Psych 215L: Language Acquisition

Lecture 16
Poverty of the Stimulus V:
Anaphoric *One*

Domain-general & domain-specific

Language acquisition may not just one or the other.

Three components of a learning theory, any of which can be either domain-general or domain-specific:

Representations of the data
- co-occurrence probabilities of acoustic signal vs.
  - phonemes, morphemes, syntactic trees

Data learned from
- filters that exclude data beyond the first 10 seconds vs.
  - filters that exclude data beyond the first clause

Updating process
- Bayesian updating vs.
  - language-specific updating process

Anaphoric *One*

Look - a red bottle!

Do you see another *one*?

Anaphoric *One*

Look - a red bottle!

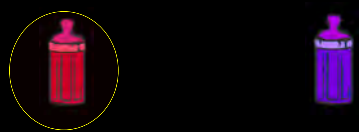
red bottle
Do you see another *one*?

Process: First determine the antecedent of *one* (what string *one* is replacing). Here, it seems to be replacing "red bottle".

Anaphoric *One*

Look - a red bottle!

Do you see another *one*? (red bottle)



Process: Because the antecedent ("red bottle") includes the modifier "red", the property RED is important for the referent of *one* to have. This is why we pick the red bottle as the intended referent of *one*.

Anaphoric *One*

Look - a red bottle!

Do you see another *one*?

Two steps:
(1)Identify syntactic antecedent
(2)Identify semantic referent

Anaphoric *One*

The second step is pretty straight-forward once you know the syntactic antecedent of *one*.

if antecedent = "red bottle", referent = RED BOTTLE 

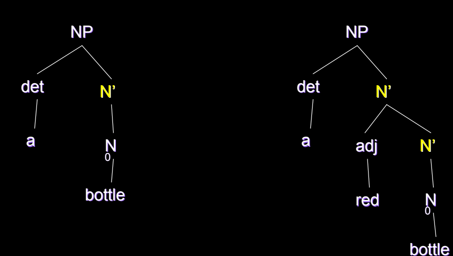
if antecedent = "bottle", referent = any BOTTLE 

As adults, we have knowledge about what the antecedent of *one* can be in various situations. In particular, we have knowledge about the syntactic category of *one*. The antecedent of *one* must be the same syntactic category of *one*, or else *one* couldn't replace it. So what syntactic category is *one*?

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

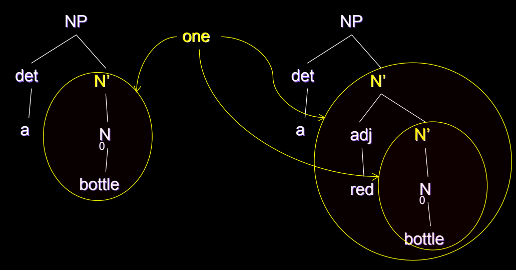
Anaphoric *One*: Syntactic Category

Many linguists believe 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".



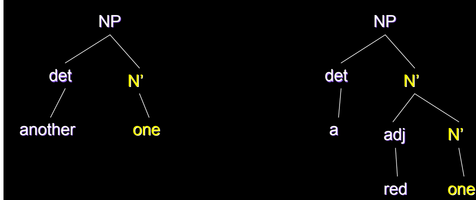
Anaphoric *One*: Syntactic Category

Many linguists believe 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".



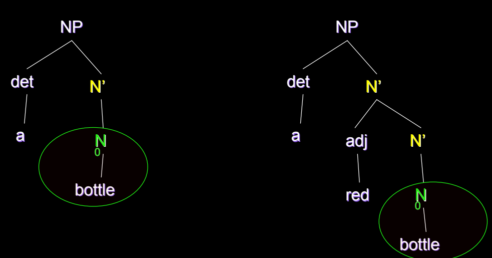
Anaphoric *One*: Syntactic Category

Many linguists believe 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".



Anaphoric *One*: Syntactic Category

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



Anaphoric *One*: Interpretations based on Syntactic Category

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

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

because *one* could only replace "bottle". We would interpret the second part as "Do you see another *bottle*?" Given this interpretation, we would consider any bottle a possible referent (like the purple bottle above), not just red bottles.

