Some Fundamental Issues in Language Acquisition Research & the Contribution of Computational Modeling

Lisa S. Pearl
Assistant Professor
Department of Cognitive Sciences
SBSG 2314
University of California
Irvine, CA 92697-5100

lpearl@uci.edu

Computation of Language Laboratory

UC Irvine

Nov 10, 2010

California State University, Fullerton

Language Acquisition as Information Extraction

Big Question:

How do children acquire knowledge about language from the language data they encounter?

Kids Do Amazing Things

Much of the linguistic system is already known by age 3.

…when most kids can't tie their own shoes or reliably recognize "4".

What kids are doing: extracting patterns and making generalizations from the surrounding data mostly without explicit instruction or correction. That is, they seem to learn mostly just by hearing examples of what is allowed in the language ("positive evidence"), rather than by being shown examples of what is not allowed in the language ("negative evidence").

How do we know they're not being taught?

Because they don't necessarily pay attention to correction.

(From Martin Braine)

Child: Want other one spoon, Daddy.
Father: You mean, you want the other spoon.
Child: Yes, I want other one spoon, please Daddy.
Father: Can you say "the other spoon"?
Child: Other…one…spoon.
Father: Say "other".
Child: Other.
Father: "Spoon."
Child: Spoon.
Father: "Other spoon."
Child: Other…spoon. Now give me other one spoon?

How do we know they're not being taught?

We mostly wouldn't know what to teach them anyway.

Once we go beyond the most superficial things (like "cat" is a furry, purring pet), most of our knowledge is subconscious. We know it – but we don't know *how* we know it or why it's so.

Knowledge of Language & Hidden Rules

Some examples from language:

You know that…

…*strop* is a possible word of English, while *stvop* isn't.

Knowledge of Language & Hidden Rules

Some examples from language:

You know that…

…"Who did you see who did that?" is not a grammatical question in English

(Instead: "Who did you see do that?")

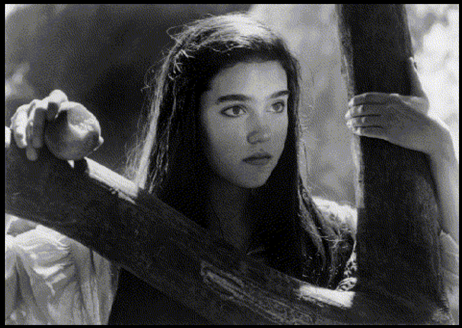
Knowledge of Language & Hidden Rules

Some examples from language:

You know that…

…In "She ate the peach while Sarah was reading", *she ≠ Sarah*

but *she* can be *Sarah* in all of these:



Sarah ate the peach while she was reading.
While she was reading, Sarah ate the peach.
While Sarah was reading, she ate the peach.

Knowledge of Language & Hidden Rules

Some examples from language:

You know that…

…the 's' in 'cats' sounds different from the 's' in goblins

Knowledge of Language & Hidden Rules

Some examples from language:

You know that…

…one structure doesn't necessarily have the same interpretation.

This is the rabbit I want to banish.
=~ I want (*me*) to banish the rabbit.
[NOT: I want the rabbit to banish (*something*).]

This is the rabbit I want to disappear.
=~ I want the rabbit to disappear.
[NOT: I want (*me*) to disappear the rabbit.]

Knowledge of Language & Hidden Rules

Some examples from language:

You know that…

… contracted forms like "wanna" and "gonna" can't always replace their respective full forms "want to" and "going to".

You get to choose who you will rescue.

"Who do you want to rescue?"

"Who do you wanna rescue?"

Knowledge of Language & Hidden Rules

Some examples from language:

You know that…

… contracted forms like "wanna" and "gonna" can't always replace their respective full forms "want to" and "going to".

You get to choose who you will rescue.

"Who do you want to rescue?"

"Who do you wanna rescue?"

You get to choose who will do the rescuing.

"Who do you want to do the rescuing?"

* "Who do you wanna do the rescuing?"

Knowledge of Language & Hidden Rules

Some examples from language:

You know that…

… contracted forms like "wanna" and "gonna" can't always replace their respective full forms "want to" and "going to".

You get to choose who you will rescue.

"Who are you going to rescue?"

"Who are you gonna   rescue?"

Knowledge of Language & Hidden Rules

Some examples from language:

You know that…

… contracted forms like "wanna" and "gonna" can't always replace their respective full forms "want to" and "going to".

You get to choose who you will rescue.

"Who are you going to rescue?"

"Who are you gonna rescue?"

"I'm going to the witch's lair to rescue her."

* "I'm gonna the witch's lair to rescue her."

About Language

One way to think about how to classify the knowledge that you have when you know a language:

You know what items (sounds, words, sentences, questions, etc.) are part of the language. You can tell whether or not a given item is grammatical in the language.

Hoggle is definitely an ornery dwarf. [grammatical]

* Hoggle an dwarf definitely ornery is. [ungrammatical]

About Language

One way to think about how to classify the knowledge that you have when you know a language:

You know what items (sounds, words, sentences, questions, etc.) are part of the language. You can tell whether or not a given item is grammatical in the language.

Hoggle is definitely an ornery dwarf. [part of English]

* Hoggle an dwarf definitely ornery is. [not part of English]

About Language

One way to think about how to classify the knowledge that you have when you know a language:

You know what items (sounds, words, sentences, questions, etc.) are part of the language. You can tell whether or not a given item is grammatical in the language.

The reason you can do this is because you know the rules & patterns that generate the items that are part of the language. (mental grammar)

About Children Learning Language

Adult knowledge: grammar that generates the items that are part of the language.

The child's job: figure out the rules that generate the items that belong in the language and that don't generate items that don't.

For example, the child wants rules to generate

"Hoggle is definitely an ornery dwarf."

but not

"Hoggle an dwarf definitely ornery is."

Not in English

Hoggle an
dwarf ornery is

Bite
adventurers
fairies

Can the girl who summon the Goblin King
can solve the Labyrinth?

In English

Hoggle is an ornery
dwarf

Fairies bite
adventurers

Can the girl who
can summon the
Goblin King solve
the Labyrinth?

Want to learn rules that generate this set of items…

Not in English

Hoggle an dwarf ornery is

Bite adventurers fairies

In English

Hoggle is an ornery dwarf

Fairies bite adventurers

Can the girl who can summon the Goblin King solve the Labyrinth?

Can the girl who summon the Goblin King can solve the Labyrinth?

…and exclude this set of items

Not in English

Hoggle an dwarf ornery is

Bite adventurers fairies

In English

Hoggle is an ornery dwarf

Fairies bite adventurers

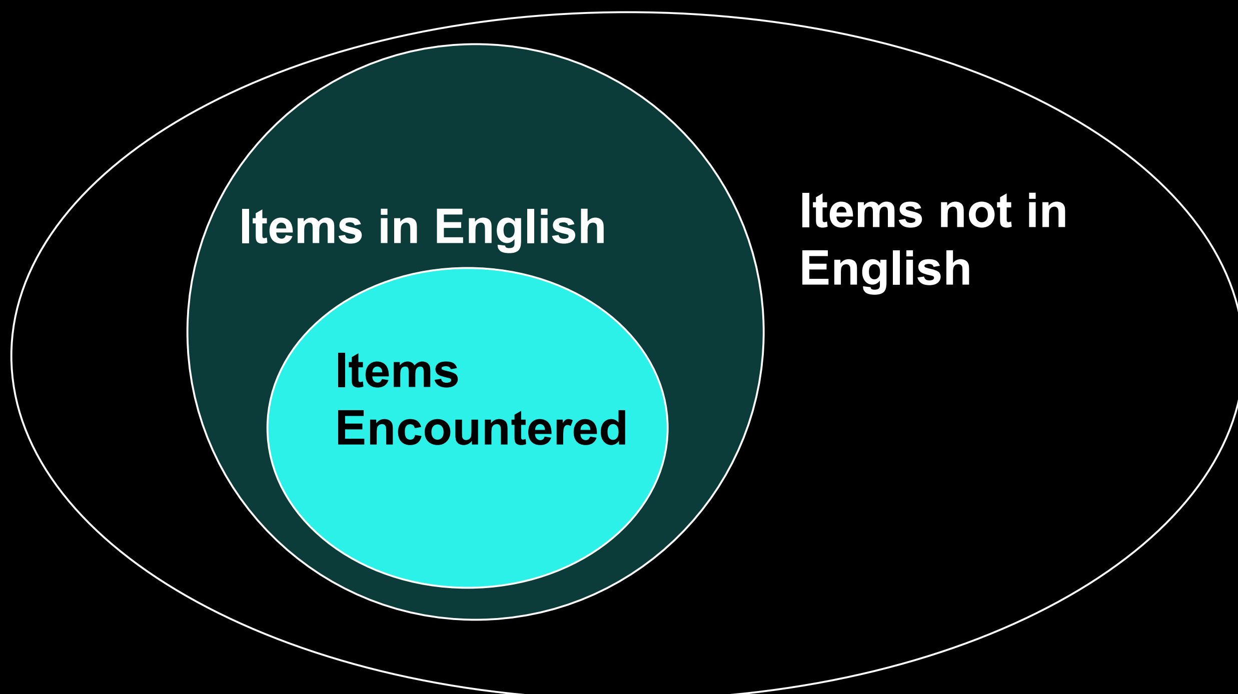
Can the girl who can summon the Goblin King solve the Labyrinth?

Can the girl who summon the Goblin King can solve the Labyrinth?

So what's the problem?

It's not clear that children encounter all the items that are part of the language.

If they only encounter a subset of the language's items, how do they know everything that belongs in the language?

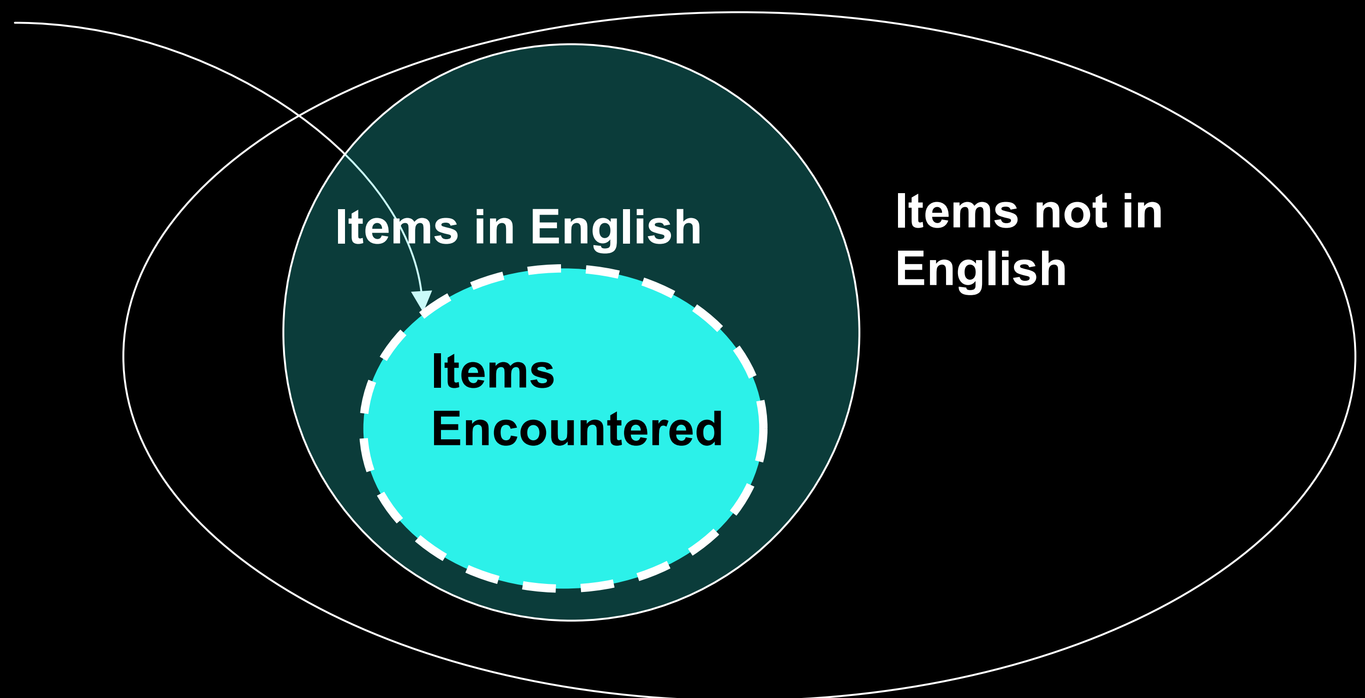


So what's the problem?

One solution: children generalize

But how do they generalize?

To here?

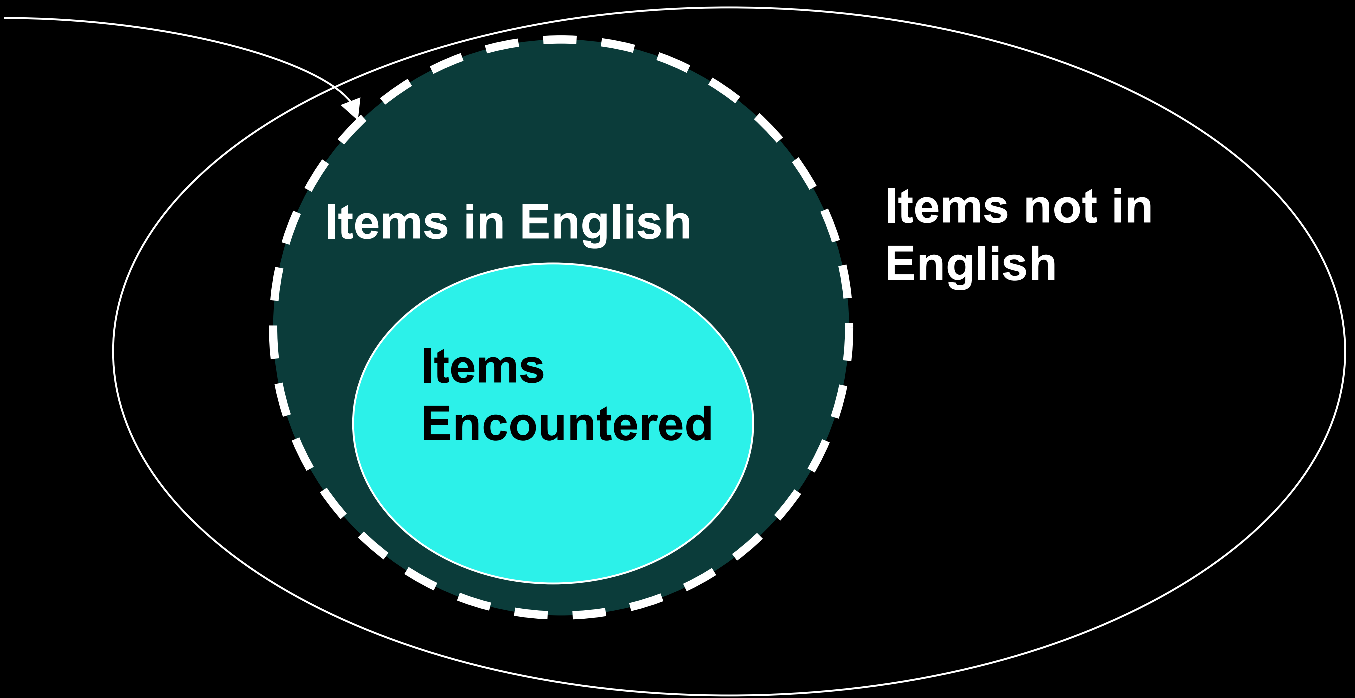


So what's the problem?

One solution: children generalize

But how do they generalize?

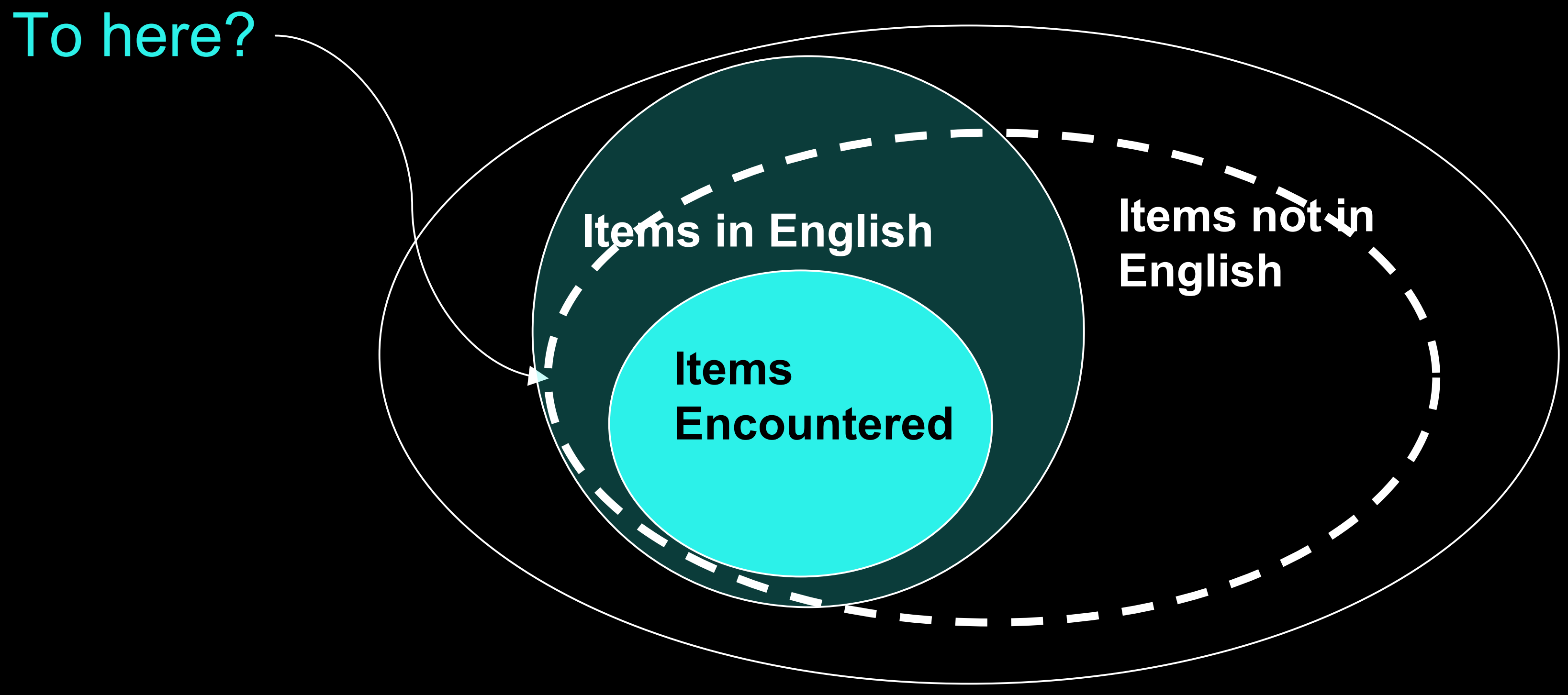
To here?



