?
no
Syntactic Usage
Pronoun used
one
Syntactic category of pronoun
N', N0
Antecedent string includes modifier?
Syntactic environment
no
<NP
Actual antecedent string
"bottle"
Object referred to
Observed
Latent


Information in the data: Unamb NP
previous context =
ex: "…a red bottle…"
current usage = pronoun
ex: "…want one…"
Referential Intent
Property mentioned?
yes
Property important?
yes
Antecedent string includes property?
yes
Syntactic Usage
Pronoun used
one
Syntactic category of pronoun
NP
Antecedent string includes modifier?
Syntactic environment
yes
NP
Actual antecedent string
"a red bottle"
Object referred to
Observed
Latent

The online probabilistic framework

Tracking the probability that a property mentioned in the potential antecedent is important: p_I

Property mentioned = yes → Property important?

Tracking the probability that the syntactic category is N' when it is smaller than NP: $p_{N'}$

Syntactic category of pronoun → Syntactic environment = <NP

The online probabilistic framework

General form of update equations for p_x (adapted from Chew 1971):

$$p_x = \frac{\overbrace{\alpha + data_x}^{\text{data seen suggesting x is true}}}{\underbrace{\alpha + \beta + totaldata_x}_{\text{total informative data seen w.r.t x}}} \qquad \alpha = \beta = 1 \text{ A very weak prior}$$

After every informative data point encountered:

$data_x = data_x + \phi_x$ Incremented by probability that data point suggests x is true

$totaldata_x = totaldata_x + 1$ One informative data point seen

The online probabilistic framework: Updating p_I

	ϕ_I	Explanation
Unamb <NP	1	Property definitely important
Unamb NP	1	Property definitely important
Syn Amb	N/A	Not informative for p_I
Sem-Syn Amb	$\frac{\rho_1}{\rho_1 + \rho_2 + \rho_3}$	Probability property is important

$\rho_1 = p_{N'} * \frac{m}{m+n} * p_I$ Category = N', choose N' with modifier, property is important

$\rho_2 = p_{N'} * \frac{n}{m+n} * (1 - p_I) * \frac{1}{t}$ Category = N', choose N' without modifier, property is not important, choose object with property by chance

$\rho_3 = (1 - p_{N'}) * (1 - p_I) * \frac{1}{t}$ Category = N^0, property is not important, choose object with property by chance

The online probabilistic framework: Updating $p_{N'}$

	$\phi_{N'}$	Explanation
Unamb <NP	1	Category definitely N'
Unamb NP	N/A	Not informative for $p_{N'}$
Syn Amb	$\frac{\rho_4}{\rho_4 + \rho_5}$	Probability category is N'
Sem-Syn Amb	$\frac{\rho_1 + \rho_2}{\rho_1 + \rho_2 + \rho_3}$	Probability category is N'

$\rho_1 = p_{N'} * \frac{m}{m+n} * p_I$ Category = N', choose N' with modifier, property is important

$\rho_2 = p_{N'} * \frac{n}{m+n} * (1 - p_I) * \frac{1}{t}$ Category = N', choose N' without modifier, property is not important, choose object with property by chance

$\rho_3 = (1 - p_{N'}) * (1 - p_I) * \frac{1}{t}$ Category = N^0, property is not important, choose object with property by chance

The online probabilistic framework: Updating $p_{N'}$

	$\phi_{N'}$	Explanation
Unamb <NP	1	Category definitely N’
Unamb NP	N/A	Not informative for $p_{N'}$
Syn Amb	$\frac{\rho_4}{\rho_4 + \rho_5}$	Probability category is N’
Sem-Syn Amb	$\frac{\rho_1 + \rho_2}{\rho_1 + \rho_2 + \rho_3}$	Probability category is N’