Anaphoric *One*: Interpretations based on Syntactic Category

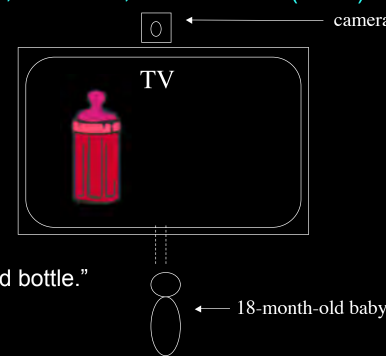
If *one* was syntactic category N^0, we would have a different interpretation of

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

because *one* could only replace "bottle". We would interpret the second part as "Do you see another *bottle*?" Given this interpretation, we would consider any bottle a possible referent (like the purple bottle above), not just red bottles.

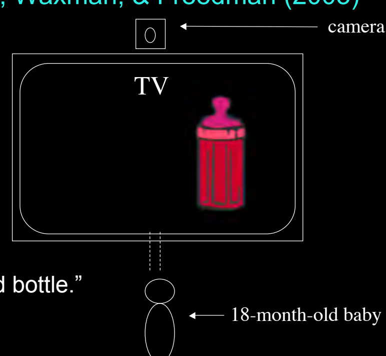
Since we allow (and in fact have a strong preference for) interpreting *one* as referring to the red bottle alone, we know that *one* cannot be syntactic category N^0. Instead, it is N' (and the antecedent in the above utterance is "red bottle").

18-month-old behavior: Lidz, Waxman, & Freedman (2003)



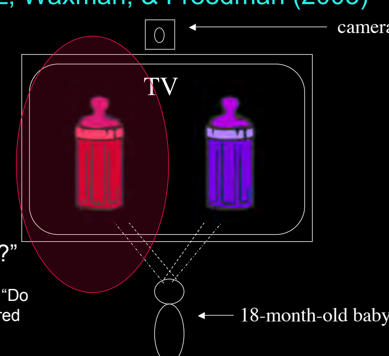
"Look! A red bottle."

18-month-old behavior: Lidz, Waxman, & Freedman (2003)



"Look! A red bottle."

18-month-old behavior: Lidz, Waxman, & Freedman (2003)



"Do you see another one?"

(Same results as "Do you see another red bottle?")

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 we saw.

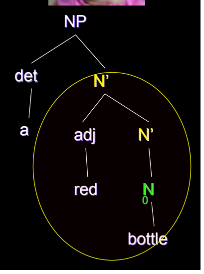
"Look – a red bottle! Do you see another one?"

LWF (2003) interpretation & conclusion:

Preference for red bottle means preferred syntactic antecedent is "red bottle".

"red bottle" can only be N' (not N^0).

Therefore, LWF concluded that 18-month-olds, like adults, believe *one* is category N' (and has antecedents that are category N').



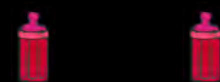
Anaphoric *One*: So what's the problem?

Acquisition: Children must learn the right syntactic category for *one*, so they end up with the right interpretation for *one*.

Problem: Most data children encounter are ambiguous for whether *one* is syntactic category N' or syntactic category N^0.

One type:

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



Incorrect hypothesis (*one* is N^0) is compatible:

If children have this incorrect hypothesis, they will interpret *one* as replacing "bottle", and look for any kind of bottle. The referent *is* a bottle, so this incorrect hypothesis is compatible with the observable data.

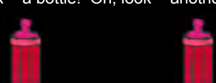
Anaphoric *One*: So what's the problem?

Acquisition: Children must learn the right syntactic category for *one*, so they end up with the right interpretation for *one*.

Problem: Most data children encounter are ambiguous for whether *one* is syntactic category N' or syntactic category N^0.

Another type:

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



Incorrect hypothesis (*one* is N^0) is compatible:

If children have this incorrect hypothesis, they will interpret *one* as replacing "bottle", and look for any kind of bottle. The referent *is* a bottle, so this incorrect hypothesis is compatible with the observable data.

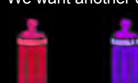
Anaphoric *One*: So what's the problem?

Acquisition: Children must learn the right syntactic category for *one*, so they end up with the right interpretation for *one*.

Problem: Lidz, Waxman, & Freedman (2003) estimate that less than 0.25% of data are unambiguous for the syntactic category of *one*.

Unambiguous data:

"Look – a red bottle! We want another *one*, but there doesn't seem to be *one* here."



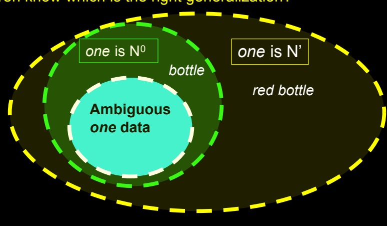
Incorrect hypothesis (*one* is N^0) is not compatible with this data point:

If children have this incorrect hypothesis, they will interpret *one* as replacing "bottle", and look for any kind of bottle. The other object present is a bottle, but the speaker claims another *one* isn't present – so *one* must be replacing "red bottle", not just "bottle" – which makes *one* an N'.

