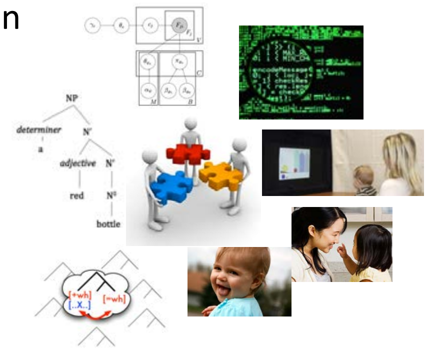
The Computation of Language: Syntactic Acquisition Edition

Computation of Language Laboratory
UC Irvine

Lisa S. Pearl
Associate Professor
Department of Cognitive Sciences
SBSG 2314
University of California, Irvine
lpearl@uci.edu



Feb 10, 2016
Department of Linguistics
UCLA

The Computation of Language: Information processing

One way to think about the computation of language is from an information processing standpoint.

The Computation of Language: Information processing

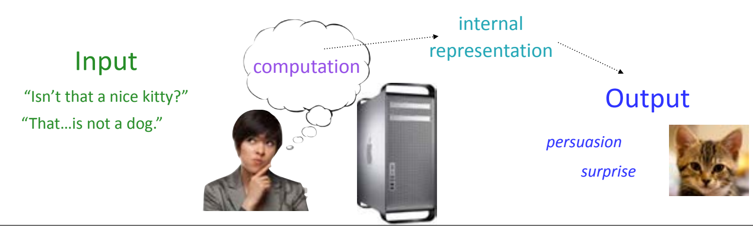
One way to think about the computation of language is from an information processing standpoint.

Natural language processing:
How do people and machines extract information about the world from the language data they encounter?

The Computation of Language: Information processing

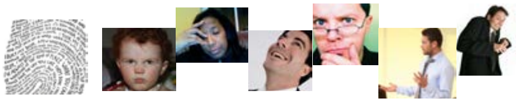
One way to think about the computation of language is from an information processing standpoint.

Natural language processing:
How do people and machines extract information about the world from the language data they encounter?



The Computation of Language: Information processing

One way to think about the computation of language is from an information processing standpoint.



Natural language processing:

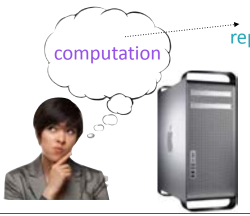
Recent work on *mindprints* and *writeprints*:

Linguistic feature-based "fingerprints" in text indicating mental states and identity.

Input

"Isn't that a nice kitty?"

"That…is not a dog."



internal representation

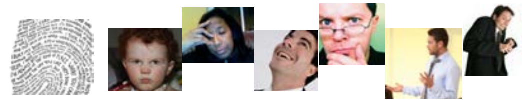
Output

persuasion

surprise

The Computation of Language: Information processing

One way to think about the computation of language is from an information processing standpoint.



Natural language processing:

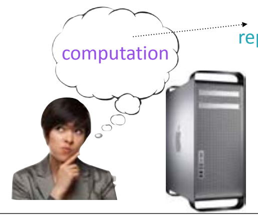
One finding: While shallow linguistic features can mimic human performance at detecting some mental states, more sophisticated syntactic and semantic features in mindprints can allow classifiers to exceed human performance in some cases

(Pearl & Steyvers 2010, 2013, Pearl & Enverga 2015, NIAAA, UCI, EU)

Input

"Isn't that a nice kitty?"

"That…is not a dog."



internal representation

Output

persuasion

surprise

The Computation of Language: Information processing

One way to think about the computation of language is from an information processing standpoint.

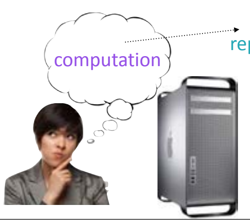
Natural language processing:

Another finding: We can use linguistically-sophisticated writeprints to identify who wrote a particular document *(Pearl & Steyvers 2012)*, and even which character written by the same author is currently being voiced in the text *(Pearl, Lu, & Haghighi in press)* — though the writeprint features that matter are different between authors vs. between characters by the same author.

Input

"Isn't that a nice kitty?"

"That…is not a dog."

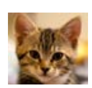


internal representation

Output

persuasion

surprise

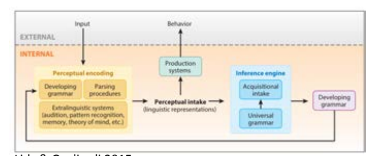


The Computation of Language: Information processing

One way to think about the computation of language is from an information processing standpoint.