So what's the problem?

One solution: children generalize

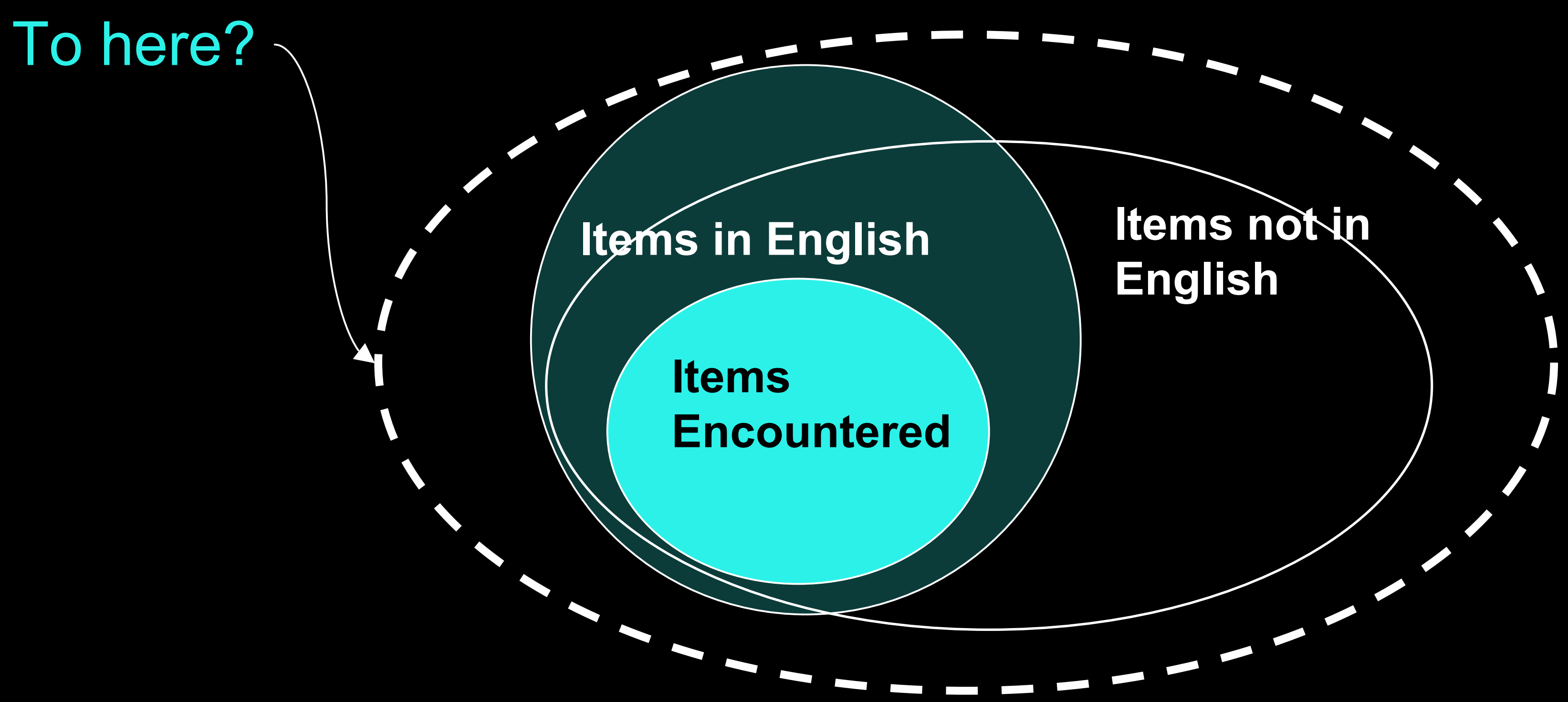
But how do they generalize?



So what's the problem?

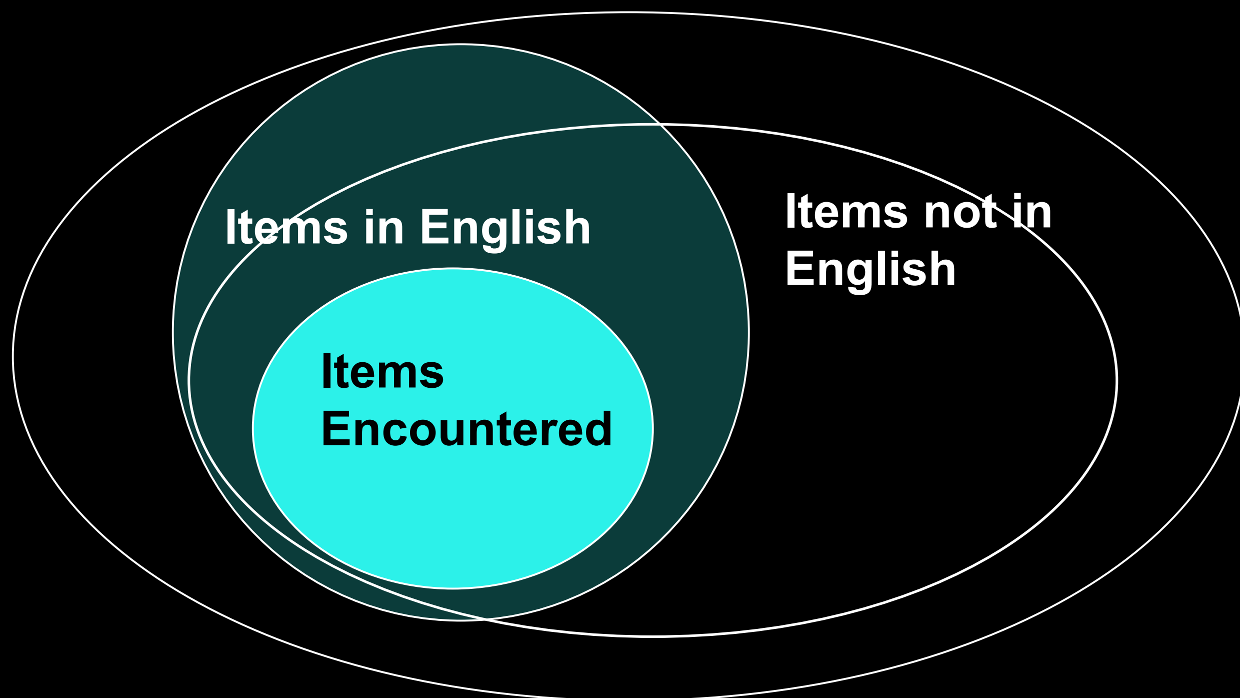
One solution: children generalize

But how do they generalize?



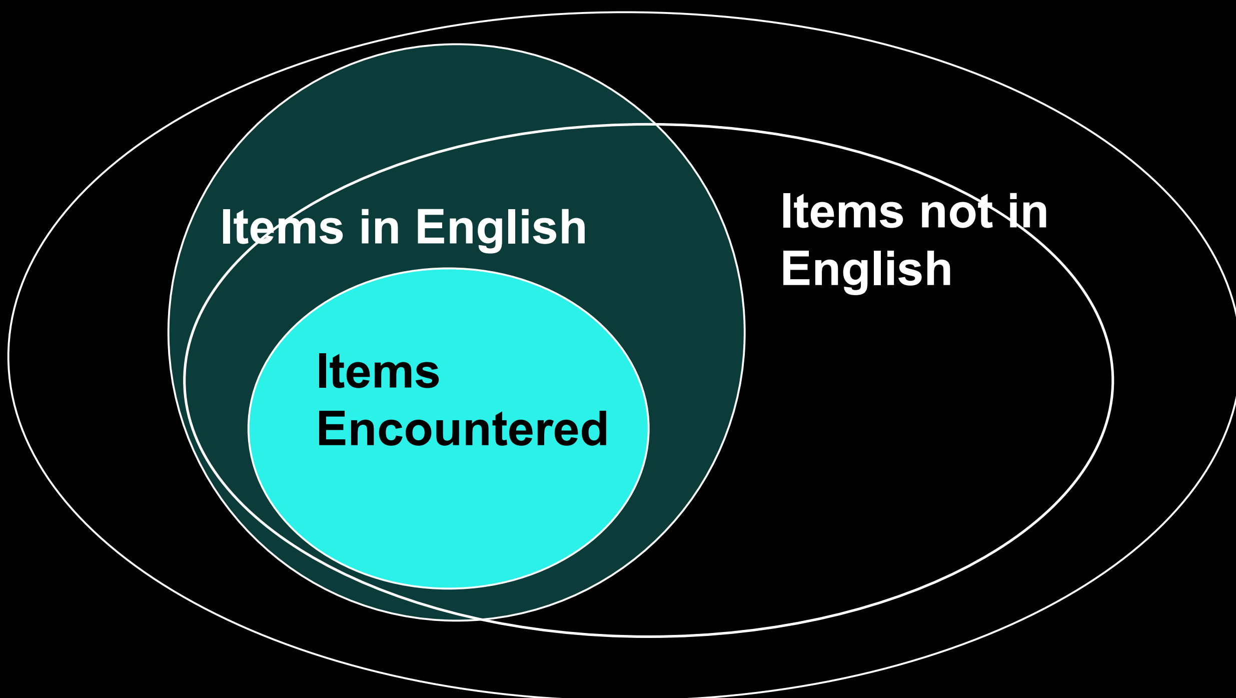
So what's the problem?

The problem is that children must make the right generalization from data that are compatible with multiple generalizations. In this sense, the data encountered are impoverished. They do not single out the correct generalization by themselves.



So what's the problem?

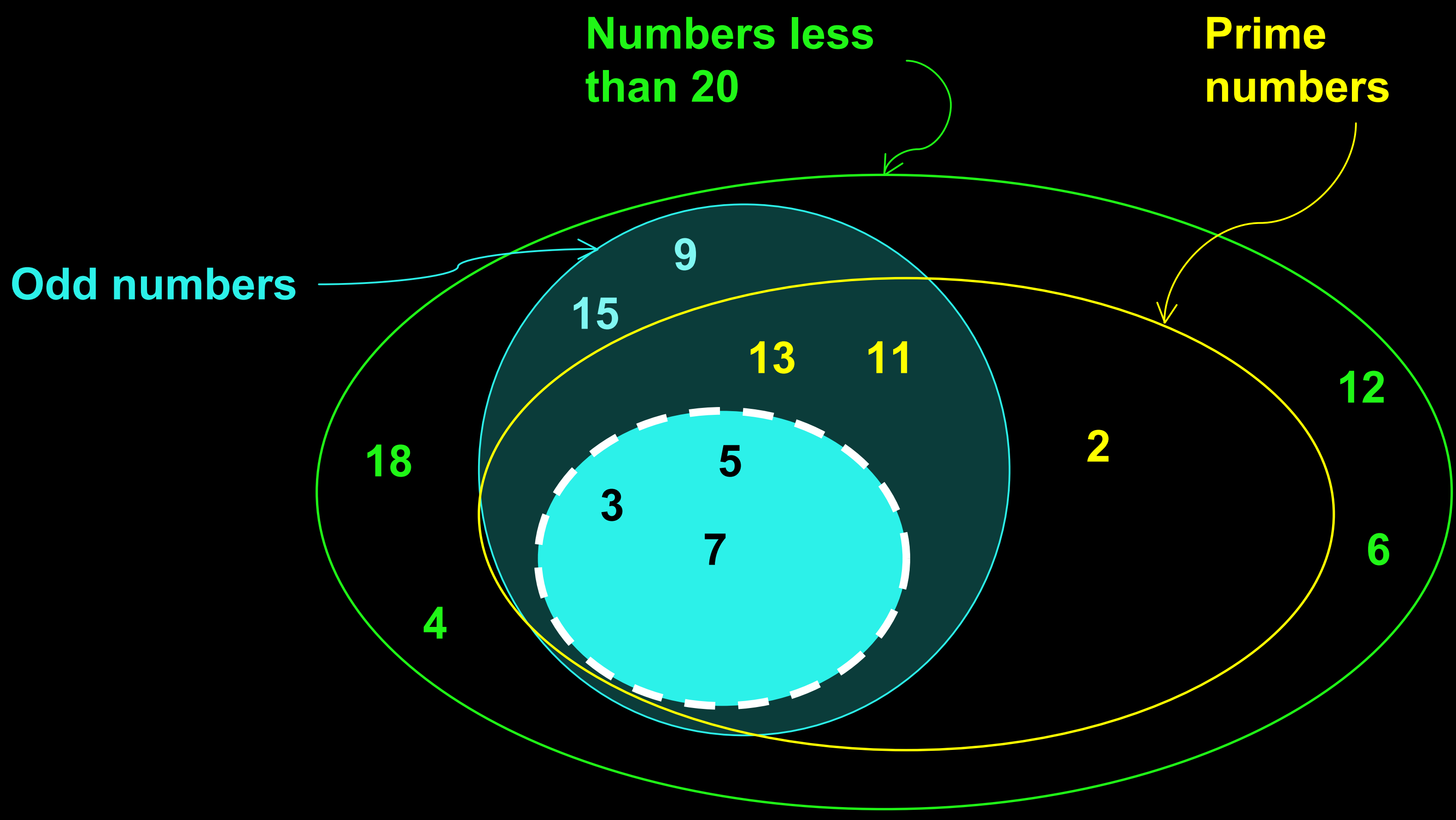
This is sometimes called the "poverty of the stimulus", the "induction problem", the "no negative evidence problem", or the "logical problem of language acquisition".



A numerical analogy

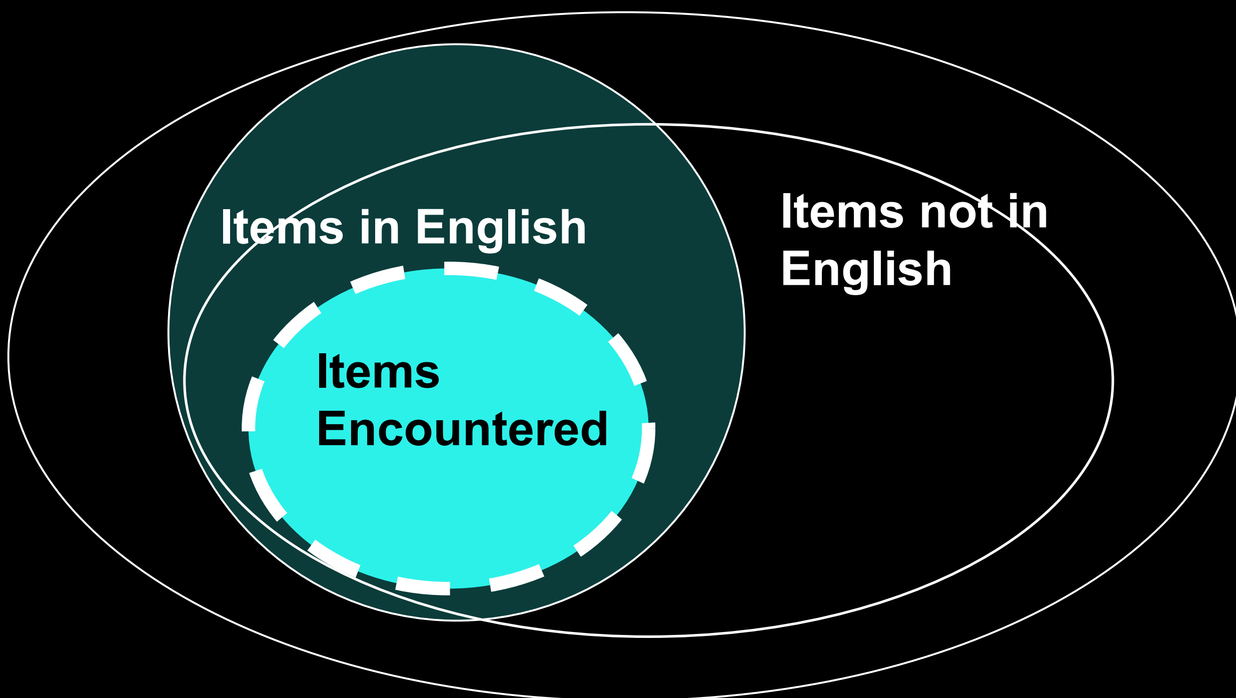
Items encountered: 3, 5, 7

What set are these numbers drawn from? That is, what is the right "number rule" for this language that will allow you to predict what numbers will appear in the future?



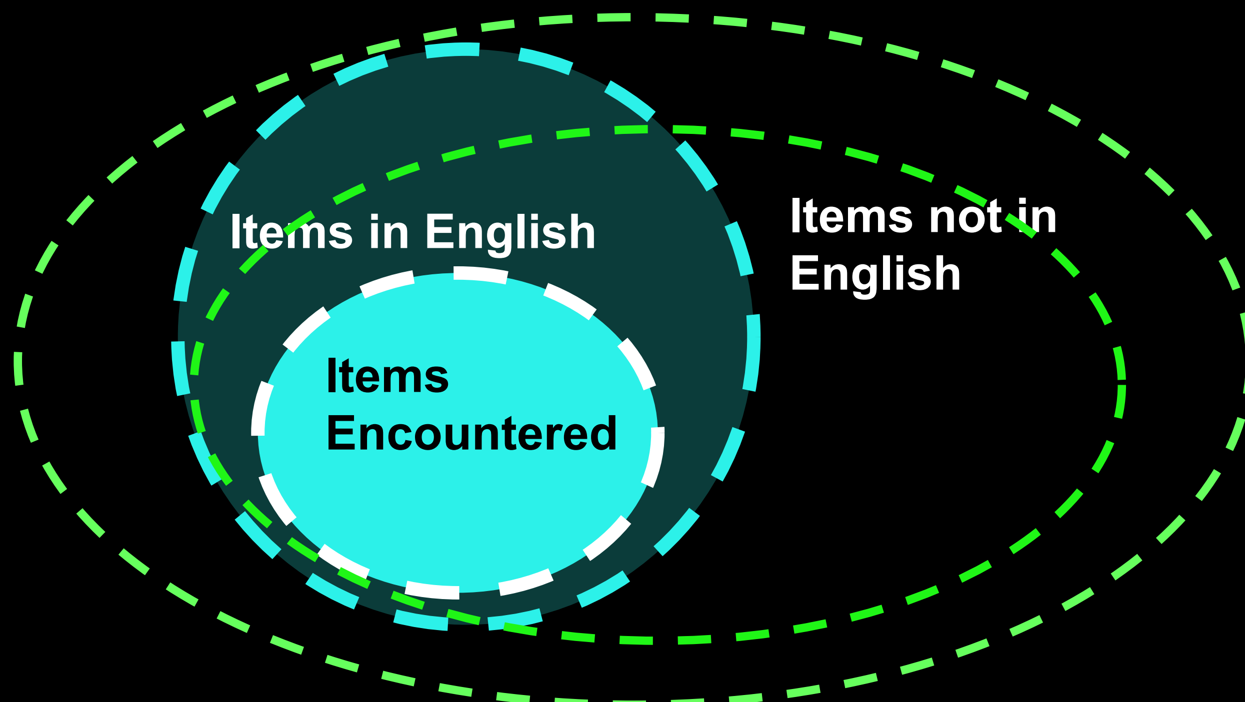
Poverty of the Stimulus: Logic

Children encounter data that are compatible with many hypotheses about the correct rules and patterns of the language.