$\rho_4 = p_{N'} * \frac{n}{m+n}$ Category = N’, choose N’ without modifier

$\rho_5 = 1 - p_{N'}$ Category = N^0

Example updates

Start with $p_{N'} = p_I = 0.50$

One Unamb <NP data point: $p_{N'} = 0.67$, $p_I = 0.67$

One Unamb NP data point: $p_{N'} = 0.50$, $p_I = 0.67$

One Sem-Syn Amb data point: $p_{N'} = 0.56$, $p_I = 0.47$
m=1, n=3, t=5 [from P&L]

One Syn Amb data point: $p_{N'} = 0.48$, $p_I = 0.50$
m=1, n=3, t=5 [from P&L]

Corpus Analysis & Learner Input

Brown/Eve corpus (CHILDES: MacWhinney 2000): starting at 18 months

17,521 utterances of child-directed speech, 2874 referential pronoun utterances

Unamb <NP	0.00%
Sem-Syn Amb	0.66%
Syn Amb	7.52%
Unamb NP	8.42%
Uninformative	83.4%

Pearl & Lidz (2009): Children learn *one*’s representation between 14 and 18 months.

Based on estimates of the number of utterances children hear from birth until 18 months (Akhtar et al., 2004), we can calculate the data distribution in their input (36,500 referential pronoun utterances total).

Corpus Analysis & Learner Input

Learner Input based on Brown/Eve corpus distributions

		Baker	**R&G**	**P&L’s EO**	**P&M**
Unamb <NP	0.00%	0	0	0	0
Sem-Syn Amb	0.66%	0	242	242	242
Syn Amb	7.52%	0	0	2743	2743
Unamb NP	8.42%	0	0	0	3073
Uninformative	83.4%	36500	36258	33515	30442

Pearl & Lidz (2009): Children learn *one*’s representation between 14 and 18 months.

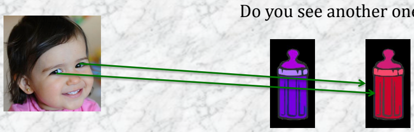
Based on estimates of the number of utterances children hear from birth until 18 months (Akhtar et al., 2004), we can calculate the data distribution in their input (36,500 referential pronoun utterances total).

Measures of Success: LWF children's behavior

In addition to directly assessing p_I and $p_{N'}$, we can measure how often a learner would reproduce the behavior in the LWF experiment.

Look – a red bottle!

Do you see another one?



Measures of Success: LWF children's behavior

In addition to directly assessing p_I and $p_{N'}$, we can measure how often a learner would reproduce the behavior in the LWF experiment.

2 choices
t = 2

$$\frac{\rho_1 + \rho_2 + \rho_3}{\rho_1 + 2 * \rho_2 + 2 * \rho_3}$$

Any outcome where learner looks at red bottle

Additional two outcomes where learner looks at other bottle

$\rho_1 = p_{N'} * \frac{m}{m+n} * p_I$ Category = N', antecedent = "red bottle"

$\rho_2 = p_{N'} * \frac{n}{m+n} * (1 - p_I) * \frac{1}{t}$ Category = N', antecedent = "bottle"

$\rho_3 = (1 - p_{N'}) * (1 - p_I) * \frac{1}{t}$ Category = N^0, antecedent = "bottle"

Testing LWF's assumption about what behavior means

In addition to directly assessing the learner's behavior, we can assess LWF's assumption that correct behavior indicates the children have the correct representation for *one*.

Is it possible to get correct behavior in the LWF experiment without having the correct representation for *one* in general (as measured by p_I and $p_{N'}$)?

Is it possible to get correct behavior in the LWF experiment without having the correct representation for *one* at the time the behavior is being produced?

$$\frac{\rho_1}{\rho_1 + \rho_2 + \rho_3}$$

the probability the look to the red bottle is because the learner has the correct representation (N', "red bottle")

given that the learner looks at the red bottle

Learner Results

Averages over 1000 simulations, standard deviations in parentheses.