Language acquisition:

How do children extract information about language from the language data they encounter?



Lidz & Gagliardi 2015

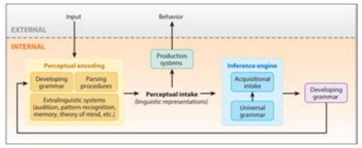
Sophisticated framework that makes explicit the different components of the acquisition process.

The Computation of Language: Information processing

One way to think about the computation of language is from an information processing standpoint.

Language acquisition:

How do children extract information about language from the language data they encounter?

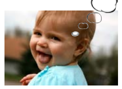


Lidz & Gagliardi 2015

Input

lʊkətðəkɪri
"What's that?"
"Do you see it?"

computation → internal representation → Output



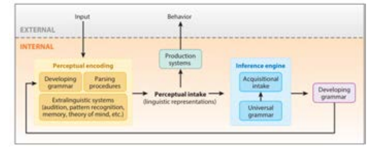
{look, at, the, kitty}
"Where's the kitty?"

The Computation of Language: Information processing

One way to think about the computation of language is from an information processing standpoint.

Language acquisition:

How do children extract information about language from the language data they encounter?

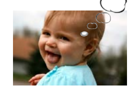


Lidz & Gagliardi 2015

Input

lʊkətðəkɪri
"What's that?"
"Do you see it?"

computation → internal representation → Output



{look, at, the, kitty}
"Where's the kitty?"

Language acquisition: Methods of investigation

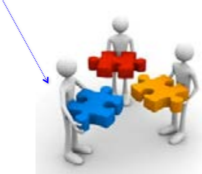
Language acquisition: Methods of investigation

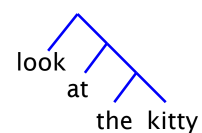
Theoretical methods:

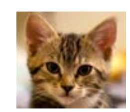
What knowledge of language is (and what children have to learn)

LOOK at the KItty

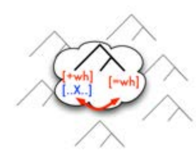
lʊkætðəkɪri





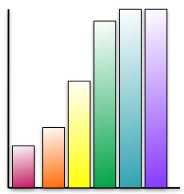


$\begin{bmatrix}+\text{stop}\\+\text{consonant}\\+\text{alveolar}\end{bmatrix} \rightarrow [\text{ɾ}] \;/\; \begin{bmatrix}+\text{vowel}\\+\text{stressed}\end{bmatrix} \_\_ \begin{bmatrix}+\text{vowel}\\-\text{stressed}\end{bmatrix}$



Language acquisition: Methods of investigation

Experimental methods:
When knowledge is acquired, what the **input** looks like, & plausible capabilities underlying **how** acquisition works



p(H1 | ) ∝ p( | H1) p(H1)

Language acquisition: Methods of investigation

Computational methods:
Strategies for **how** children acquire knowledge, sophisticated **quantitative analysis** of children's input & output

lʊkətðəkɪri

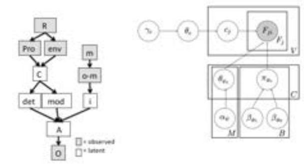
lʊkət ðəkɪri

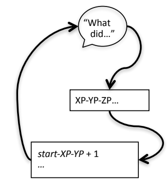
lʊ kət ðəkɪri

lʊkət ðə kɪri

lʊ kətðə kɪri

lʊk ət ðə kɪri

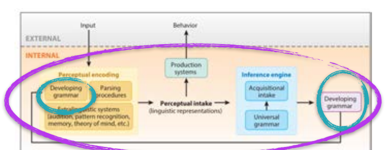




Language acquisition: Representation & Development

Language acquisition involves complex knowledge that builds on itself over the course of linguistic development, embedded in a developing cognitive system.

This means there's a natural dependence between theories of knowledge representation and theories of knowledge development.



Lidz & Gagliardi 2015

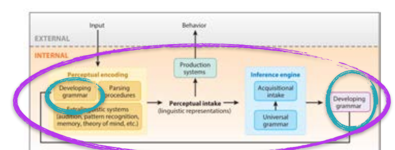
Language acquisition: Foundational knowledge

Language acquisition involves complex knowledge that builds on itself over the course of linguistic development, embedded in a developing cognitive system.