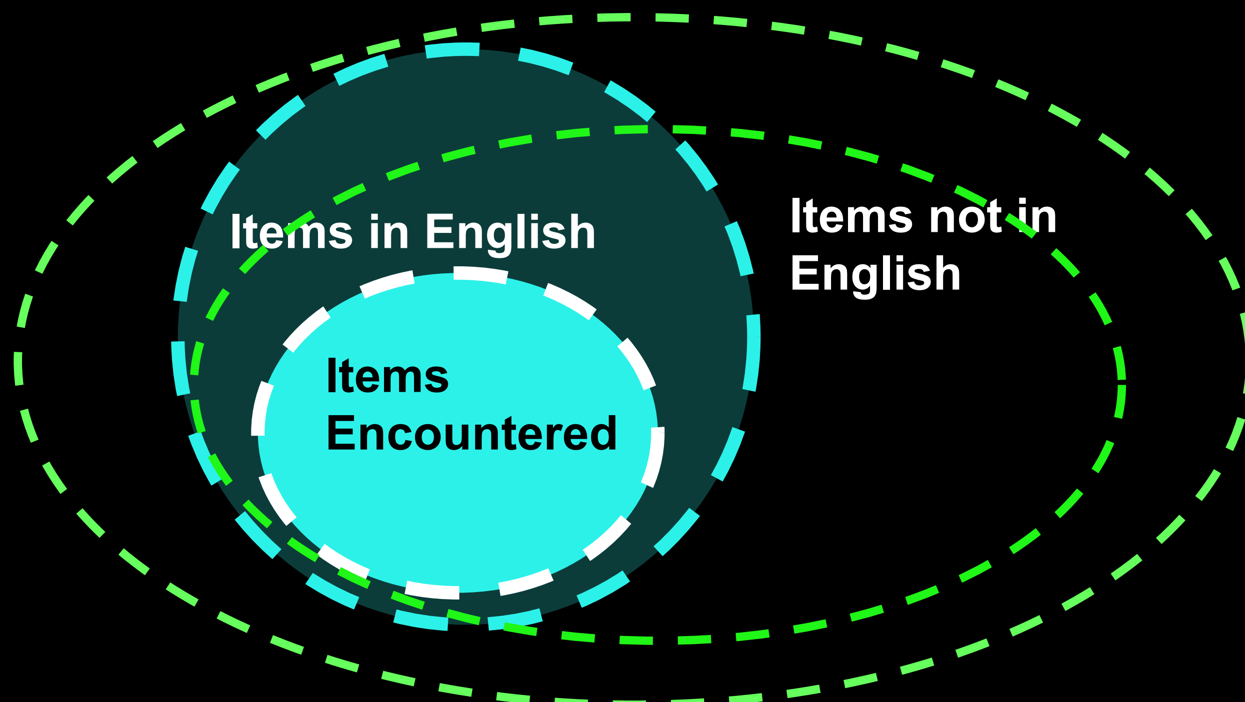
Poverty of the Stimulus: Logic

Specifically, the data encountered are compatible with both the correct hypothesis and other, incorrect hypotheses about the rules and patterns of the language.



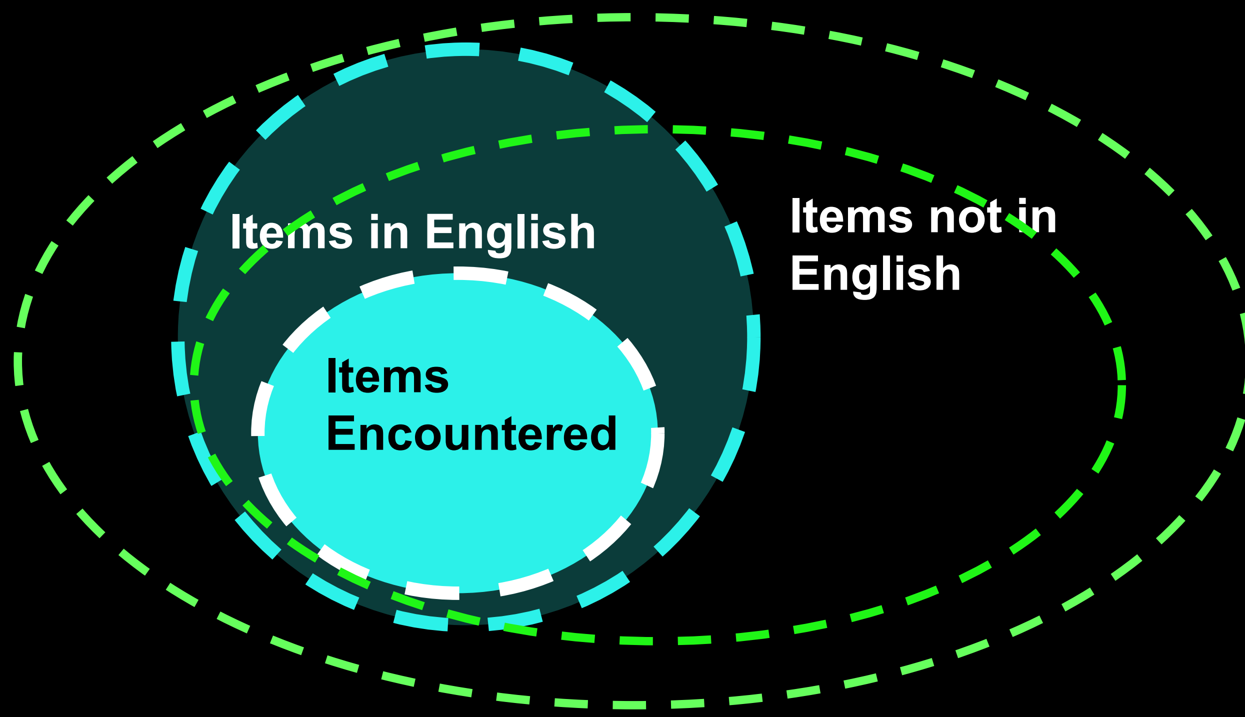
Poverty of the Stimulus: Logic

A rational learner would consider all compatible hypotheses, and perhaps make errors before choosing the correct hypothesis. Maybe some rational learners would choose an incorrect hypothesis in the end.



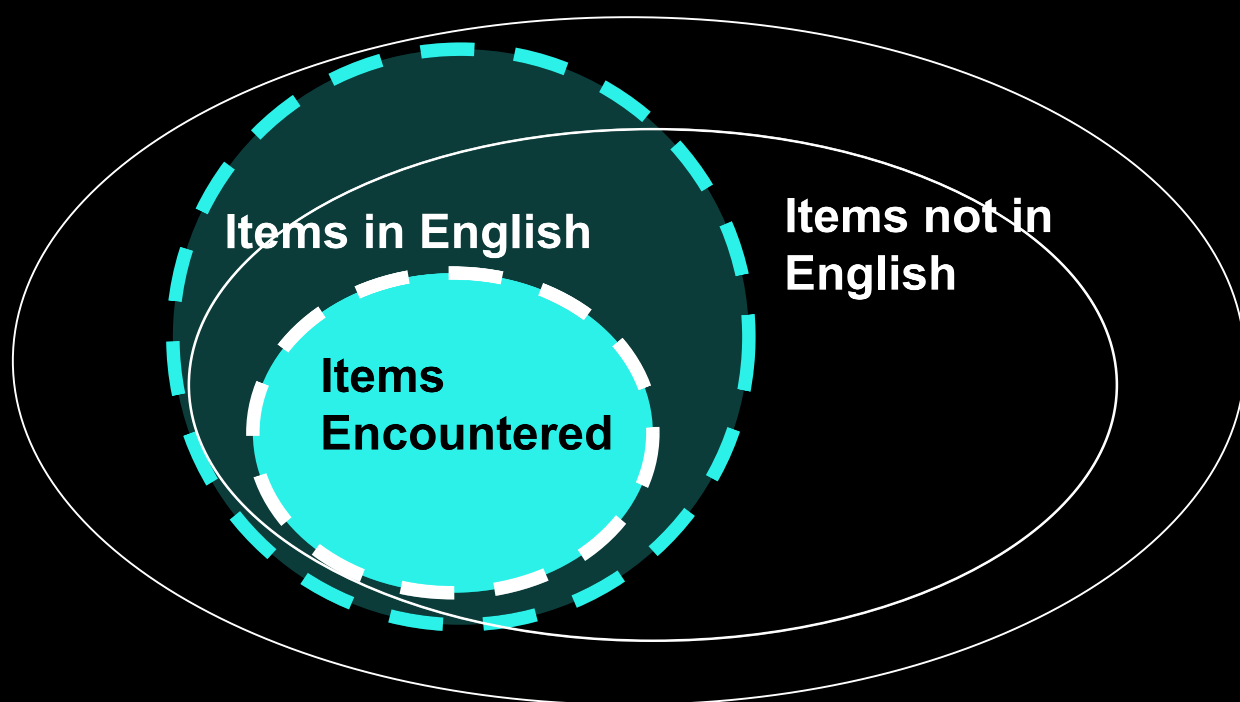
Poverty of the Stimulus: Logic

Expectation for rational learners: errors in performance. Children will behave as if they think ungrammatical items are part of the language at some point.



Argument for Prior Knowledge

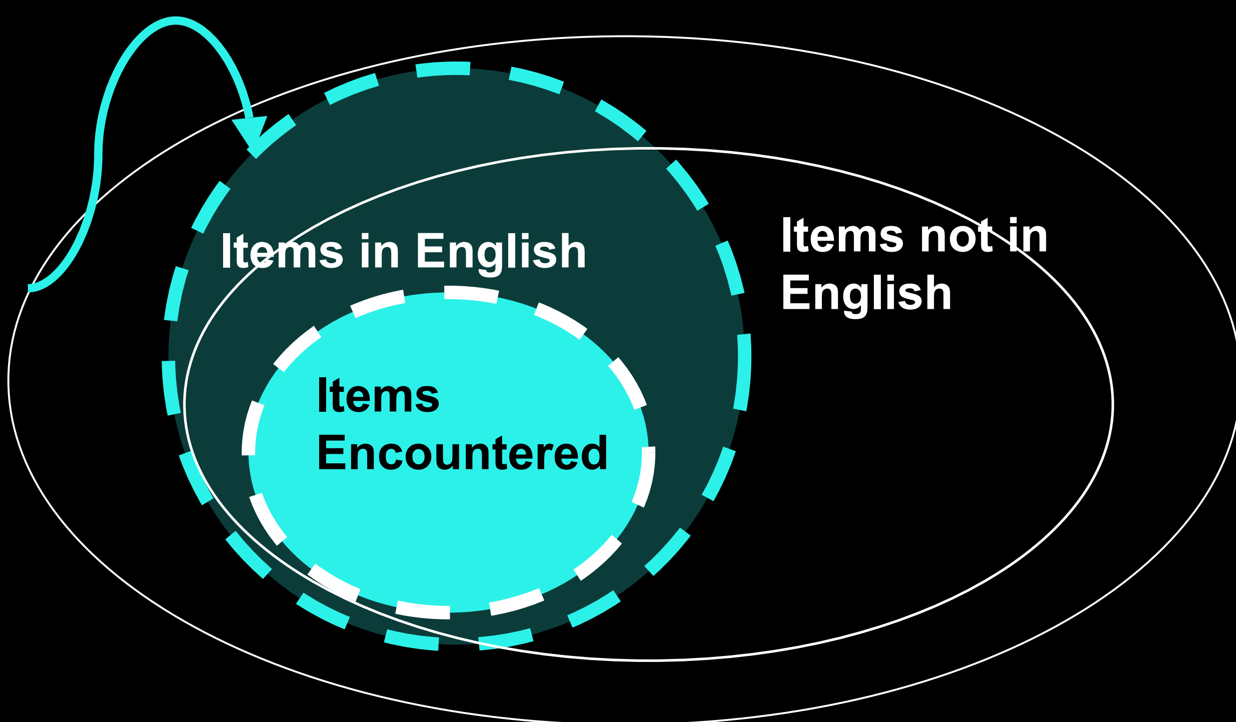
But what if children never behave as if they consider the incorrect hypotheses? That is, they never produce errors compatible with the incorrect hypotheses or accept items that are compatible with the incorrect hypotheses.



Argument for Prior Knowledge

Conclusion: children have some prior knowledge that causes them never to consider the incorrect hypotheses. Instead, they only consider the correct hypothesis for what the rules and patterns of the language might be.

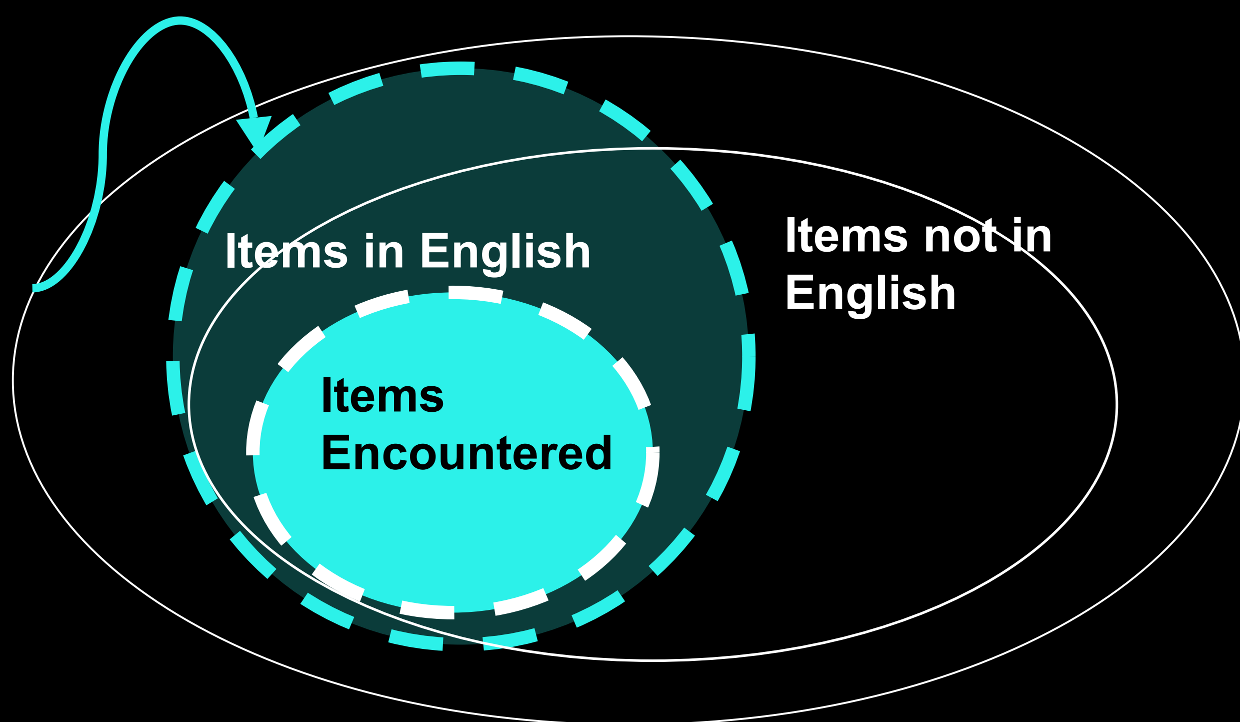
Prior knowledge restricts children's hypothesis to this as the set of items allowed in the language



Argument for Prior Knowledge

Nativist conclusion: the prior knowledge children have is specific to language (domain-specific) and is innate (not derivable from the child's experience). This kind of prior knowledge is sometimes called "Universal Grammar".

Innate knowledge restricts children's hypothesis to this as the set of items allowed in the language



Reasonable questions

- What are some examples of linguistic knowledge that seem to present a poverty of the stimulus situation?

- When there is a poverty of the stimulus situation, what kind of "knowledge" do children need to end up with the right answer? How can we test different ideas of what this knowledge might be?

Reasonable questions

- What are some examples of linguistic knowledge that seem to present a poverty of the stimulus situation?
  - Structure dependence of complex yes/no questions
  - Anaphoric *one*

- When there is a poverty of the stimulus situation, what kind of "knowledge" do children need to end up with the right answer? How can we test different ideas of what this knowledge might be?

Reasonable questions

- What are some examples of linguistic knowledge that seem to present a poverty of the stimulus situation?
  - Structure dependence of complex yes/no questions
  - Anaphoric *one*
- When there is a poverty of the stimulus situation, what kind of "knowledge" do children need to end up with the right answer? How can we test different ideas of what this knowledge might be?
  - Knowledge kinds:
    - Domain-specific or domain-general?
    - Knowledge about linguistic structure or knowledge about how to learn?
  - Computational modeling studies to identify the necessary knowledge

Reasonable questions

- What are some examples of linguistic knowledge that seem to present a poverty of the stimulus situation?
  - Structure dependence of complex yes/no questions
  - Anaphoric *one*

- When there is a poverty of the stimulus situation, what kind of “knowledge” do children need to end up with the right answer? How can we test different ideas of what this knowledge might be?
  - Knowledge kinds:
    - Domain-specific or domain-general?
    - Knowledge about linguistic structure or knowledge about how to learn?
  - Computational modeling studies to identify the necessary knowledge

Structure-Dependence of
Complex Yes/No questions

Structure-Dependence of Complex Yes/No questions

Jareth can alter time.

↓

Can Jareth alter time?

To turn the sentence into a yes/no question, move the auxiliary verb ("can") to the front.

The child's task: figure out a rule that will form yes/no questions from their corresponding sentences.

Structure-Dependence of Complex Yes/No questions

Jareth can alter time.
Can Jareth alter time?

Rule?

Structure-Dependence of Complex Yes/No questions

Jareth can alter time.
Can Jareth alter time?

Rule: Move first auxiliary?

Structure-Dependence of Complex Yes/No questions

Jareth can alter time.
Can Jareth alter time?

Rule: Move first auxiliary?

Anyone who can wish away their brother would be tempted to do it.
Would anyone who can wish away their brother be tempted to do it?

Structure-Dependence of Complex Yes/No questions

Jareth can alter time.
Can Jareth alter time?

Rule: Move first auxiliary?

Rule?

Anyone who can wish away their brother would be tempted to do it.
Would anyone who can wish away their brother be tempted to do it?

Structure-Dependence of Complex Yes/No questions

Jareth can alter time.
Can Jareth alter time?

Rule: Move first auxiliary?

Rule: Move last auxiliary?

Anyone who can wish away their brother would be tempted to do it.
Would anyone who can wish away their brother be tempted to do it?

Structure-Dependence of Complex Yes/No questions

Jareth can alter time.
Can Jareth alter time?

Rule: Move first auxiliary?

Rule: Move last auxiliary?

Anyone who can wish away their brother would be tempted to do it.
Would anyone who can wish away their brother be tempted to do it?

Someone who can solve the labyrinth should show someone else who can't how.
Should someone who can solve the labyrinth show someone else who can't how?