Anaphoric *One*: So what's the problem?

Acquisition: Children must learn the right syntactic category for *one*, so they end up with the right interpretation for *one*.

Problem: If children don't encounter unambiguous data often enough to notice them (Baker 1978, Hornstein & Lightfoot 1981, Crain 1991), they are left with data that are compatible with both hypotheses – that *one* is N^0 and that *one* is N'. How do children know which is the right generalization?



Recent Response from Regier & Gahl (2004)

Actually, children can learn this syntactic category information from the available data if the hypothesis space is simply *one* refers to N^0 vs. *one* refers to N'. The key is to cleverly use data that are ambiguous between the two hypotheses, instead of only using unambiguous data for what *one* refers to N'.

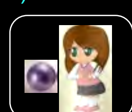
Estimates of children's data (what 18-month-olds heard)

Unambiguous data 10

"Jack wants a red ball. Lily doesn't have one."

Lily has a ball, but not a red ball.

one = *red ball*, and *one* refers to N'

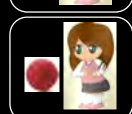


Type I Ambiguous data 183

"Jack wants a red ball. Lily has one."

Lily has a red ball.

one = *ball* OR *one* = *red ball*, *one* refers to N^0 OR N'

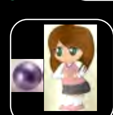


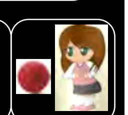
Type II Ambiguous Data 3805

"Jack wants a ball. Lily has one."

Lily has a ball.

one = *ball*, *one* refers to N^0 OR N'





Regier & Gahl (2004): A Model for how to learn the interpretation of *one*

Main idea: A Bayesian learner is a domain-general learning mechanism that would be able to use both Unambiguous and Type I Ambiguous data

Using Type I Ambiguous data:

"Jack wants a red ball, and Lily has *one* for him."

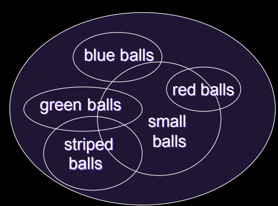
All the relevant knowledge for anaphoric *one* can be derived from knowing whether the property red is important for the referent (the ball, in this case) to have. (If the ball is always red, *red* is important and part of the string *one* refers to - and *red ball* is unequivocally N'.)

Basic strategy: Keep track of how often the referent that *one* refers to has the property mentioned in the potential antecedent (e.g. How often is the ball red?)

Bayesian expectations: The referents of *one*

If the property mentioned in the potential antecedent (e.g. *red*) is not important, the set of objects (e.g. balls) that *one* refers to should look something like this.

all balls

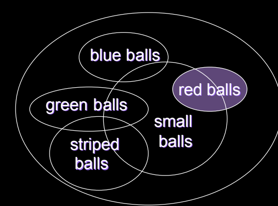


"…red ball…one…"

Bayesian expectations: The referents of *one*

If instead the property mentioned in the potential antecedent (e.g. *red*) is important, the set of objects (e.g. balls) that *one* refers to should look something like this.

all balls



"…red ball…one…"

Bayesian reasoning about referents

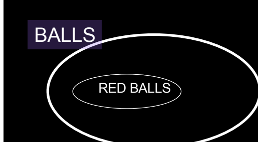
If the referents of *one* keep having the property mentioned in the potential antecedent (e.g. the balls keep being red when the phrase *red ball* is the potential antecedent), this is a conspicuous coincidence if the property isn't actually important. The Bayesian learner encodes this automatically and rewards the hypothesis that thinks the referent of *one* should be a red ball.

The reward is based on the relative size of the sets of potential referents (e.g., all balls vs. red balls).

Bayesian reasoning about referents

If red balls are a really small part of all the balls, it's really conspicuous that red balls keep being picked out. So, the Bayesian learner strongly rewards the hypothesis that the property red is actually important (i.e., that *red ball* is the antecedent).

"…red ball…one…"

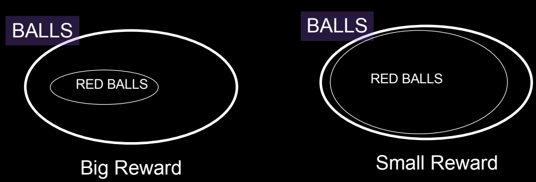


Big Reward

Bayesian reasoning about referents

If instead red balls are a really large part of all the balls, it's not really that conspicuous that red balls keep being picked out. So, the Bayesian learner weakly rewards the hypothesis that the property red is actually important (i.e., that *red ball* is the antecedent).