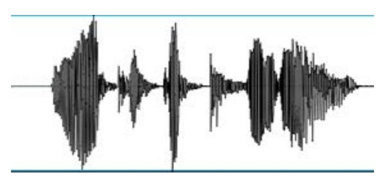
Examples of "foundational" processes that children use for building more sophisticated knowledge:

speech segmentation

syntactic categorization



Lidz & Gagliardi 2015



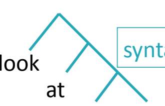
phonology

LOOK at the KItty

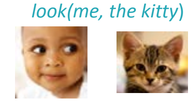
Nouns = Xx

syntax

look at the kitty



look(me, the kitty)



semantics

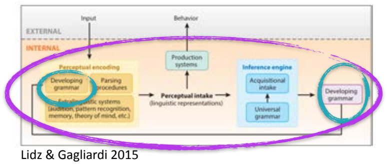
Language acquisition: Foundational knowledge

Language acquisition involves complex knowledge that builds on itself over the course of linguistic development, embedded in a developing cognitive system.

Examples of "foundational" processes that children use for building more sophisticated knowledge:

speech segmentation

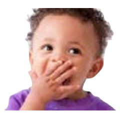
syntactic categorization



Lidz & Gagliardi 2015

A recent finding: When the underlying representation (i.e., assumptions about language structure) is immature, immature processing capabilities may be helpful rather than harmful

speech segmentation: Pearl, Goldwater and Steyvers 2010, 2011, Phillips and Pearl 2012, 2015b



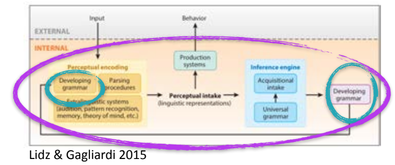
Language acquisition: Foundational knowledge

Language acquisition involves complex knowledge that builds on itself over the course of linguistic development, embedded in a developing cognitive system.

Examples of "foundational" processes that children use for building more sophisticated knowledge:

speech segmentation

syntactic categorization



Lidz & Gagliardi 2015

A recent finding: Developing representations are often "good enough" for scaffolding other acquisition processing even when they don't match adult representations (*Pearl 2014, Pearl & Sprouse 2015, Pearl under review*)

speech segmentation: Phillips and Pearl 2012, 2014a,b, 2015a,b, Pearl and Phillips under review, Phillips and Pearl under revision

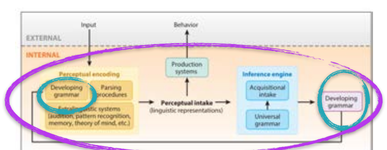
syntactic categorization: Bar-Sever and Pearl 2016

Language acquisition: More sophisticated knowledge

Language acquisition involves complex knowledge that builds on itself over the course of linguistic development, embedded in a developing cognitive system.

Examples of more sophisticated knowledge that depends on the foundational knowledge:

metrical stress



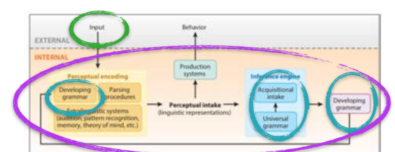
Lidz & Gagliardi 2015

Language acquisition: More sophisticated knowledge

Language acquisition involves complex knowledge that builds on itself over the course of linguistic development, embedded in a developing cognitive system.

Examples of more sophisticated knowledge that depends on the foundational knowledge:

metrical stress



Lidz & Gagliardi 2015

A current finding: Some linguistic representations may be less acquirable from cognitively plausible child-directed input than previously assumed unless certain learning biases are in place

Pearl 2007, 2008, 2009, 2011, Pearl, Ho, & Detrano 2014, under review, Pearl under review

Language acquisition: More sophisticated knowledge

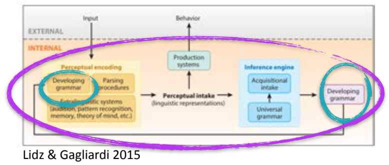
Language acquisition involves complex knowledge that builds on itself over the course of linguistic development, embedded in a developing cognitive system.

Examples of more sophisticated knowledge that depends on the foundational knowledge:

syntactic islands

English anaphoric *one*

where arguments appear syntactically



Lidz & Gagliardi 2015

Language acquisition: More sophisticated knowledge

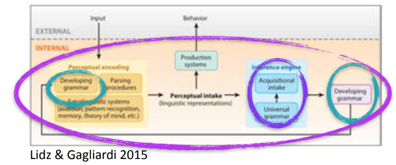
Language acquisition involves complex knowledge that builds on itself over the course of linguistic development, embedded in a developing cognitive system.