Structure-Dependence of Complex Yes/No questions

Jareth can alter time.
Can Jareth alter time?

Rule: Move first auxiliary?

Rule: Move last auxiliary?

Anyone who can wish away their brother would be tempted to do it.
Would anyone who can wish away their brother be tempted to do it?

Rule???

Someone who can solve the labyrinth should show someone else who can't how.
Should someone who can solve the labyrinth show someone else who can't how?

Structure-Dependence of Complex Yes/No questions

Jareth can alter time.
Can Jareth alter time?

Rule: Move first auxiliary?

Rule: Move last auxiliary?

Anyone who can wish away their brother would be tempted to do it.
Would anyone who can wish away their brother be tempted to do it?

Rule???

Someone who can solve the labyrinth should show someone else who can't how.
Should someone who can solve the labyrinth show someone else who can't how?

Need a rule that is compatible with *all* of these, since they're all grammatical English questions.

Structure-Dependence of Complex Yes/No questions

Jareth can alter time.
Can Jareth alter time?

Anyone who can wish away their brother would be tempted to do it.
Would anyone who can wish away their brother be tempted to do it?

Someone who can solve the labyrinth should show someone else who can't how.
Should someone who can solve the labyrinth show someone else who can't how?

Idea: Try looking at the sentence structure, not just the linear order of the words in the sentences.

Structure-Dependence of Complex Yes/No questions

Jareth can alter time.
Can Jareth alter time?

embedded clauses = additional descriptive sentences that are not part of the main clause

Anyone who <u>can</u> wish away their brother would be tempted to do it.
Would anyone who <u>can</u> wish away their brother be tempted to do it?

Someone who <u>can</u> solve the labyrinth should show someone else who <u>can't</u> how.
Should someone who <u>can</u> solve the labyrinth show someone else who <u>can't</u> how?

Idea: Try looking at the sentence structure, not just the linear order of the words in the sentences.

Structure-Dependence of Complex Yes/No questions

Jareth can alter time.
Can Jareth alter time?

embedded clauses = additional descriptive sentences that are not part of the main clause

Anyone who can wish away their brother would be tempted to do it.
Would anyone who can wish away their brother be tempted to do it?

Someone who can solve the labyrinth should show someone else who can't how.
Should someone who can solve the labyrinth show someone else who can't how?

Let's look just at the main clauses in these examples

Structure-Dependence of Complex Yes/No questions

Jareth can alter time.
Can Jareth alter time?

Anyone would be tempted to do it.
Would anyone be tempted to do it?

Someone should show someone else how.
Should someone show someone else how?

Let's look just at the main clauses in these examples

Structure-Dependence of Complex Yes/No questions

Jareth can alter time.
Can Jareth alter time?

Anyone would be tempted to do it.
Would anyone be tempted to do it?

Someone can show someone else how.
Can someone show someone else how?

Let’s look just at the main clauses in these examples

Structure-Dependence of Complex Yes/No questions

Jareth can alter time.
Can Jareth alter time?

Anyone would be tempted to do it.
Would anyone be tempted to do it?

Someone can show someone else how.
Can someone show someone else how?

Rule that works for all of these examples (and all English examples):
Move the auxiliary verb in the main clause to make a yes/no question.

This is a rule dependent on the structure of the sentences.

Children's Knowledge of Complex Yes/No Questions

Children seem to know this rule by the age of 3.
(Crain & Nakayama 1987)

Learning problem: Children don't encounter all the examples we saw. They encounter a subset of the possible yes/no questions in English.

Most of the data they encounter (particularly before the age of 3) consists of simple yes/no questions.

Jareth can alter time.
Can Jareth alter time?

Legate & Yang (2002) estimate less than 1% of child-directed speech data consist of informative complex questions. This may be too few for children to notice.

Poverty of the Stimulus: Yes/No Questions

If children only encounter (or notice) simple yes/no questions, the problem is that these simple yes/no questions are compatible with a lot of different rules. How do children know the right way to generalize?

Rule: Move first auxiliary?

Jareth can alter time.
Can Jareth alter time?

Rule: Move last auxiliary?

Rule: Move main clause auxiliary?

Rule: Move auxiliary in even-numbered position in sentence?

Rule: Move auxiliary closest to a noun?

Anaphoric *One*

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!

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. 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

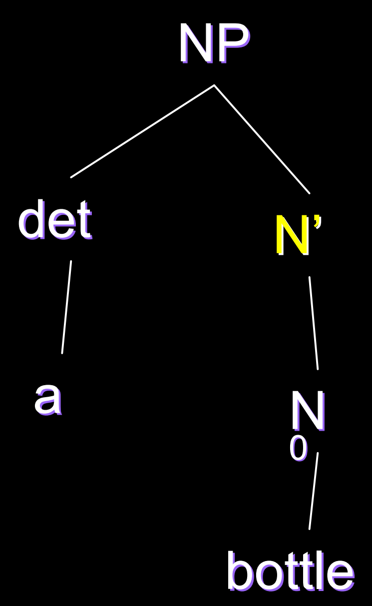
As adults, we have knowledge about what the antecedent of *one* can be in various situations. In particular, we have knowledge about the syntactic category of *one*. The antecedent of *one* must be the same syntactic category of *one*, or else *one* couldn't replace it. So what syntactic category is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

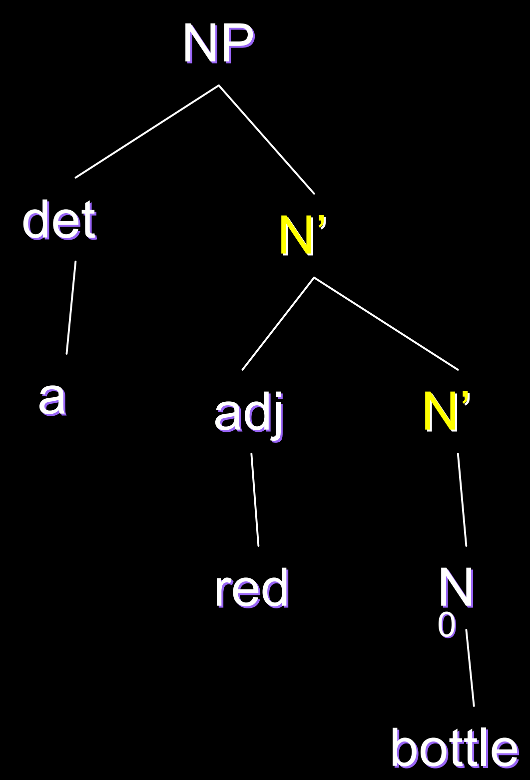
"Look – a red bottle! Do you see another *one*?"

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

Anaphoric *One*: Syntactic Category

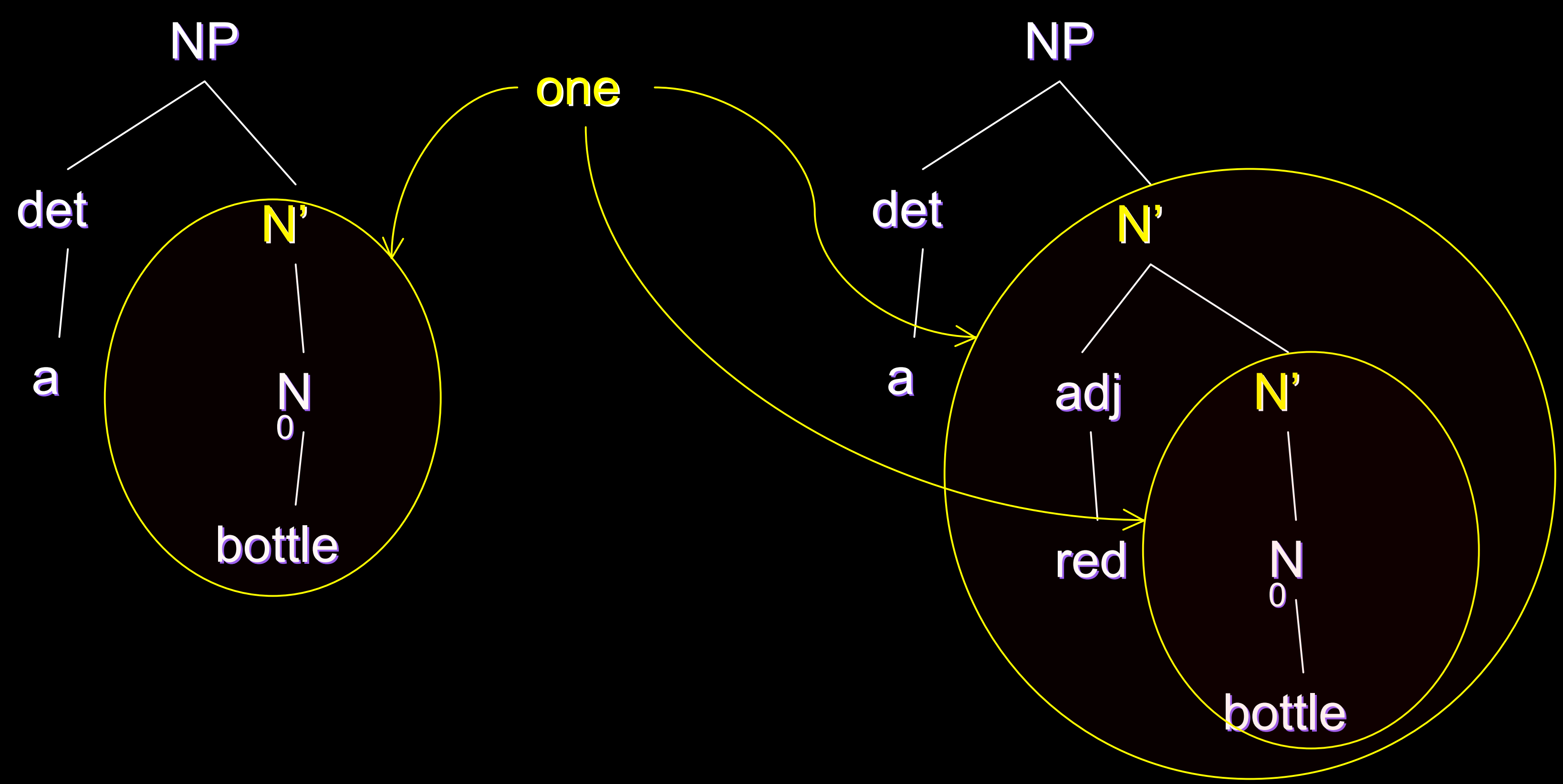
Many linguists believe that *one* in these kind of utterances is a syntactic category smaller than an entire noun phrase, but larger than just a noun (N^0). This category is sometimes called N'. This category includes strings like "bottle" and "red bottle".





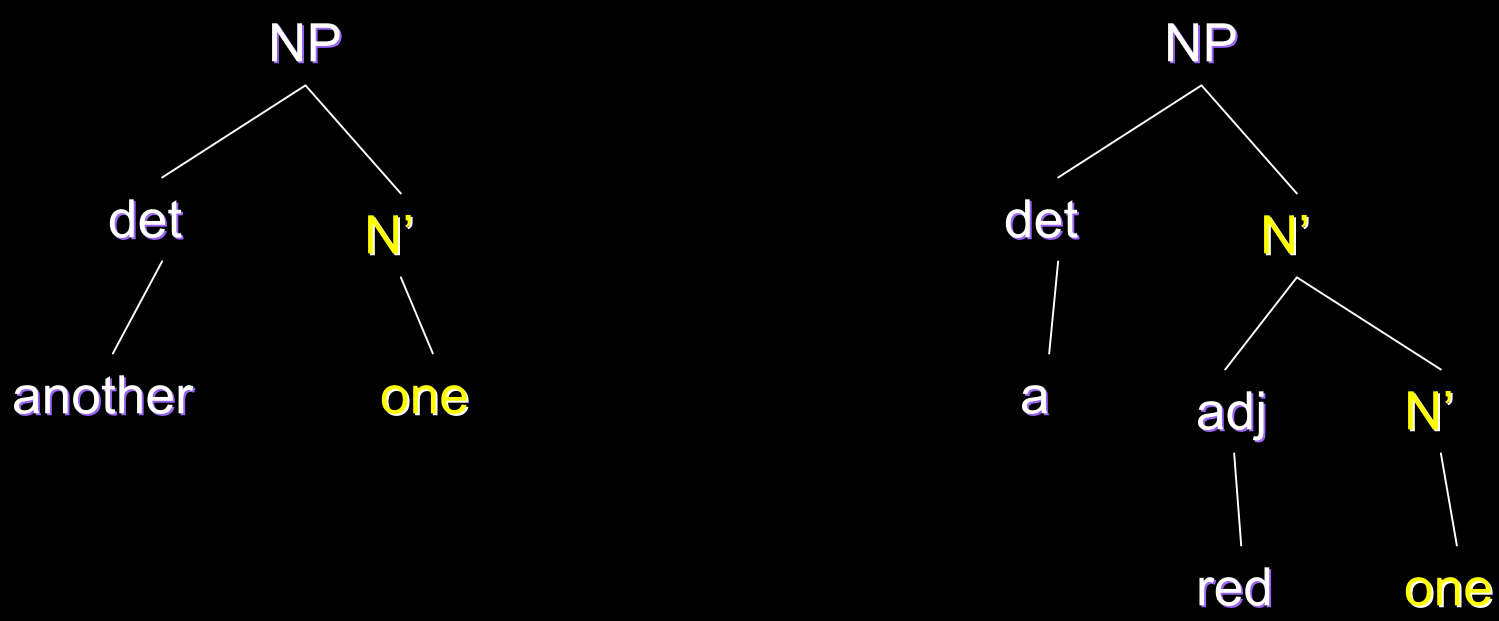
Anaphoric *One*: Syntactic Category

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".



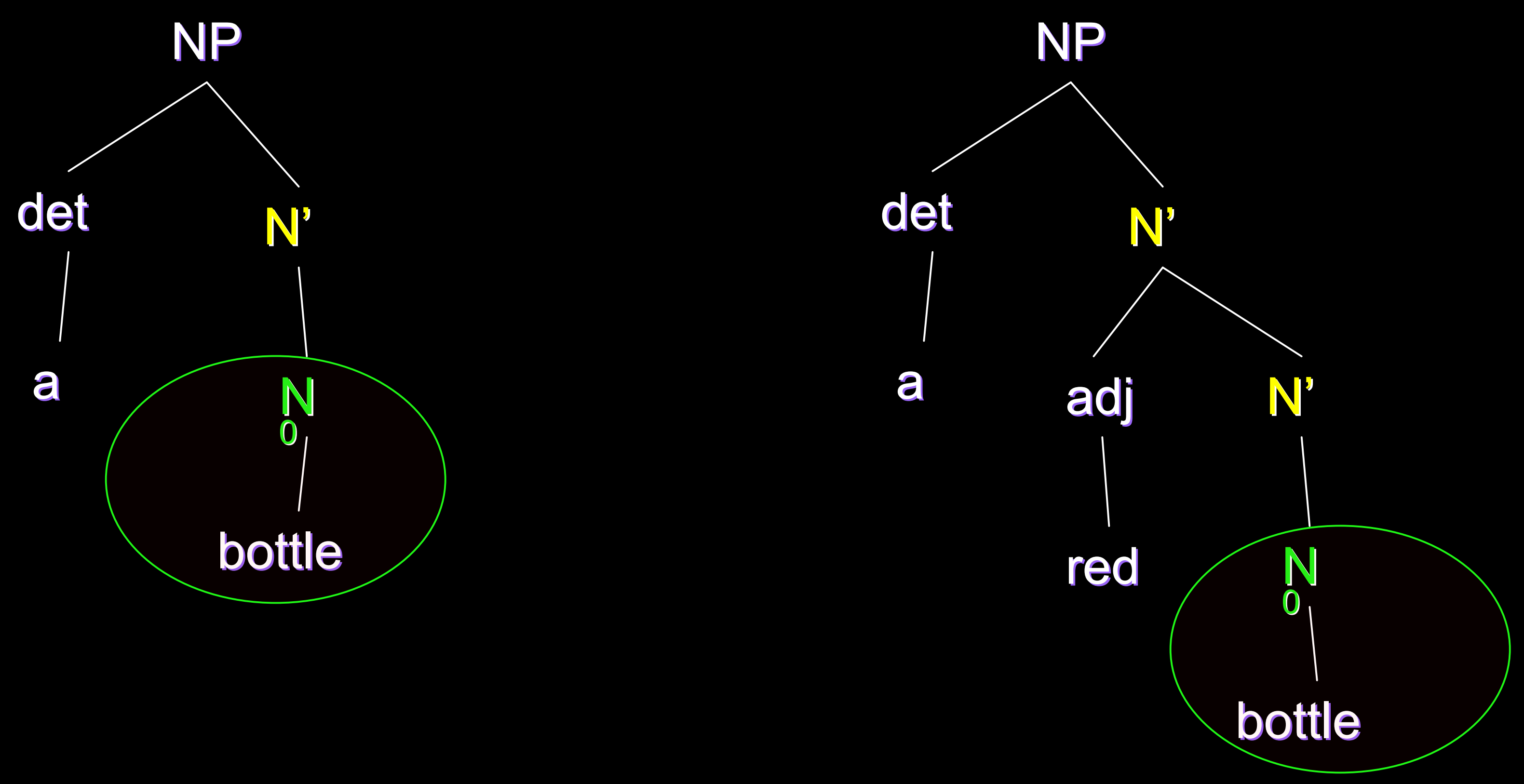
Anaphoric *One*: Syntactic Category

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".



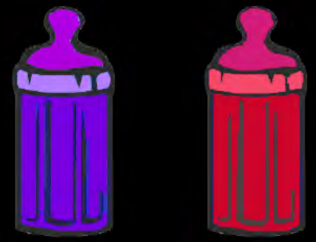
Anaphoric *One*: Syntactic Category

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”).

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.

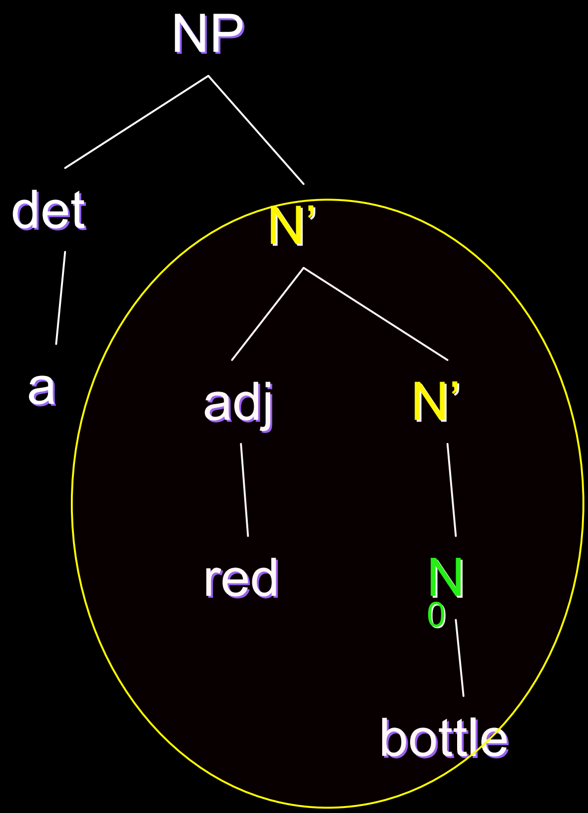
"Look – a red bottle! Do you see another one?"