"…*red ball*…*one*…"



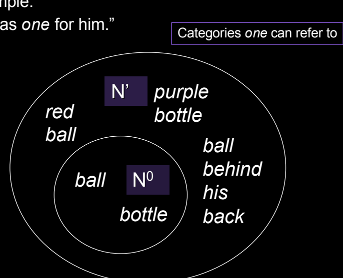
But what about the rest of the data?

One strength of Bayesian models are their ability to use all kinds of data, as long as the data are evenly mildly informative. So what about the Type II ambiguous data? Are these data informative? If so, it seems like a domain-general learner would use them as they make up the bulk of the data.

But what about the rest of the data?

Type II ambiguous data are informative if we think about the hypothesis space of potential antecedent strings for anaphoric *one*.

Type II Ambiguous data example:
"Jack wants a ball, and Lily has *one* for him."
one = *ball*, *one* = N^0 OR N'



Because of the layout of the hypothesis space (one hypothesis covers a subset of the strings the other covers), the Size Principle will favor the smaller hypothesis when the data are ambiguous.

Upshot: Type II Ambiguous data are informative about the syntactic category *one* refers to.

But maybe we wish they weren't…

Important Caveat: The smaller syntactic category hypothesis is that *one* refers to the category N^0. (Oops!) This means that the Type II Ambiguous data favor the incorrect syntactic hypothesis. Semantic consequence: any property that might be mentioned in the potential antecedent (e.g. *red*) won't matter because that property would be part of the larger N' category, not the N^0 category.

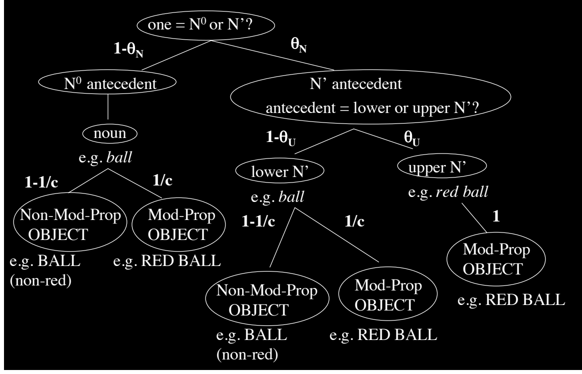
More pointedly, these data make up the bulk of the data to children - what would happen if a Bayesian learner used all the available informative data (Unambiguous, Ambiguous Type I, and Ambiguous Type II)?

An Equal-Opportunity Model

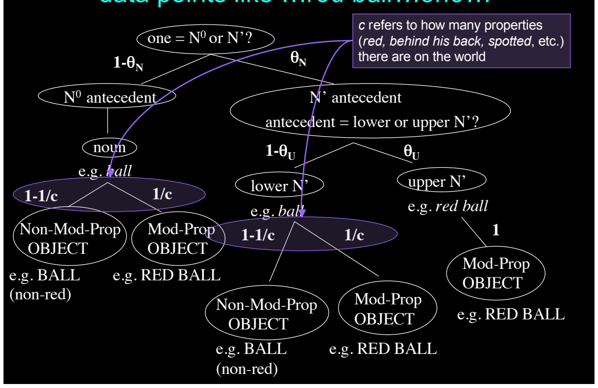
Generative model that learns by trying to construct the grammar that was used to generate the data ("analysis by synthesis").

Assumption: All data are generated by having *one* refer to an antecedent that is either an N^0 or N' string (θ_N). If an N' string is chosen and a property is mentioned in a potential antecedent, *one* can refer either to the smaller/lower N' (without the property, e.g. *ball*) or the larger/upper N' (with the property, e.g. *red ball*) (θ_U).