Examples of more sophisticated knowledge that depends on the foundational knowledge:

syntactic islands

English anaphoric *one*

where arguments appear syntactically



Lidz & Gagliardi 2015

A current finding: The knowledge needed to create the right acquisitional intake may not necessarily look like we thought it did (e.g., what's in Universal Grammar).

syntactic islands: Pearl & Sprouse 2013a, 2013b, Pearl 2014, Pearl & Sprouse 2015, Pearl under rev.

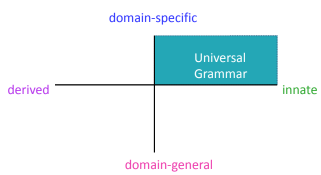
English anaphoric one: Pearl 2007, Pearl & Lidz 2009, Pearl & Mis 2011, Pearl 2014, Pearl & Mis in press

where arguments appear: Pearl & Sprouse in progress

NSF: "Testing the Universal Grammar Hypothesis", "An Integrated Theory of Syntactic Acquisition"

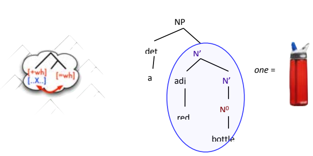
Today's Plan

Investigating Universal Grammar (UG)



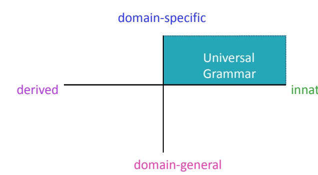
Characterizing learning problems precisely enough to informatively model them

UG modeling forays



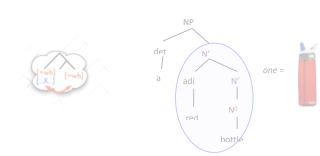
Today's Plan

Investigating Universal Grammar (UG)



Characterizing learning problems precisely enough to informatively model them

UG modeling forays



Motivating Universal Grammar

The argument from acquisition: one explicit motivation that highlights the natural link between linguistic representation and language acquisition.

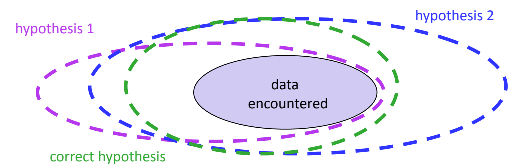
Universal Grammar (UG) allows children to acquire knowledge about language as effectively and rapidly as they do (Chomsky 1980, Crain 1991, Hornstein & Lightfoot 1981, Lightfoot 1982b, Legate & Yang 2002, among many others).

Motivating Universal Grammar

What's so hard about acquiring language?

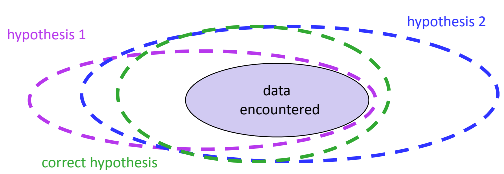
There seem to be induction problems, given the available data.

(Poverty of the Stimulus, Logical Problem of Language Acquisition, Plato's Problem)



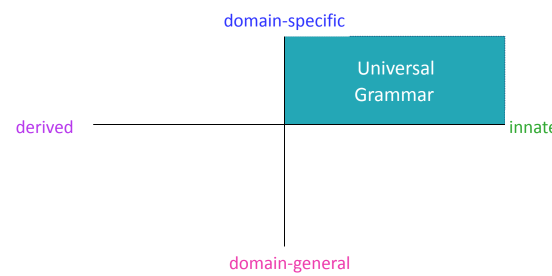
Motivating Universal Grammar

So if the data themselves don't pick out the right answer (and children all seem to), something internal to children must be guiding them.



Motivating Universal Grammar

If that something is both innate and domain-specific, we consider it part of Universal Grammar (UG) (Chomsky 1965, Chomsky 1975, Pearl & Sprouse 2013).



Motivating the contents of UG

Proposals have traditionally come from characterizing a specific acquisition problem for a particular linguistic phenomenon, and describing the (UG) solution to that specific characterization.

Motivating the contents of UG

Proposals have traditionally come from characterizing a specific acquisition problem for a particular linguistic phenomenon, and describing the (UG) solution to that specific characterization.