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:

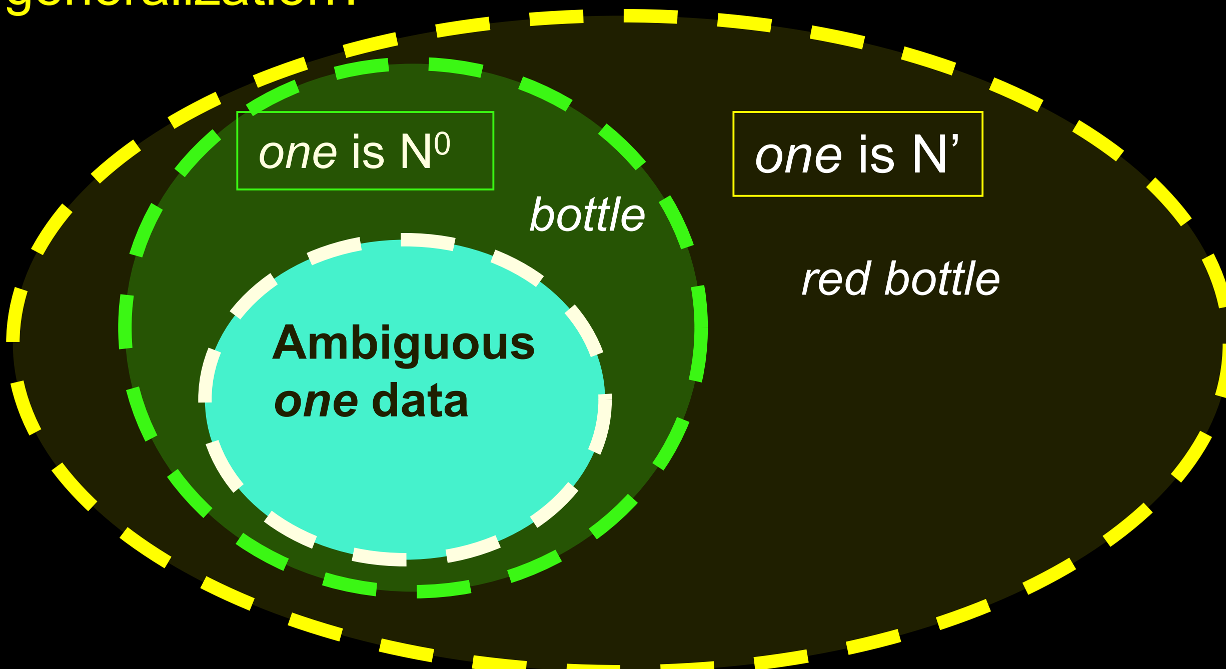
"Look – a red bottle! We want another *one*, but there doesn't seem to be *one* here."

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, 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?



Reasonable questions

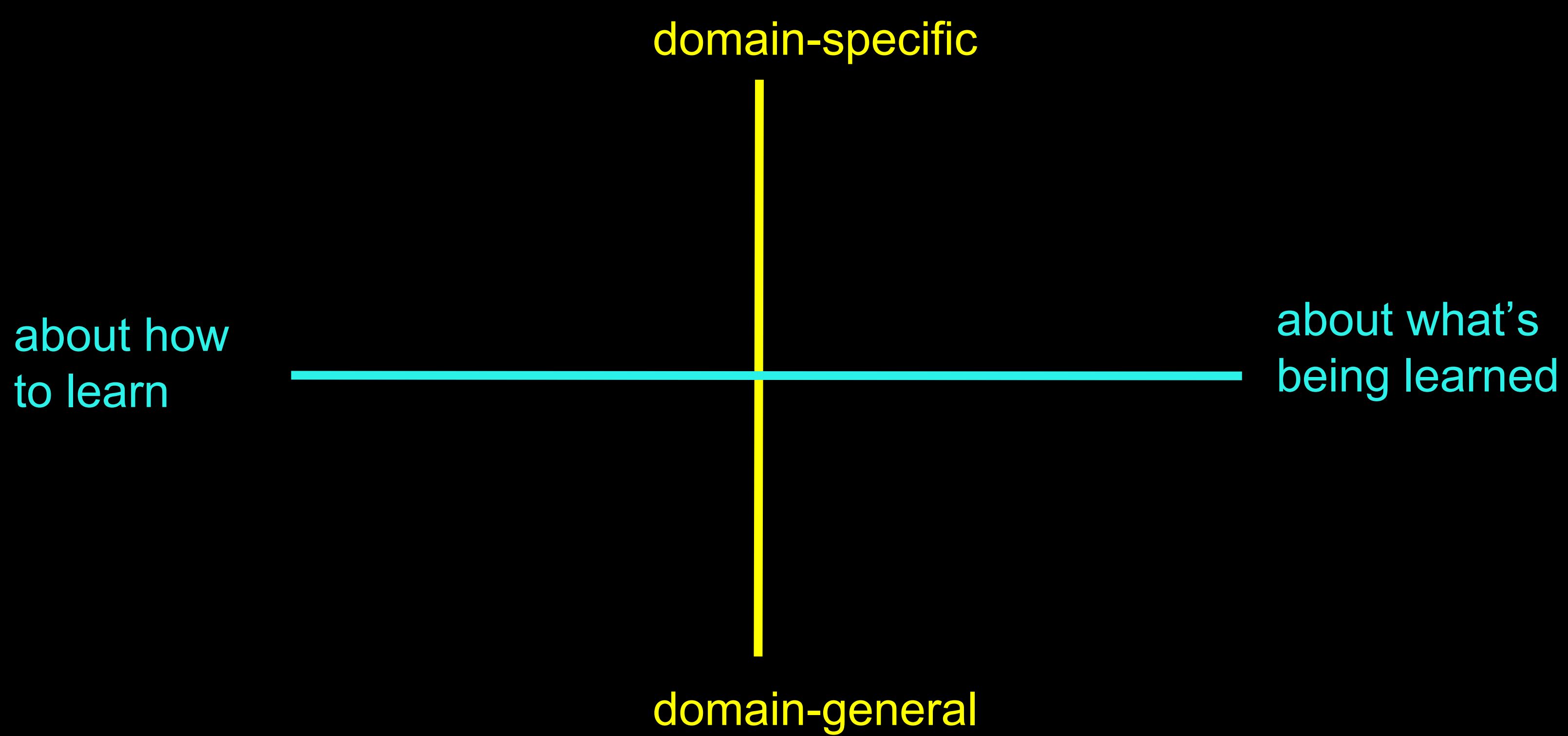
- What are some examples of linguistic knowledge that seem to present a poverty of the stimulus situation?
  - Structure dependence of complex yes/no questions
  - Anaphoric *one*

- When there is a poverty of the stimulus situation, what kind of "knowledge" do children need to end up with the right answer? How can we test different ideas of what this knowledge might be?
  - Knowledge kinds:
    - Domain-specific or domain-general?
    - Knowledge about linguistic structure or knowledge about how to learn?
  - Computational modeling studies to identify the necessary knowledge

Knowledge kinds

Two different examples: complex yes/no question formation, anaphoric *one*

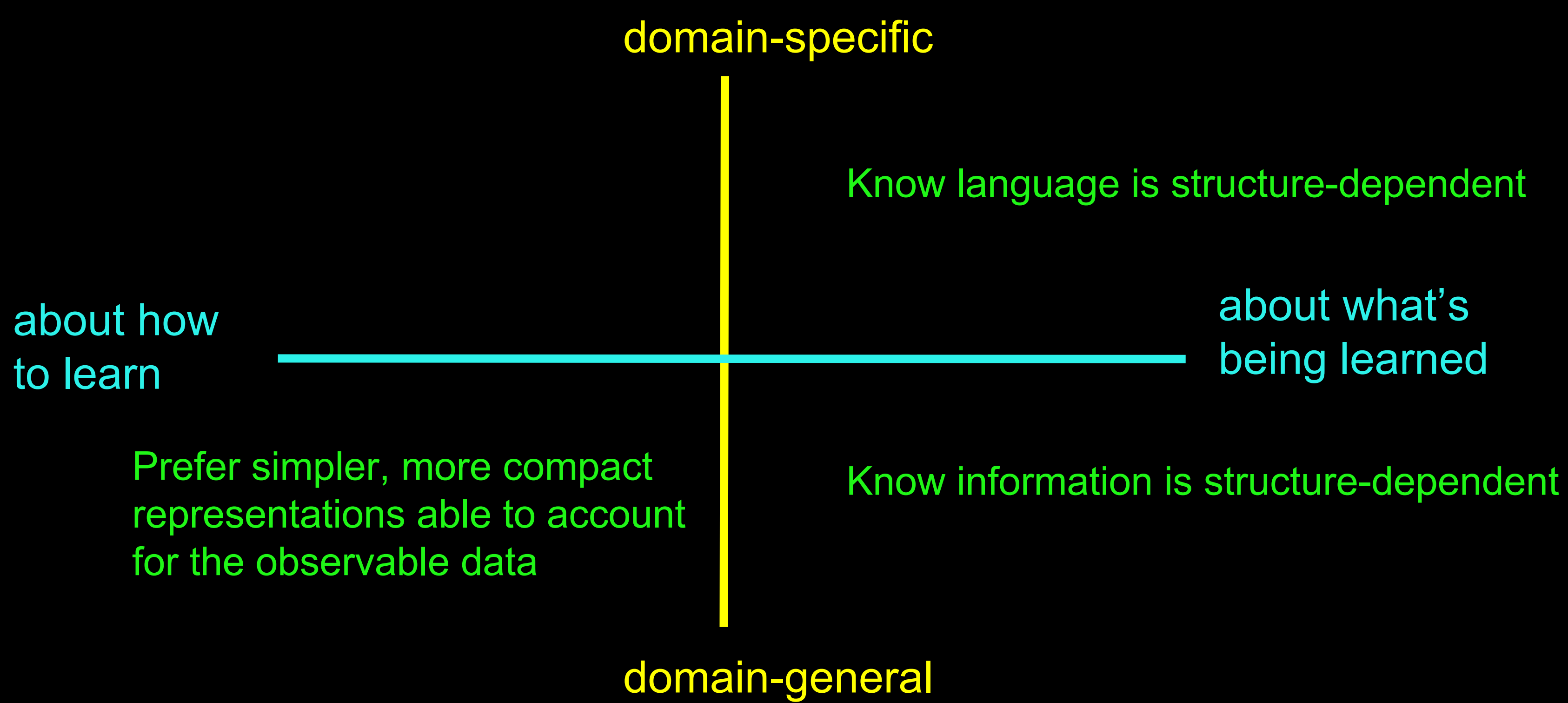
For each one, we want to know what kind of knowledge is needed to learn the correct generalizations:



Knowledge kinds

Two different examples: complex yes/no question formation, anaphoric *one*

For each one, we want to know what kind of knowledge is needed to learn the correct generalizations:



Knowledge kinds

Two different examples: complex yes/no question formation, anaphoric *one*

For each one, we want to know what kind of knowledge is needed to learn the correct generalizations:

domain-specific

Prefer syntactic information over semantic information

Know *one* cannot be N^0

about how to learn

about what's being learned

Prefer unambiguous data

Prefer more probable representations, based on data

domain-general

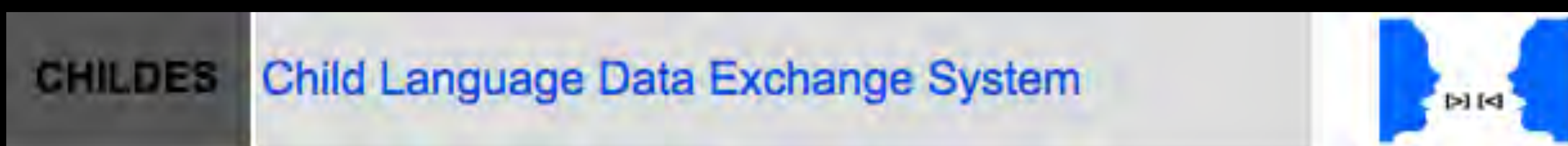
Computational Modeling: What a "Digital" Child Can Tell Us

If we know the problem the child must solve, we have information about these:

– what the target knowledge state is
  - Form complex yes/no questions correctly
  - Interpret anaphoric *one* correctly

– what data the child is learning from
  - Child-directed speech data

http://childes.psy.cmu.edu/



This is what led us to know we have a poverty of the stimulus situation in the first place.

Computational Modeling: What a "Digital" Child Can Tell Us

We can then construct a model where we have precise control over these:

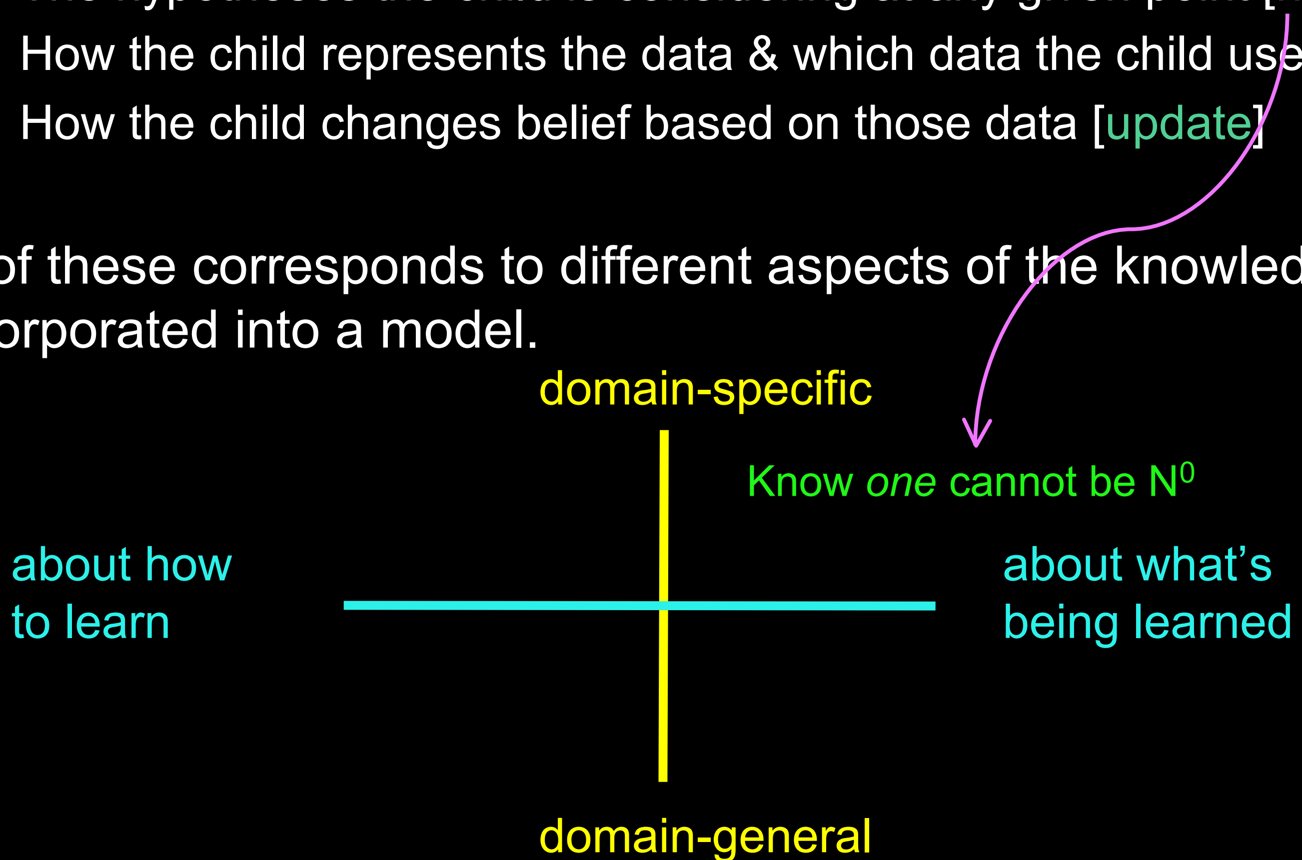
- The hypotheses the child is considering at any given point [hypothesis space]
- How the child represents the data & which data the child uses [data intake]
- How the child changes belief based on those data [update]

Computational Modeling: What a "Digital" Child Can Tell Us

We can then construct a model where we have precise control over these:

- The hypotheses the child is considering at any given point [hypothesis space]
- How the child represents the data & which data the child uses [data intake]
- How the child changes belief based on those data [update]

Each of these corresponds to different aspects of the knowledge that can be incorporated into a model.

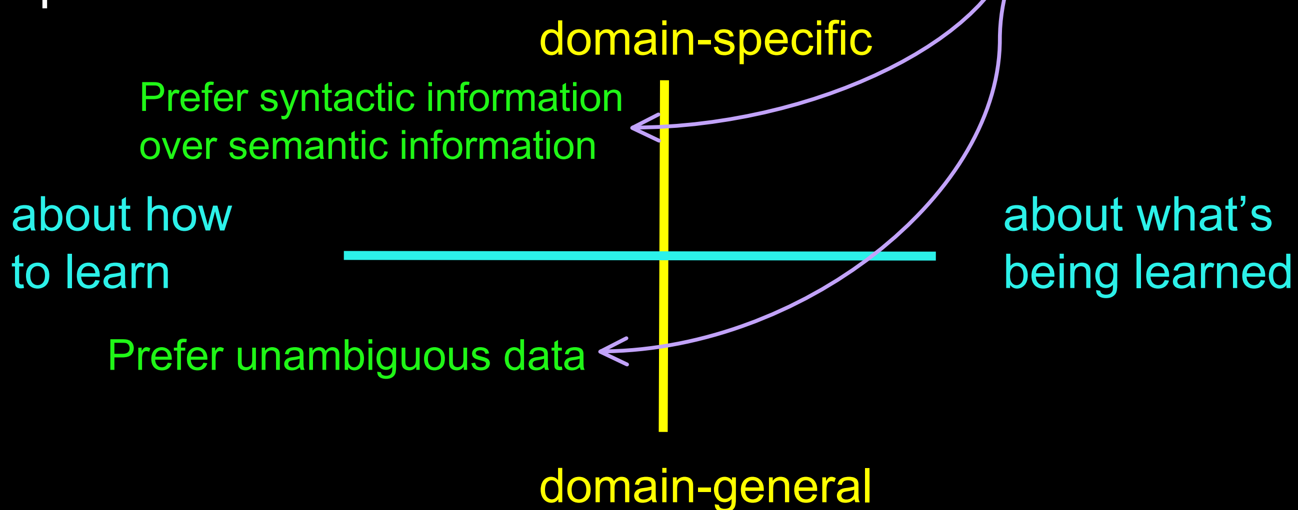


Computational Modeling:
What a "Digital" Child Can Tell Us

We can then construct a model where we have precise control over these:

- The hypotheses the child is considering at any given point [hypothesis space]
- How the child represents the data & which data the child uses [data intake]
- How the child changes belief based on those data [update]

Each of these corresponds to different aspects of the knowledge that can be incorporated into a model.