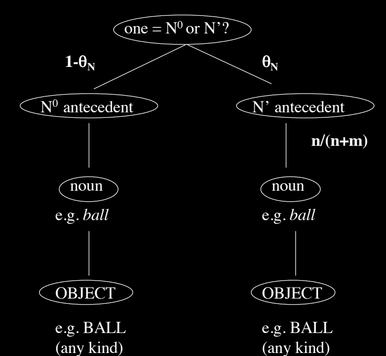
An Equal-Opportunity Model: Generating data points like …*red ball*…*one*…



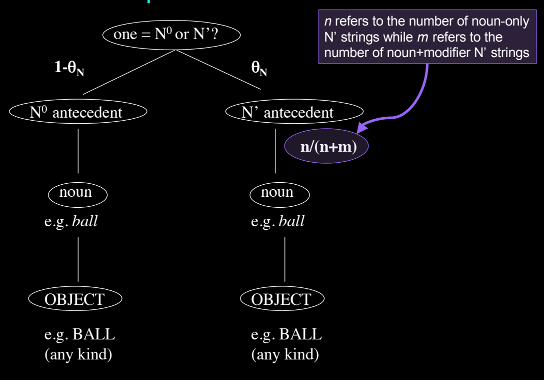
An Equal-Opportunity Model: Generating data points like …*red ball*…*one*…



An Equal-Opportunity Model: Generating data points like …*ball*…*one*…



An Equal-Opportunity Model: Generating data points like *…ball…one…*



Updating the Equal-Opportunity Learner

Unambiguous Data

$$\theta_N = \frac{\alpha + (data_N + 1)}{\alpha + \beta + (totaldata_N + 1)}, \alpha = \beta = 0.5 \qquad \theta_U = \frac{\alpha + (data_U + 1)}{\alpha + \beta + (totaldata_U + 1)}, \alpha = \beta = 0.5$$

Type I Ambiguous Data

$$\theta_N = \frac{\alpha + (data_N + I_{N-Type1})}{\alpha + \beta + (totaldata_N + 1)}, \alpha = \beta = 0.5 \qquad \theta_U = \frac{\alpha + (data_U + I_{U-Type1})}{\alpha + \beta + (totaldata_U + 1)}, \alpha = \beta = 0.5$$

Type II Ambiguous Data

$$\theta_N = \frac{\alpha + (data_N + I_{N-Type2})}{\alpha + \beta + (totaldata_N + 1)}, \alpha = \beta = 0.5$$

Updating the Equal-Opportunity Learner

Unambiguous Data, from $\theta_N = \theta_U = 0.5$

1 unambiguous data point, c = 5 (5 potential properties in the world)

$\theta_N = \theta_U = 0.75$

Type I Ambiguous Data, from $\theta_N = \theta_U = 0.5$

1 type I ambiguous data point, c = 5 (5 potential properties in the world)

$\theta_N = 0.625$, $\theta_U = 0.666$

Type II Ambiguous Data, from $\theta_N = \theta_U = 0.5$

1 type II ambiguous data point, n/m = 1/2 (two string types, of which a simple noun string (*ball*) is 1)

$\theta_N = 0.417$, $\theta_U = 0.5$

EO Model: Interpreting Anaphoric *One*

For a given utterance involving anaphoric *one* where there is more than one potential N' antecedent (e.g., …*red ball…one*…):

(1) Decide if the antecedent should be N^0 or N', using θ_N.
(2) If the antecedent is N^0, the referent is any object regardless of property (e.g., any ball)
(3) If the antecedent is N', decide if the antecedent is the smaller/lower or larger/upper N', using θ_U.
(4) Based on this decision, pick out the appropriate referent (e.g., lower = *ball*, so referent is any ball; upper = *red ball*, so referent is a red ball)

Initial probability of adult interpretation (choose N', choose upper N'):

$\theta_N * \theta_U = 0.5 * 0.5 = 0.25$.

Good learning means this probability increases over time.

EO Model: Results with generous parameter value estimates

Probability of choosing *one* anaphoric to N' is low. But if the learner happens to do that, probability of choosing the correct N' is high. Making the parameter values less generous only exacerbates the problem. Upshot: Equal-Opportunity Learner has a problem.

Back to models that don't use all the available informative data

Main point: Using some of the ambiguous data is better than ignoring it all (similar to what Regier & Gahl 2004 found). A data filter is useful for the learner…so how could a learner implement one sensibly?

About the data filter

Ignore some of the ambiguous data, but not all of it.

Domain-specific or domain-general?

Pearl & Lidz say: "Given that this filter requires the learner to single out a specific type of potentially informative data to ignore, and the property of this ignored data involves whether the potential linguistic antecedent has a modifier, we consider this filter to be specific to language learning. As such, it seems reasonable to consider it a domain-specific filter."

About a child implementing the data filter

Pearl & Lidz say: "It seems fairly obvious that the learner cannot (and probably should not) come equipped with a filter that says 'ignore type II ambiguous data' without some procedure for identifying this data. What we really want to know is whether there is a principled way to derive this filter. Specifically, we want the filter that ignores type II ambiguous data to be a consequence of some other principled learning strategy."