	Baker	R&G	P&L's EO	P&M
p_I	0.50 (<.01)	0.95 (<.01)	0.02 (<.01)	>0.99 (<.01)
$p_{N'}$	0.50 (<.01)	0.97 (<.01)	0.17 (.02)	0.37 (.04)
p(LWF behavior)				
p(correct representation when producing LWF behavior)				

As previous studies found:
Traditional unambiguous data alone fails (Baker).
Leveraging Sem-Syn ambiguous data succeeds (R&G, P&L).
Leveraging Syn ambiguous data in addition fails (P&L's EO).

New result: Leveraging Unamb NP data (P&M) does not yield the correct representation in general ($p_{N'}$ is low), but…

Learner Results

Averages over 1000 simulations, standard deviations in parentheses.

	Baker	R&G	P&L's EO	P&M
p_I	0.50 (<.01)	0.95 (<.01)	0.02 (<.01)	>0.99 (<.01)
$p_{N'}$	0.50 (<.01)	0.97 (<.01)	0.17 (.02)	0.37 (.04)
p(LWF behavior)	0.53 (<.01)	0.93 (<.01)	0.50 (<.01)	>0.99 (<.01)
p(correct representation when producing LWF behavior)				

New result:
The probability of producing the LWF behavior with this incorrect representation is high.

How does this work?
If p_I is high, then when a property is mentioned (like "red"), the learner believes that property is relevant – which means the referent must include that property (RED BOTTLE).

Learner Results

Averages over 1000 simulations, standard deviations in parentheses.

	Baker	R&G	P&L's EO	P&M
p_I	0.50 (<.01)	0.95 (<.01)	0.02 (<.01)	>0.99 (<.01)
$p_{N'}$	0.50 (<.01)	0.97 (<.01)	0.17 (.02)	0.37 (.04)
p(LWF behavior)	0.53 (<.01)	0.93 (<.01)	0.50 (<.01)	>0.99 (<.01)
p(correct representation when producing LWF behavior)				

What this means:
LWF's assumption that correct behavior indicates the child has the correct representation does not seem to hold.

Learner Results

Averages over 1000 simulations, standard deviations in parentheses.

	Baker	R&G	P&L's EO	P&M
p_I	0.50 (<.01)	0.95 (<.01)	0.02 (<.01)	>0.99 (<.01)
$p_{N'}$	0.50 (<.01)	0.97 (<.01)	0.17 (.02)	0.37 (.04)
p(LWF behavior)	0.53 (<.01)	0.93 (<.01)	0.50 (<.01)	>0.99 (<.01)
p(correct representation when producing LWF behavior)	0.22 (<.01)	0.92 (<.01)	<0.01 (<.01)	>0.99 (<.01)

Or does it?
When the child produces the correct behavior in the LWF experiment, the probability that the child has the correct representation *when making that interpretation* is very high, even if the probability for the correct representation *in general* (e.g., when there is no modifier present) is very low.

Upshot: LWF were not wrong about children's representation when interpreting utterances like those in their experiment.

Learner Results

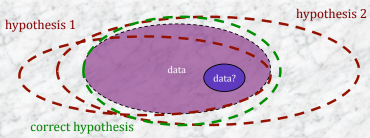
Averages over 1000 simulations, standard deviations in parentheses.

	Baker	R&G	P&L's EO	P&M
p_I	0.50 (<.01)	0.95 (<.01)	0.02 (<.01)	>0.99 (<.01)
$p_{N'}$	0.50 (<.01)	0.97 (<.01)	0.17 (.02)	0.37 (.04)
p(LWF behavior)	0.53 (<.01)	0.93 (<.01)	0.50 (<.01)	>0.99 (<.01)
p(correct representation when producing LWF behavior)	0.22 (<.01)	0.92 (<.01)	<0.01 (<.01)	>0.99 (<.01)

Also, the other learners behave as LWF expect:
When they show the correct behavior, they have the correct representation.
When they show incorrect behavior, they have the incorrect representation.

Recap & Implications

• Children may be able to learn the correct interpretation for *one* in certain situations (such as the LWF experiment) by broadening the set of data they consider relevant.

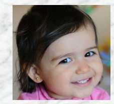


• Just because children demonstrate that they have the correct interpretation some of the time does not mean they have the correct representation all of the time.