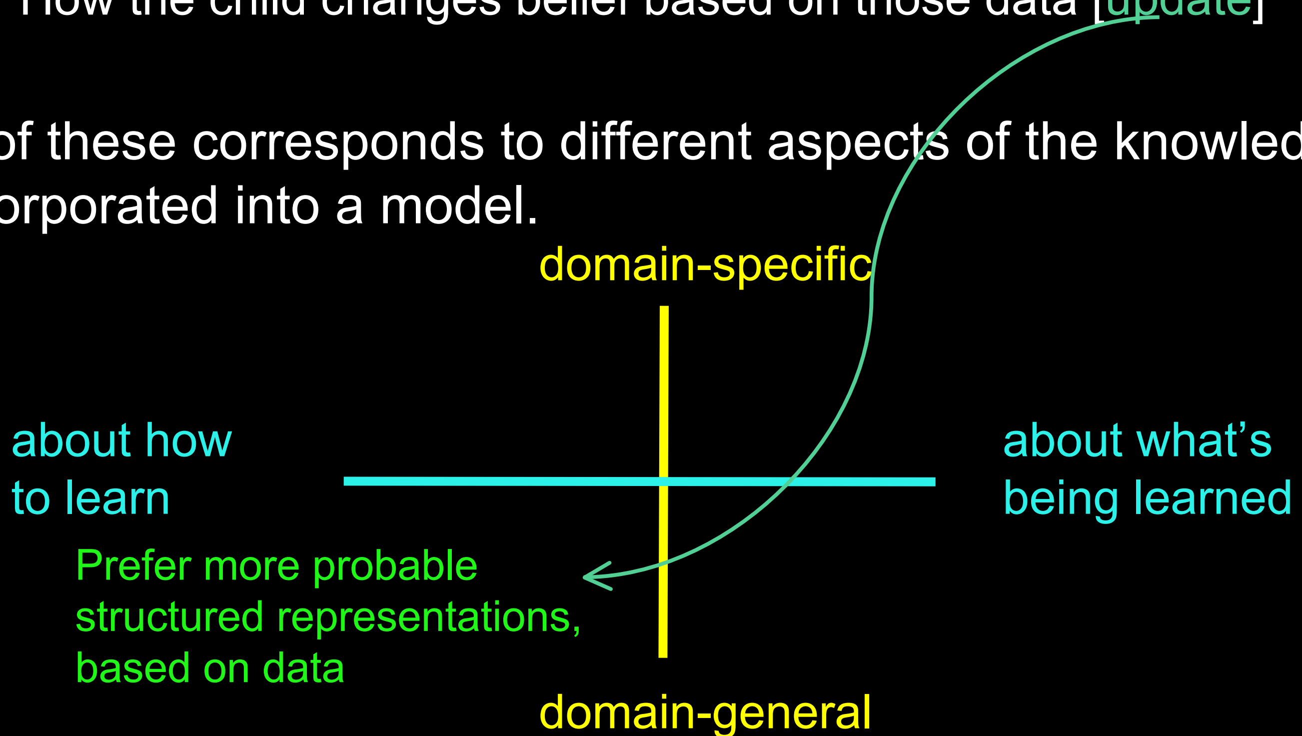


Computational Modeling:
What a "Digital" Child Can Tell Us

We can then construct a model where we have precise control over these:

- The hypotheses the child is considering at any given point [hypothesis space]
- How the child represents the data & which data the child uses [data intake]
- How the child changes belief based on those data [update]

Each of these corresponds to different aspects of the knowledge that can be incorporated into a model.



Computational Modeling:
What a "Digital" Child Can Tell Us

Models are most informative when they're grounded empirically.

This is why most models make use of the child-directed speech data available through databases like CHILDES.

Many models will try to make cognitively plausible assumptions about how the child is representing and processing input data:

- Processing data points as they are encountered
- Assuming children have memory limitations (ex: memory of data points may decay over time)

General Modeling Process

(1) Decide what kind of learner the model represents (ex: normally developing 14- to 18-month-old child learning first language)

General Modeling Process

(1) Decide what kind of learner the model represents (ex: normally developing 14- to 18-month-old child learning first language)

(2) Decide what data the child learns from (ex: Suppes corpus from CHILDES) and how the child processes that data (ex: anaphoric *one* data points identified as unambiguous or ambiguous)

General Modeling Process

(1) Decide what kind of learner the model represents (ex: normally developing 14- to 18-month-old child learning first language)

(2) Decide what data the child learns from (ex: Suppes corpus from CHILDES) and how the child processes that data (ex: anaphoric *one* data points identified as unambiguous or ambiguous)

(3) Decide what hypotheses the child has (ex: *one* is N^0 or *one* is N' for these data points) and what information is being tracked in the input (ex: frequency of informative *one* data points)

General Modeling Process

(1) Decide what kind of learner the model represents (ex: normally developing 14- to 18-month-old child learning first language)

(2) Decide what data the child learns from (ex: Suppes corpus from CHILDES) and how the child processes that data (ex: anaphoric *one* data points identified as unambiguous or ambiguous)

(3) Decide what hypotheses the child has (ex: *one* is N^0 or *one* is N' for these data points) and what information is being tracked in the input (ex: frequency of informative *one* data points)

(4) Decide how belief in different hypotheses is updated (ex: use Bayesian inference to infer most likely hypothesis, given observed data)

General Modeling Process

(5) Decide what the measure of success is

ex: making correct generalizations

- preferring a correctly formed complex yes/no question over an incorrectly formed one

ex: achieving a certain knowledge state by the end of the learning period

- knowing that *one* is syntactic category N'
- matching child behavioral data (ex: preference for looking at the red bottle in "Look – a red bottle! Do you see another one?")

Some example computational models

A basis for many recent computational models: Probabilistic reasoning by children

- Recent experimental studies have shown that infants and young children seem to behave as if they are capable of a type of domain-general probabilistic reasoning known as Bayesian inference

  - Gerken (2006): 9-month-old infants

  - Xu & Tenenbaum (2007): 3- and 4-year-old children

Bayesian Inference: Size Principle

- A Bayesian learner can assign a probability to any hypothesis under consideration by balancing two things:
  - The prior probability of that hypothesis being correct
  - The likelihood of that hypothesis producing the observed data

$$P(\text{hypothesis} \mid \text{data}) \propto P(\text{hypothesis}) * P(\text{data} \mid \text{hypothesis})$$

Bayesian Inference: Size Principle

- A Bayesian learner can assign a probability to any hypothesis under consideration by balancing two things:
  - The prior probability of that hypothesis being correct
  - The likelihood of that hypothesis producing the observed data

$$P(\text{hypothesis} \mid \text{data}) \propto P(\text{hypothesis}) * P(\text{data} \mid \text{hypothesis})$$

The likelihood calculation allows a Bayesian learner to follow the Size Principle (Tenenbaum & Griffiths 2001), and automatically prefer less-general hypotheses (which correspond to sets of smaller size) to more-general hypotheses (which correspond to sets of larger size). This is sometimes referred to as a sensitivity to “suspicious coincidences” (Xu & Tenenbaum 2007).

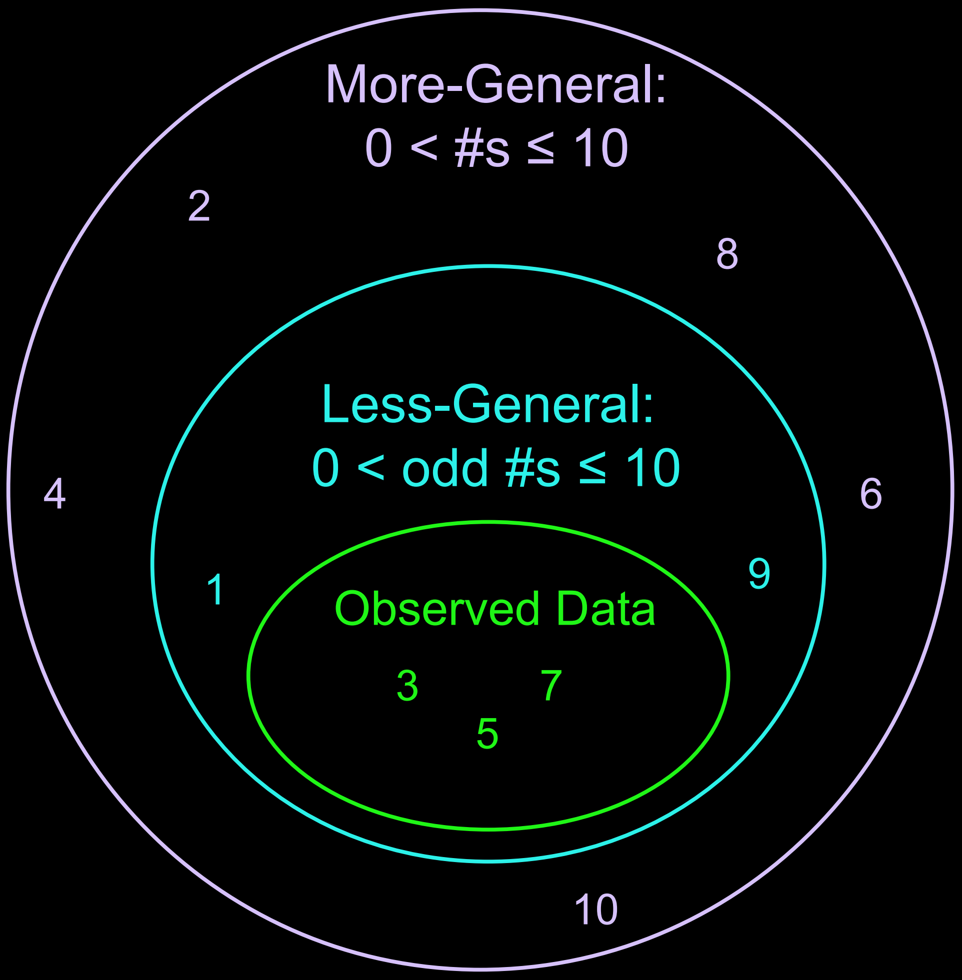
Size Principle: A Numerical Example

Suppose we observe the numbers 3, 5, and 7,

and we're considering two hypotheses for the right "number rule":

(1) Odd numbers between 0 and 10

(2) All numbers between 0 and 10



Size Principle: A Numerical Example

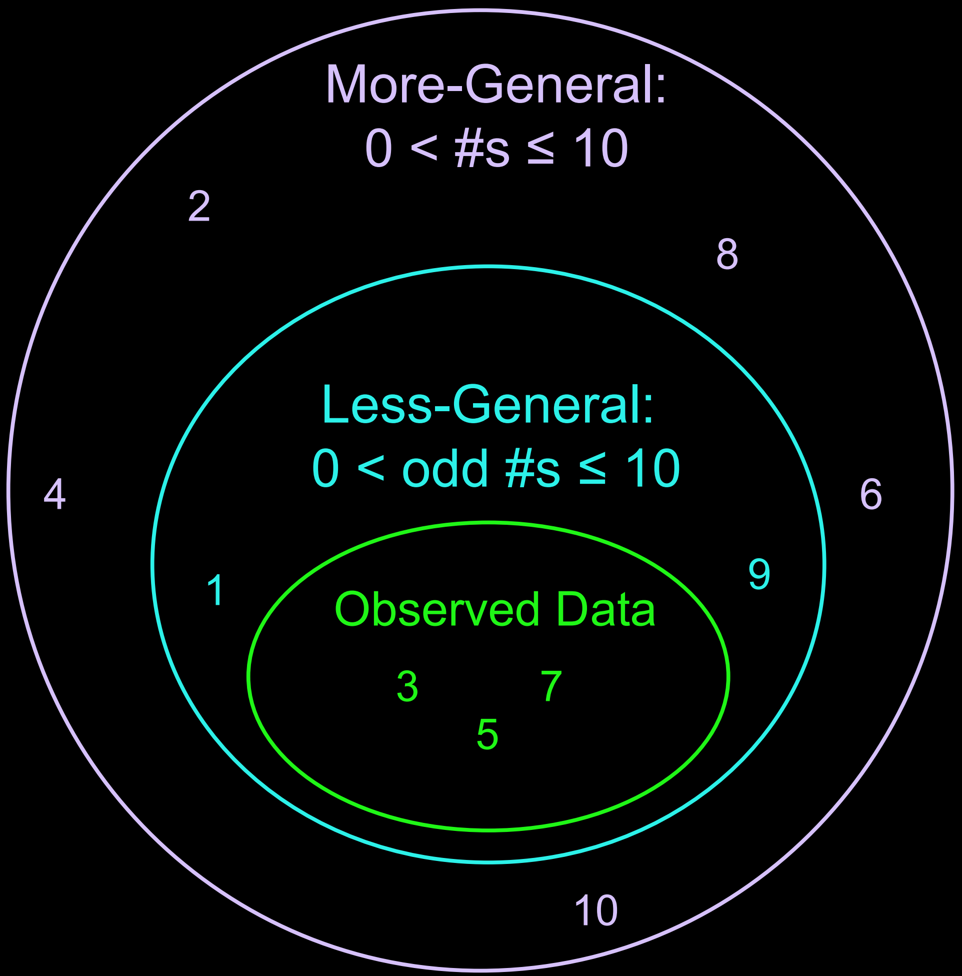
What's the likelihood that we would observe these three numbers if we were drawing from the less-general hypothesis of odd numbers between 1 and 10?

P(3) = 1/5

P(5) = 1/5

P(7) = 1/5

Total likelihood of observed data = 1/5*1/5*1/5 = 1/125 = .008



Size Principle: A Numerical Example

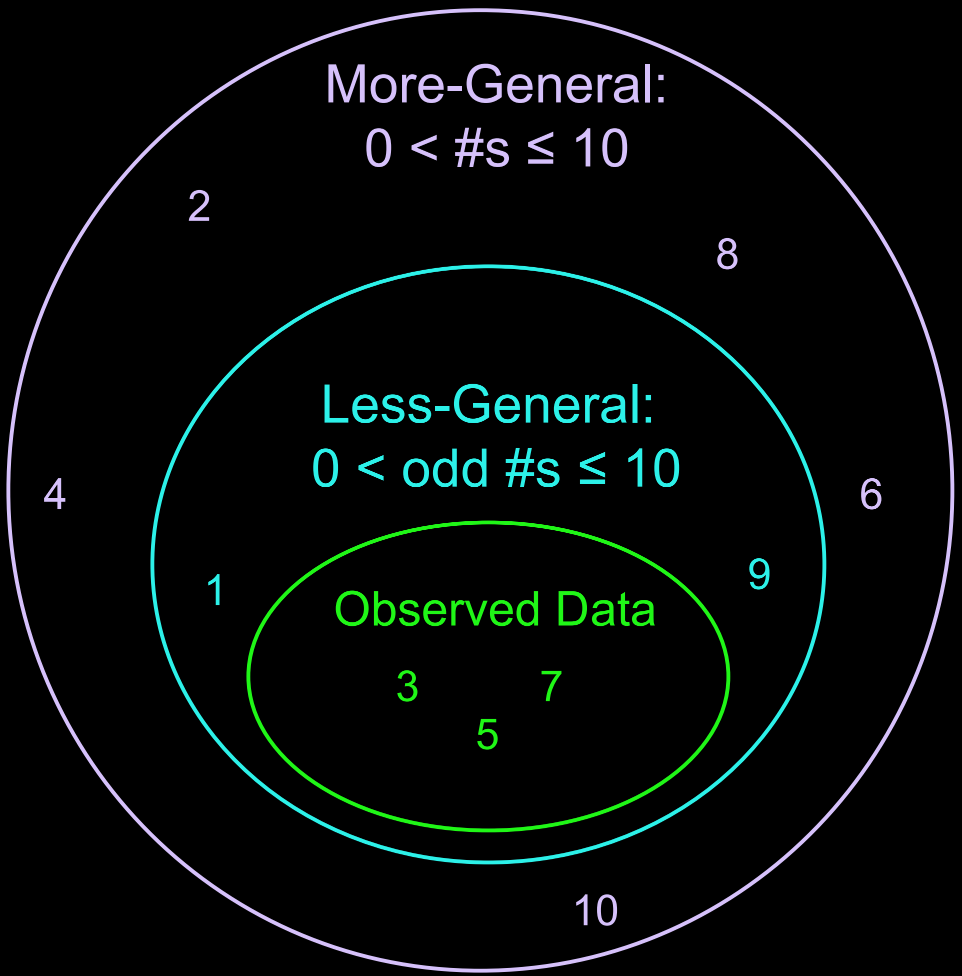
What's the likelihood that we would observe these three numbers if we were drawing from the more-general hypothesis of all numbers between 1 and 10?

P(3) = 1/10

P(5) = 1/10

P(7) = 1/10

Total likelihood of observed data = 1/10*1/10*1/10 = 1/1000 = .001



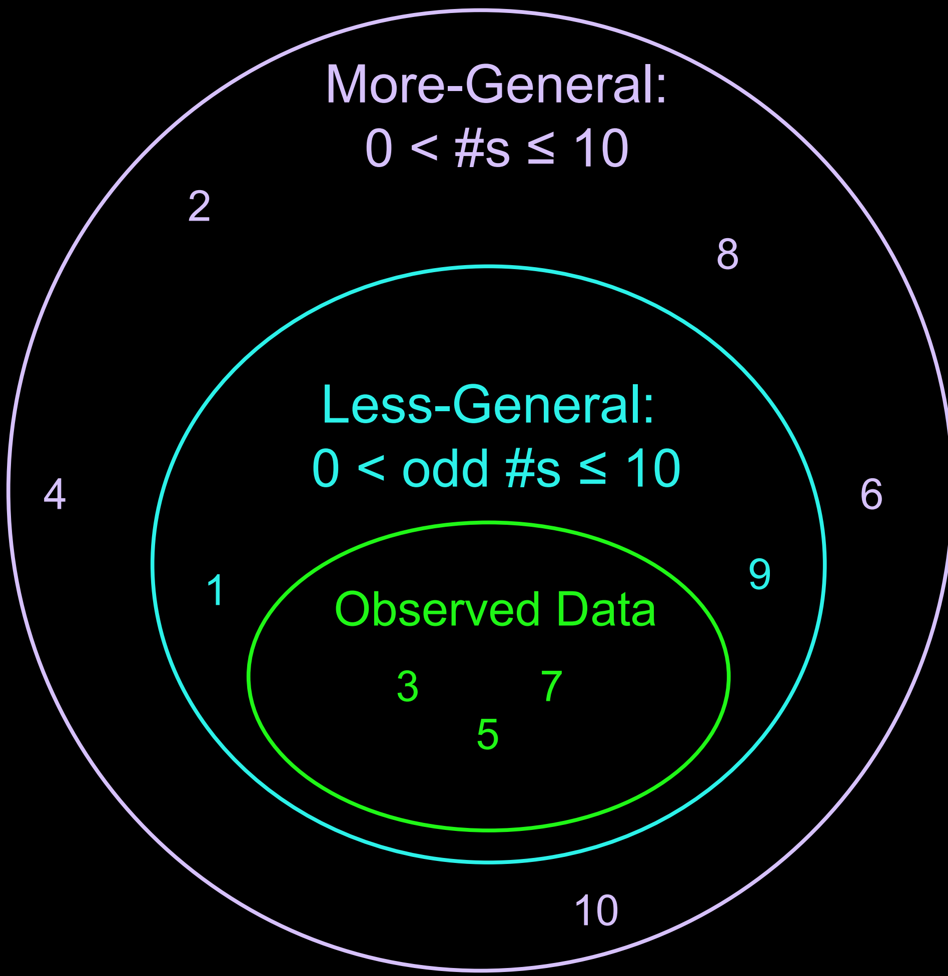
Size Principle: A Numerical Example

A Bayesian learner sensitive to the likelihood of observing the data, given each hypothesis, would thus prefer the less-general hypothesis to the more-general hypothesis.