About a child implementing the data filter

A domain-general idea: Learn in cases of uncertainty.

Type II Ambiguous data (…*ball*…*one*…) doesn't count as uncertain because in the local context (that is, for that one data point), the referent of *one* isn't uncertain - the antecedent is the simple noun (*ball*) and the referent is the object corresponding to that noun (ball). (However, at the global level (for deciding the syntactic category *one* is anaphoric with), this data point is uncertain.)

Type I Ambiguous data (…*red ball*…*one*…), however, is uncertain in the local context because it is unclear which string *one* is anaphoric with (*red ball, ball*) and so unclear what the referent is.

Upshot: "Learn in cases of local uncertainty" would cause the child to use Type I Ambiguous data and ignore Type II Ambiguous data…which then makes it possible to learn anaphoric *one*.

Pearl & Lidz conclusions

"The case of anaphoric *one* demonstrates the interplay between domain-specificity and domain-generality in language learning. What we have seen is that a domain-general learning procedure can be successful in this case, but crucially only when paired with domain-specific filters on data intake. Moreover, we have suggested that the particular domain-specific filter that yields the best result can plausibly be derived from a domain-general learning strategy."

"…emphasized the efficacy of data intake filtering on learners. Filtering the data is, in some sense, a counterintuitive approach to learning because it discards potentially informative data. Moreover, eliminating data can lead to a data sparseness problem. However, in order to find the correct generalizations in the data in our case, we found that eliminating some data was more effective than using it all. The right generalizations are hiding in the data, but paying attention to all of the data will make them harder to find."

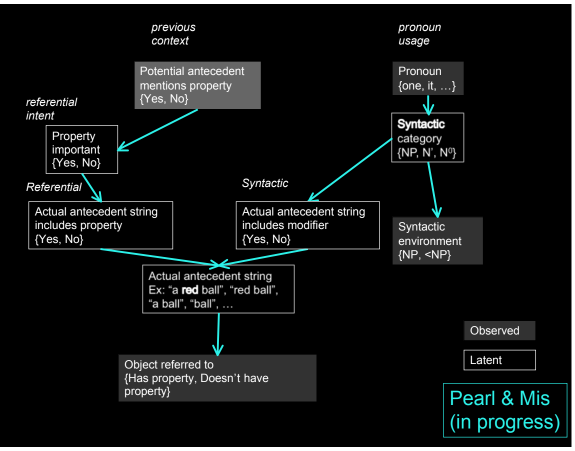
Pearl & Mis (in progress)

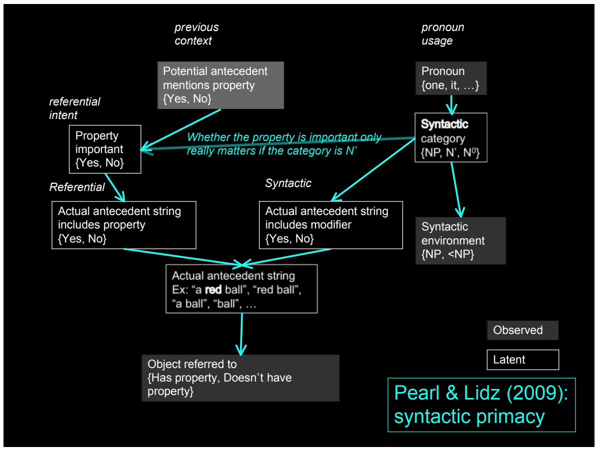
Taking "use all the informative data" one step further – worth considering that *one* is a referential element, and there are other referential elements in the language (like *it*) that might be informative for learning about *one*.

In particular, the decision about whether a mentioned property is important in the potential antecedent is applicable whether the referential element is *it* or *one*.

"….a red ball…want *it.*"

"…a red ball…want another *one*."




previous context
pronoun usage
Potential antecedent mentions property {Yes, No}
Pronoun {one, it, …}
referential intent
Property important {Yes, No}
Whether the property is important only really matters if the category is N'
Syntactic category {NP, N', N0}
Referential
Syntactic
Actual antecedent string includes property {Yes, No}
Actual antecedent string includes modifier {Yes, No}
Syntactic environment {NP, <NP}
Actual antecedent string Ex: "a red ball", "red ball", "a ball", "ball", …
Observed
Latent
Object referred to {Has property, Doesn't have property}
Pearl & Lidz (2009): syntactic primacy