Recap & Implications

• While children must eventually learn the correct representation of *one,* they do not necessarily need to do so by 18 months.

• Instead, they may realize that *one*'s category is N' (rather than N^0) at some later point.



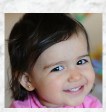
One possibility:
[F&al] Leverage the syntactic distribution of *one* with innate domain-general statistical learning, by using <u>subtle domain-specific semantic distinctions</u> that indicate syntactic category.

"ball with stripes"	"side of the road"
"one with dots"	*"one of the river"
[modifiers]	[complements = conceptually evoked by head noun]
[head noun = N']	[head noun = N^0]

The Acquisition Trajectory

I want *it.*
I want *one.* Another *one*!

Do you see *him*?
Do you see *one*?

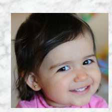


Before 18 months:
Need domain-specific knowledge
Recognize that *one* is similar to other anaphoric elements (*it, him,* etc.).

How to get it?
Derive it by using innate domain-general statistical learning abilities to observe the distribution of *one* compared to these other elements.

The Acquisition Trajectory

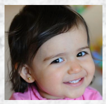

"Look – a red bottle! Oh, look – another one!"



Before 18 months:
Track how often a mentioned property is important for a referent to have.

How to get it?
Use innate domain-general statistical learning abilities to track this.

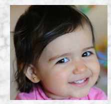
The Acquisition Trajectory



18 months:
Be able to assign the correct interpretation to utterances like those in the LWF experiment. (Know that *one* is N' in these cases.)

"Look – a red bottle! Do you see another one?"

The Acquisition Trajectory



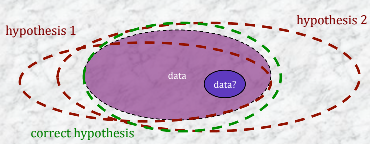
"ball with stripes"
"side of the road"
"one with dots"
*"one of the river"

After 18 months:
Need domain-specific knowledge about subtle semantic distinctions that indicate syntactic category in order to leverage the syntactic distribution of *one* with innate domain-general statistical learning.

How?
May come from innate domain-specific knowledge (UG) about language.

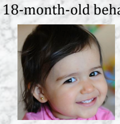
Back to the bigger questions

- When induction problems exist, what does it take to solve them?
  - What indirect evidence is available? How might a child leverage this evidence?
  - → Broader data sets that are identifiable via innate domain-general learning abilities may be additional sources of useful information.



Back to the bigger questions

- When induction problems exist, what does it take to solve them?
  - What learning biases can get the job done, given the available data? Are they necessarily innate and domain-specific (UG)?
  - → In this case study, the first step may not involve this kind of knowledge, although achieving the final adult knowledge state may.

Stage I	18-month-old behavior	Stage II
derived domain-specific knowledge		innate domain-specific knowledge?
innate domain-general statistical learning		innate domain-general statistical learning

Back to the bigger questions

- How can the necessary learning biases inform us about how the acquisition process works?
  - → identify learning biases needed to achieve 18-month-old behavior
  - → identify knowledge state those biases suggest
  - → suggest a two stage acquisition process for learning anaphoric *one*

Stage I

derived domain-specific knowledge

innate domain-general statistical learning

18-month-old behavior

Stage II

innate domain-specific knowledge?

innate domain-general statistical learning

The big picture

- Indirect evidence does not necessarily mean indirect negative evidence – it can come from considering a broader pool of informative data
- Indirect evidence does not necessarily negate the need for learning biases (of whatever kind)
- Considering indirect evidence and its impact on acquisition can help define concrete proposals about what is necessarily innate and domain-specific, and thus what is in Universal Grammar
- Knowing the impact of the necessary learning biases on acquisition may also inform us about the acquisition trajectory

Thank you

Ben Mis

The members of CoLa Lab at UCI

Computation of Language Laboratory UC Irvine

Erika Webb

Vance Chung

UCI Computational Models of Language Learning Seminar 2010

The National Science Foundation, grant BCS-0843896

NSF