Likelihood of less-general hypothesis = .008

Likelihood of more-general hypothesis = .001

It is a “suspicious coincidence” to keep seeing odd numbers if the correct hypothesis is that both odd and even numbers are allowed.



Size Principle: A Numerical Example

Moreover, the more odd numbers we observe, the more certain a Bayesian learner is that the less-general hypothesis is correct.

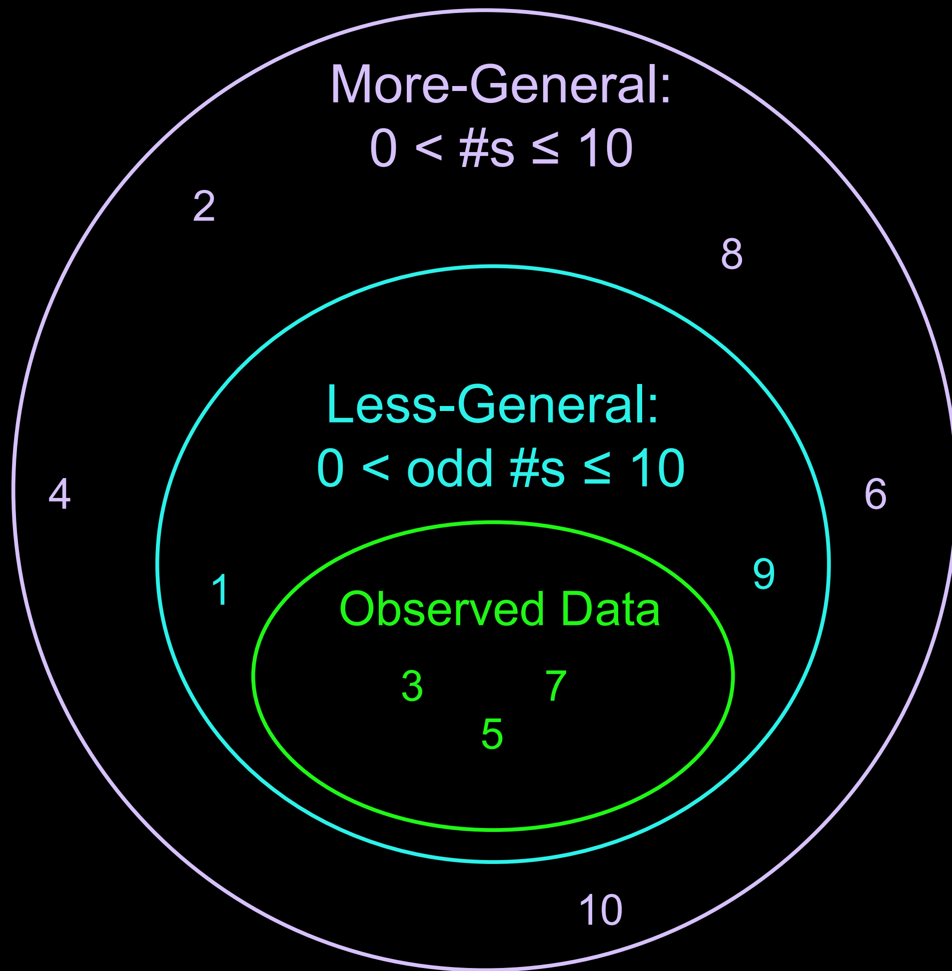
Suppose we see 10 odd numbers:

Likelihood of less-general hypothesis = $(1/5)^{10}$

= .00032

Likelihood of more-general hypothesis = $(1/10)^{10}$

= .0000000001



Gerken (2006): Artificial Language Data

	di	je	li	we
le	leledi	leleje	leleli	lelewe
wi	wiwidi	wiwije	wiwili	wiwiwe
ji	jijidi	jijije	jijili	jijiwe
de	dededi	dedeje	dedeli	dedewe

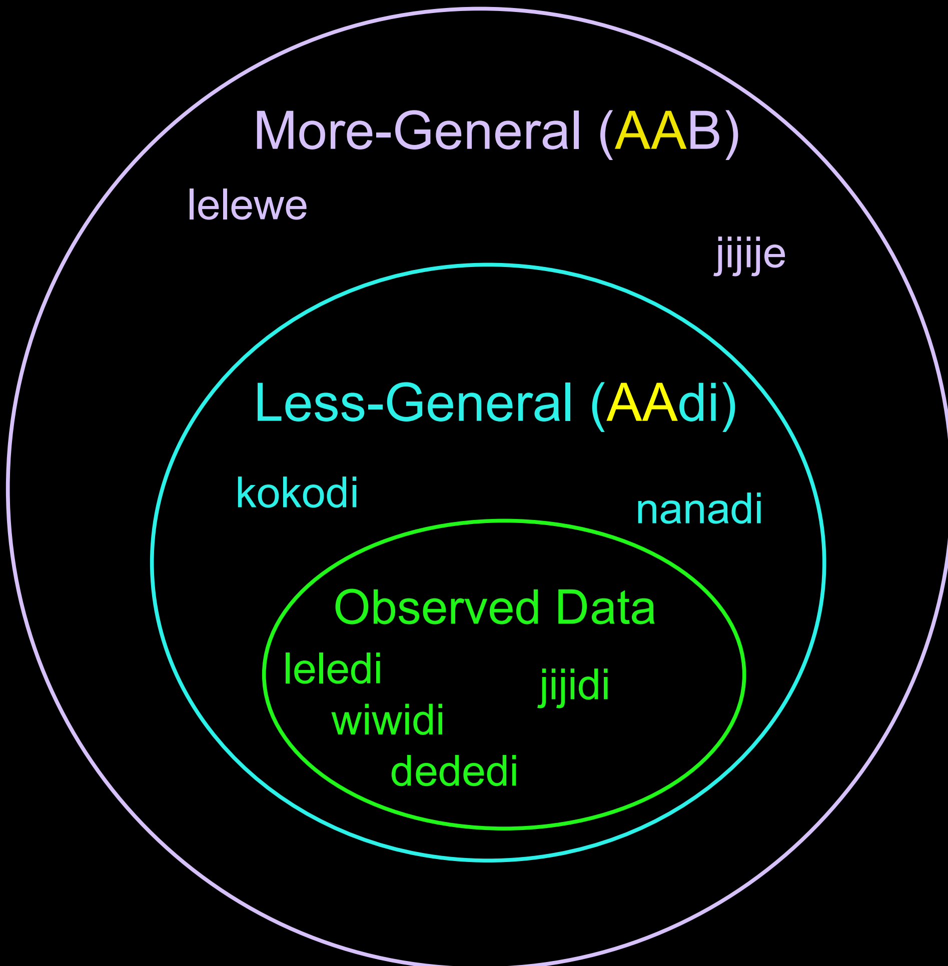
Gerken (2006): Data Subset

	di	je	li	we
le	leledi	leleje	leleli	lelewe
wi	wiwidi	wiwije	wiwili	wiwiwe
ji	jijidi	jijije	jijili	jijiwe
de	dededi	dedeje	dedeli	dedewe

9-month-olds infants exposed to a subset of this dataset which was compatible both with a less-general pattern (AAdi) and a more-general pattern (AAB).

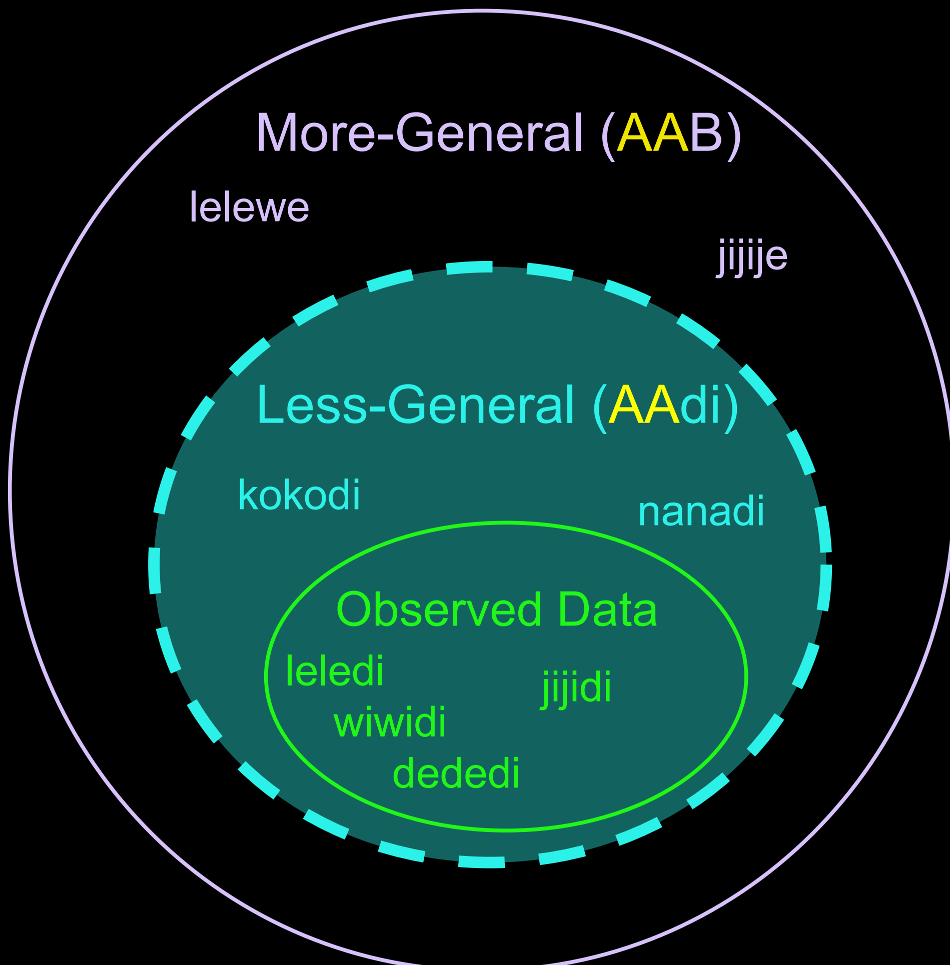
Gerken (2006): Infant Choices

The hypotheses infants could have, given this data subset



Gerken (2006): Infant Choices

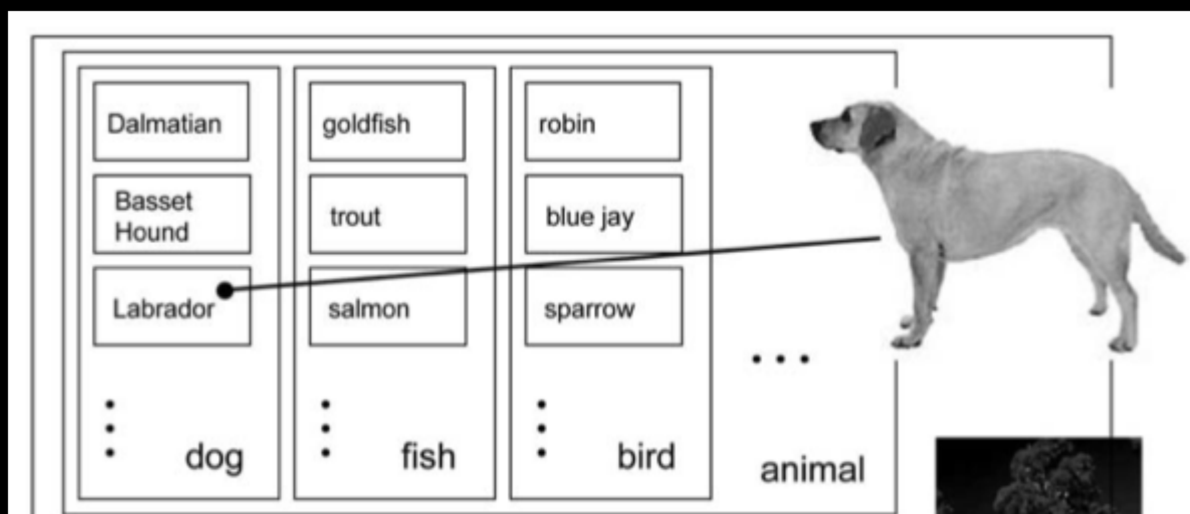
Gerken found that infants prefer to be conservative and make the less-general generalization, given the data subset. This is what the Size Principle predicts.



Xu & Tenenbaum (2007): Word-Meaning Mapping

Object-kind labels: *dog* vs. *dalmatian*

Issue: clearly overlapping labels – a dalmatian is a dog, but not all dogs are dalmatians. Which conceptual level does each label apply to?

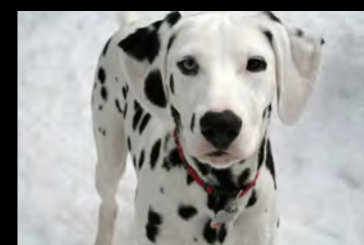


Xu & Tenenbaum (2007): Experiment

3- and 4-year-old children were told they were learning a new language that had different names for things. They were shown examples of objects with nonsense names like *fep* and later asked to show what they thought those novel names referred to.

fep

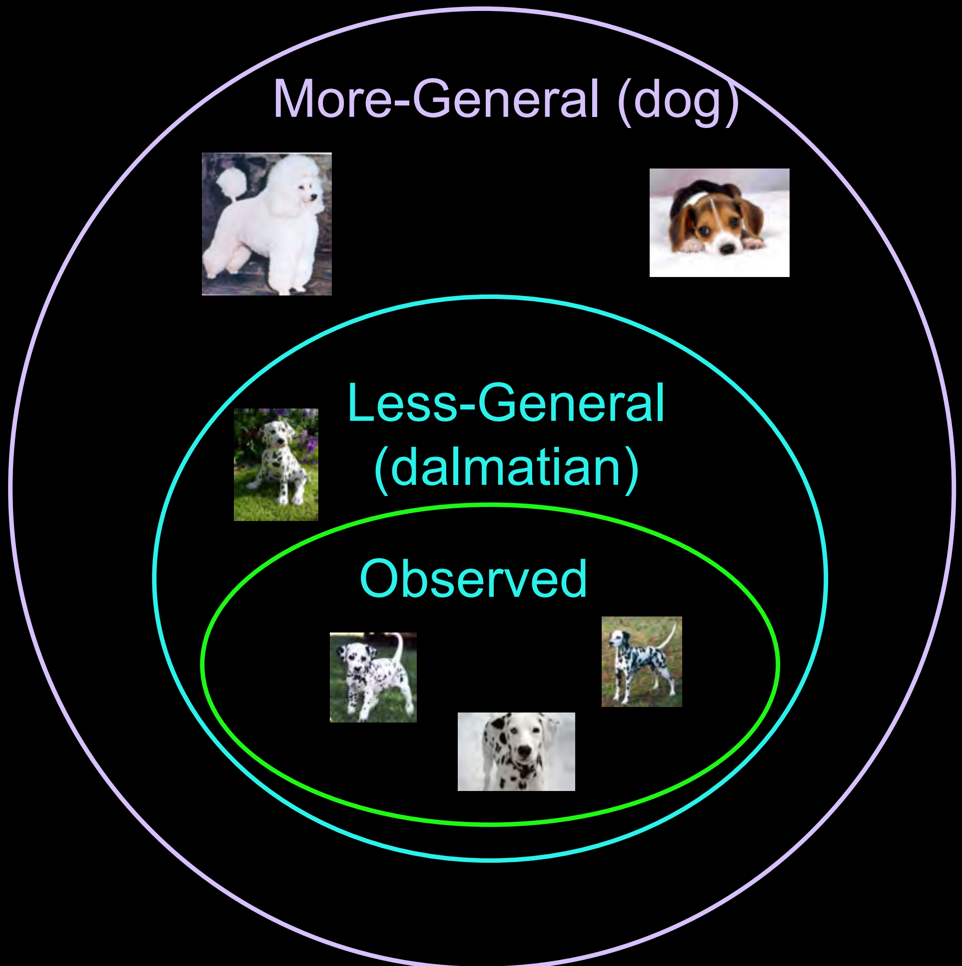
fep



fep

What does *fep* mean?

One way to think about children's choices



Children’s choices consistent with Bayesian reasoning

Xu & Tenenbaum (2007) found that children seemed sensitive to suspicious coincidences, and followed the Size Principle by choosing the less-general hypothesis.

Back to the models

Several recent computational models have attempted to address poverty of the stimulus questions, and rely on Bayesian inference as the main method of learning. By modeling the acquisition process for these linguistic phenomena, these models hope to pinpoint the kind of knowledge required for language acquisition.

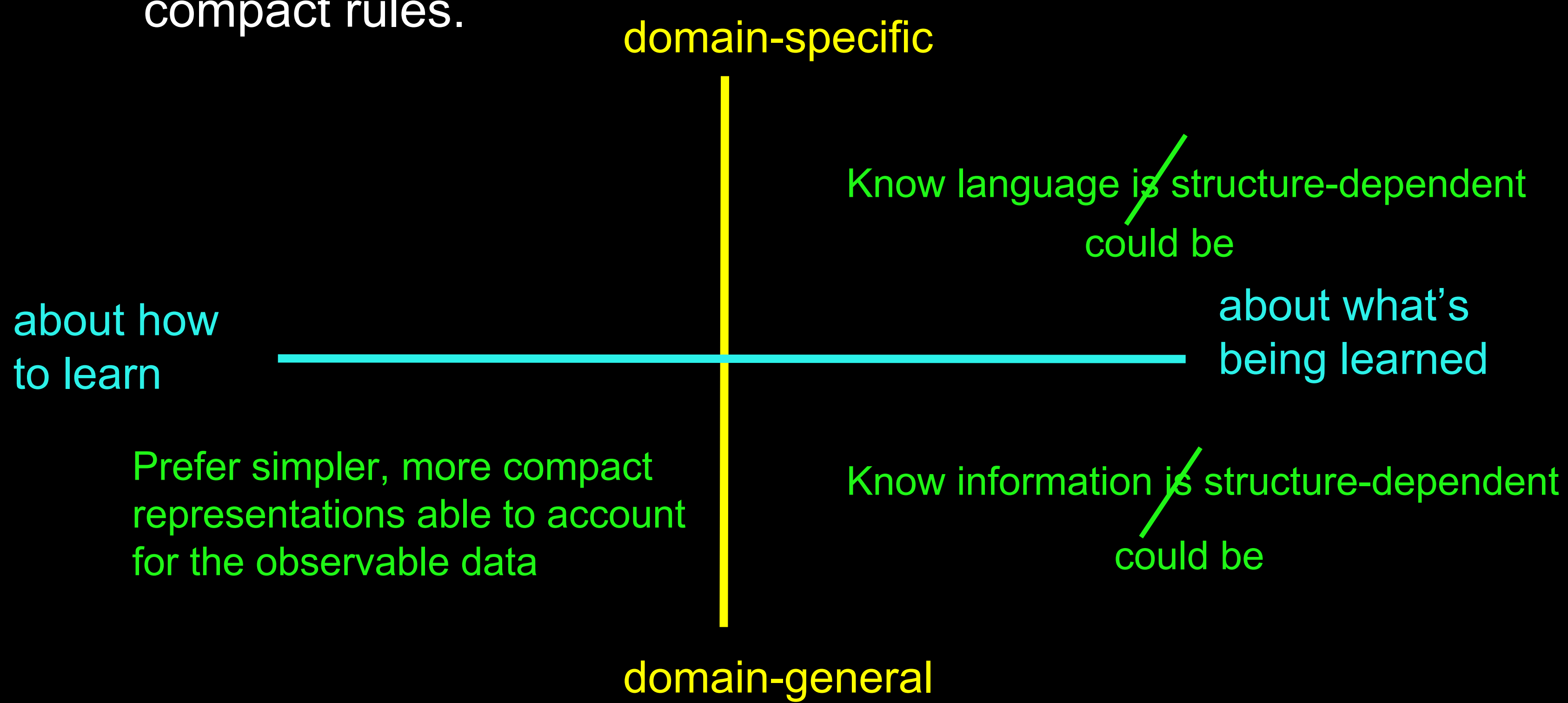
- Structure-dependence of yes-no questions: Perfors, Tenenbaum, & Regier (2006, under review)
- Anaphoric *one*: Regier & Gahl (2004), Foraker et al. (2009), Pearl & Lidz (2009), Pearl & Mis (in prep)

What the models say about different phenomena

Structure-dependence of yes/no questions:

– Perfors, Tenenbaum & Regier (2006, under review):

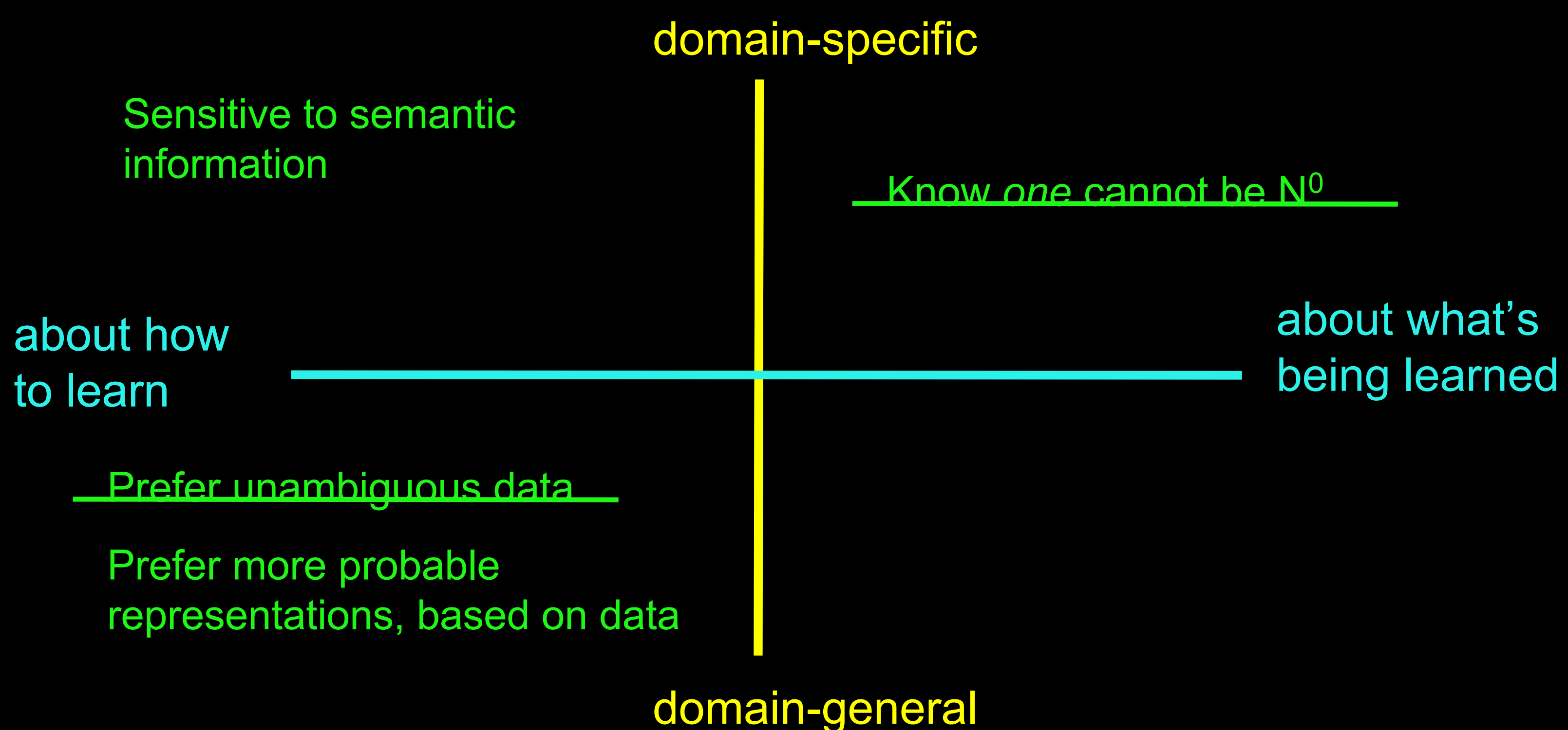
The learner needs to allow structure-dependent rules in the hypothesis space, but doesn't need to explicitly disallow structure-independent rules as long as the learner prefers simpler, more compact rules.



What the models say about different phenomena

Anaphoric *one*:

- Regier & Gahl (2004): Bayesian learners don't need to know *one* cannot be N^0 as long as they are sensitive to semantic information in the data and learn from other informative data besides unambiguous data.



What the models say about different phenomena

Anaphoric *one*:

- Foraker et al. (2009): Bayesian learners don't need to know *one* cannot be N^0 as long as they are only trying to learn the syntactic category of *one* and are sensitive to subtle semantic distinctions indicated by different syntactic categories.

domain-specific

Focus on syntactic category

~~Know *one* cannot be N^0~~

Sensitive to subtle semantic distinctions

about how to learn

about what's being learned

~~Prefer unambiguous data~~

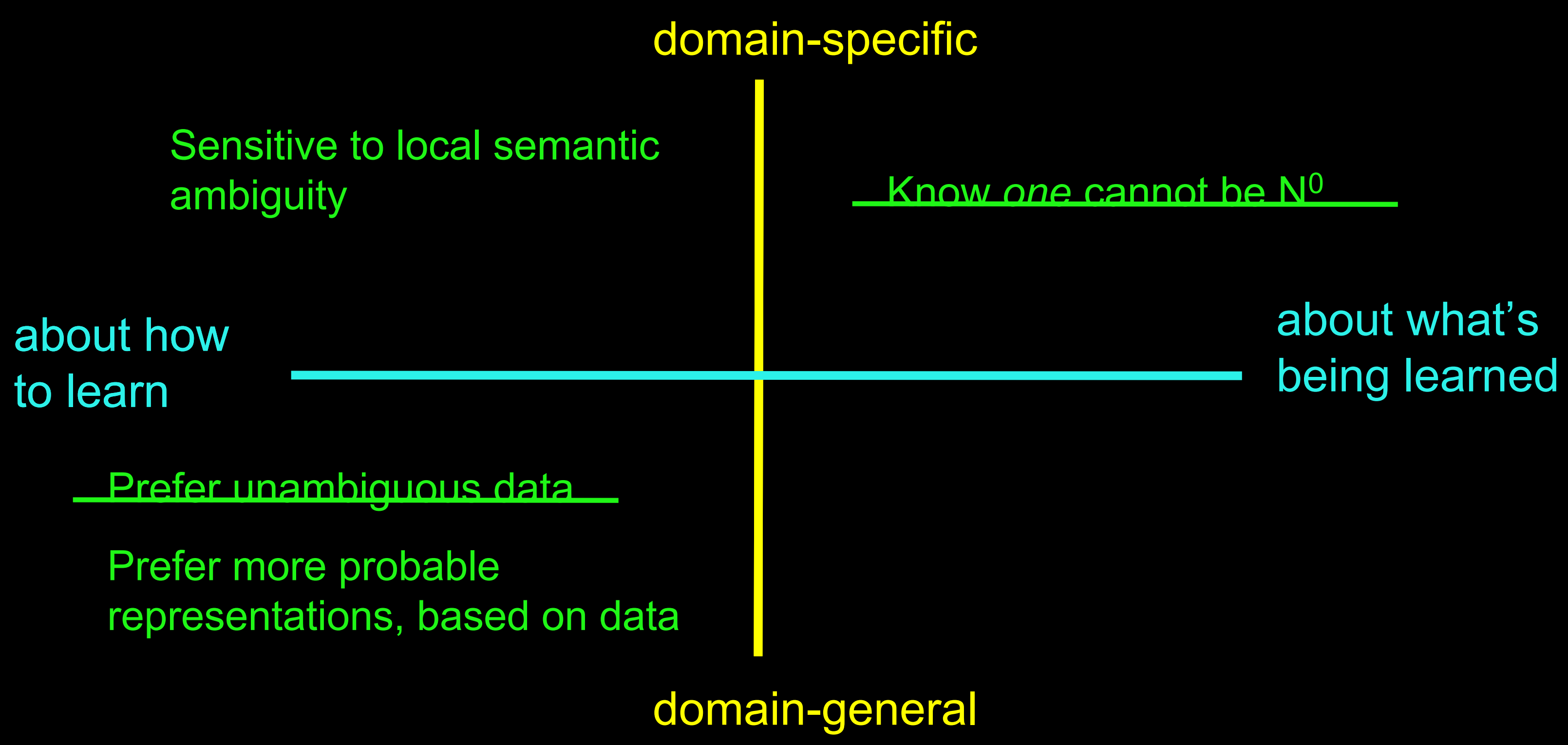
Prefer more probable representations, based on data

domain-general

What the models say about different phenomena

Anaphoric *one*:

- Pearl & Lidz (2009): Bayesian learners don't need to know *one* cannot be N^0 as long as they learn from a subset of the informative data that are either unambiguous or semantically ambiguous.



What the models say about different phenomena

Anaphoric *one*:

- Pearl & Mis (in prep): Bayesian learners don't need to know *one* cannot be N^0 as long as they recognize a larger set of data as informative, due to the fact that *one* is a referential element and other items in the language are too (ex: pronouns like *it*).

domain-specific

~~Know *one* cannot be N^0~~

One is one of many
referential elements

about how
to learn

about what's
being learned

~~Prefer unambiguous data~~

Prefer more probable
representations, based on data

domain-general

Big Picture Summary

- Poverty of the Stimulus: fundamental argument for Universal Grammar, which is often characterized as innate domain-specific knowledge required for children to learn language.

- Some examples of potential poverty of the stimulus situations
  - Structure dependence of complex yes/no questions
  - Anaphoric *one*

- Using computational modeling to determine the nature of the knowledge required when there is a poverty of the stimulus situation
    - Domain-specific or domain-general?
    - Knowledge about linguistic structure or knowledge about how to learn?

Computation of Language Laboratory

UC Irvine

Thank you!

Lisa S. Pearl
Assistant Professor
Department of Cognitive Sciences
SBSG 2314
University of California
Irvine, CA 92697-5100

lpearl@uci.edu

Extra Material

Knowledge of Language & Hidden Rules

Some examples from language:

You know that…

…these two statements mean fairly different things:

"Not even ten years ago you could see *Labyrinth* in theaters."

Could you see Labyrinth in theaters within the last ten years?

"Not even ten years ago could you see *Labyrinth* in theaters."

Could you see Labyrinth in theaters ten years ago?

Knowledge of Language & Hidden Rules

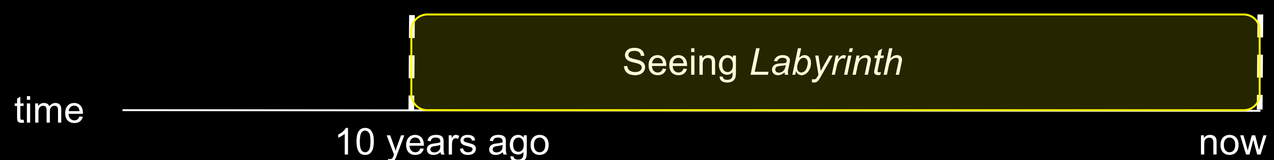
Some examples from language:

You know that…

…these two statements mean fairly different things:

"Not even ten years ago you could see *Labyrinth* in theaters."

Could you see Labyrinth in theaters within the last ten years?



"Not even ten years ago could you see *Labyrinth* in theaters."

Could you see Labyrinth in theaters ten years ago?

Knowledge of Language & Hidden Rules

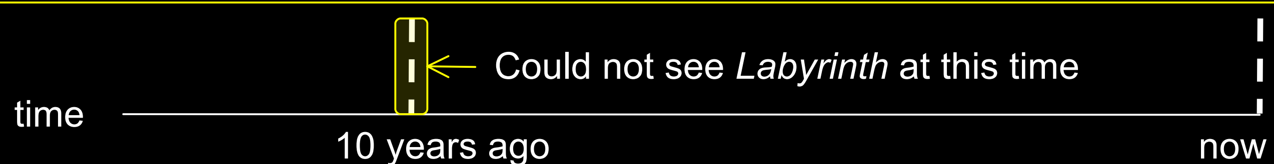
Some examples from language:

You know that…

…these two statements mean fairly different things:

"Not even ten years ago you could see *Labyrinth* in theaters."

Could you see Labyrinth in theaters within the last ten years?



"Not even ten years ago could you see *Labyrinth* in theaters."

Could you see Labyrinth in theaters ten years ago?

Other Poverty of the Stimulus Examples

Syntactic Islands

Syntactic Islands

Lily thought the letter (from the soldier) inspired the students.

Syntactic Islands

Lily thought the letter (from the soldier) inspired the students.

Questions formed from this statement:

Human ratings: Sprouse, Wagers, & Phillips (submitted 2010)

Ask about Lily (simple subject extraction: main clause)

Who __ thought [the letter inspired the students]?

Who __ thought [the letter from the soldier inspired the students]?

Syntactic Islands

Lily thought the letter (from the soldier) inspired the students.

Questions formed from this statement:

Human ratings: Sprouse, Wagers, & Phillips (submitted 2010)

Ask about Lily (simple subject extraction: main clause)

Who __ thought [the letter inspired the students]?

Who __ thought [the letter from the soldier inspired the students]?

Ask about the letter (simple subject extraction: embedded clause)

What did Lily think [__ inspired the students]?

Syntactic Islands

Lily thought the letter (from the soldier) inspired the students.

Questions formed from this statement:

Human ratings: Sprouse, Wagers, & Phillips (submitted 2010)

Ask about Lily (simple subject extraction: main clause)

Who __ thought [the letter inspired the students]?

Who __ thought [the letter from the soldier inspired the students]?

Ask about the letter (simple subject extraction: embedded clause)

What did Lily think [__ inspired the students]?

Ask about the soldier (complex subject extraction: embedded clause)

* Who did Lily think [the letter from __] inspired the students?

Syntactic Islands

Lily thought the letter (from the soldier) inspired the students.

Questions formed from this statement:

Human ratings: Sprouse, Wagers, & Phillips (submitted 2010)

Ask about Lily (simple subject extraction: main clause)

Who __ thought [the letter inspired the students]?

Who __ thought [the letter from the soldier inspired the students]?

Ask about the letter (simple subject extraction: embedded clause)

What did Lily think [__ inspired the students]?

Ask about the soldier (complex subject extraction: embedded clause)

* Who did Lily think [the letter from __] inspired the students?

Why can't we ask the last question? Conventional answer: You're trying to extract the question word from a "subject island". This is bad.

Syntactic Islands: The Problem

How do we learn to have these intuitions? Maybe we just like things we hear a lot and don't like things we never hear.

Questions formed from this statement:

Human ratings: Sprouse, Wagers, & Phillips (submitted 2010)

Ask about Lily (simple subject extraction: main clause)

Who __ thought [the letter inspired the students]?

Who __ thought [the letter from the soldier inspired the students]?

Ask about the letter (simple subject extraction: embedded clause)

What did Lily think [__ inspired the students]?

Ask about the soldier (complex subject extraction: embedded clause)

* Who did Lily think [the letter from __] inspired the students?

Syntactic Islands: The Problem

Let's look at how often these kind of questions appear in child-directed speech data.

Questions formed from this statement:

Pearl & Sprouse (in prep.) – from 12,000 wh-utterances of child-directed speech

2 Ask about Lily (simple subject extraction: main clause)
Who __ thought [the letter inspired the students]?

0 Who __ thought [the letter from the soldier inspired the students]?

13 Ask about the letter (simple subject extraction: embedded clause)
What did Lily think [__ inspired the students]?

0 Ask about the soldier (complex subject extraction: embedded clause)
* Who did Lily think [the letter from __] inspired the students?

Syntactic Islands: The Problem

Even the questions that sound perfectly grammatical are very rare. More importantly, people have different intuitions about the questions that never appear – one question sounds pretty good while the other sounds terrible.

Questions formed from this statement:

Pearl & Sprouse (in prep.) – from 12,000 wh-utterances of child-directed speech

2	Ask about Lily (simple subject extraction: main clause) Who __ thought [the letter inspired the students]?
0	Who __ thought [the letter from the soldier inspired the students]?
13	Ask about the letter (simple subject extraction: embedded clause) What did Lily think [__ inspired the students]?
0	Ask about the soldier (complex subject extraction: embedded clause) * Who did Lily think [the letter from __] inspired the students?

Syntactic Islands: The Problem

How do children develop the correct intuitions about these questions if they never hear them? How do they know to make generalizations about the language that include one question but not the other?

Questions formed from this statement:

Pearl & Sprouse (in prep.) – from 12,000 wh-utterances of child-directed speech

2 Ask about Lily (simple subject extraction: main clause)

Who __ thought [the letter inspired the students]?

0 Who __ thought [the letter from the soldier inspired the students]?

13 Ask about the letter (simple subject extraction: embedded clause)

What did Lily think [__ inspired the students]?

0 Ask about the soldier (complex subject extraction: embedded clause)

* Who did Lily think [the letter from __] inspired the students?