Structure-dependent rules

(Chomsky 1980, Anderson & Lightfoot 2000; Fodor & Crowther 2002; Berwick et al. 2011; Anderson 2013)

Pirates who can dance can often fight well.

Can pirates who can dance __ often fight well?

Motivating the contents of UG

Proposals have traditionally come from characterizing a specific acquisition problem for a particular linguistic phenomenon, and describing the (UG) solution to that specific characterization.

Syntactic islands: Constraints on long-distance dependencies

(Chomsky 1973, Huang 1982, Lasnik & Saito 1984, Pearl & Sprouse 2013a, 2013b, 2015)

Where did Jack think Lily bought the necklace from __?

*Where did Jack think the necklace from __ was too expensive?

Motivating the contents of UG

Proposals have traditionally come from characterizing a specific acquisition problem for a particular linguistic phenomenon, and describing the (UG) solution to that specific characterization.

English anaphoric *one* representation

(Baker 1978, Pearl & Mis 2011, 2016)

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

UG proposals: Generation & evaluation

How to generate a learning theory proposal:

Characterize the learning problem precisely and identify a potential solution.

UG proposals: Generation & evaluation

How to generate a learning theory proposal:

Characterize the learning problem precisely and identify a potential solution.

Benefit of computational modeling:
We can make sure the learning problem is characterized precisely enough to implement. It's not always obvious what pieces are missing until you try to build a model of the learning process.
(Pearl 2014, Pearl & Sprouse 2015)

UG proposals: Generation & evaluation

How to generate a learning theory proposal:

Characterize the learning problem precisely and identify a potential solution.

How to evaluate a learning theory proposal:

See if it's successful when embedded in a model of the acquisition process for that learning problem.

UG proposals: Generation & evaluation

How to generate a learning theory proposal:

Characterize the learning problem precisely and identify a potential solution.

How to evaluate a learning theory proposal:

See if it's successful when embedded in a model of the acquisition process for that learning problem.

Recently, in computational modeling, we've seen the integration of rich hypothesis spaces with probabilistic/statistical learning mechanisms (Sakas & Fodor 2001, Yang 2004, Pearl 2011, Dillon et al. 2013, Pearl & Sprouse 2013, Pearl et al. 2014, Pearl & Mis 2016, among many others).

UG proposals: Generation & evaluation

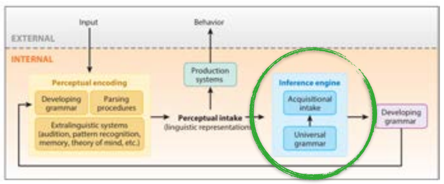
How to generate a learning theory proposal:

Characterize the learning problem precisely and identify a potential solution.

How to evaluate a learning theory proposal:

See if it's successful when embedded in a model of the acquisition process for that learning problem.

We've also seen the development of more sophisticated acquisition frameworks that highlight the precise role of UG (Lidz & Gagliardi 2015).



Example: UG determines what data from the perceived input are relevant (acquisitional intake)

UG proposals: Generation & evaluation

How to generate a learning theory proposal:

Characterize the learning problem precisely and identify a potential solution.

How to evaluate a learning theory proposal:

See if it's successful when embedded in a model of the acquisition process for that learning problem.

This computational modeling feedback helps us refine our theories about both the knowledge representation the learning theory relies on and the acquisition process that uses that representation.

UG proposals: Generation & evaluation

How to generate a learning theory proposal:

Characterize the learning problem precisely and identify a potential solution.

How to evaluate a learning theory proposal:

See if it's successful when embedded in a model of the acquisition process for that learning problem.

How to decide if any components of the proposal are UG:

Examine the components of the successful learning solution.

UG proposals: Generation & evaluation

How to generate a learning theory proposal:

Characterize the learning problem precisely and identify a potential solution.

How to evaluate a learning theory proposal:

See if it's successful when embedded in a model of the acquisition process for that learning problem.

How to decide if any components of the proposal are UG:

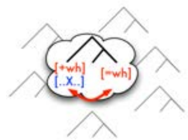
Examine the components of the successful learning solution.

Are they necessarily both domain-specific and innate?

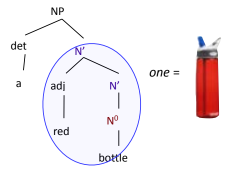
Note: We may use "innate" as a placeholder until we can determine if it's impossible to derive the relevant component (Pearl 2014, Pearl & Mis 2016).

UG proposal refinement: Recent successful forays

Syntactic islands (constraints on *wh*-dependencies):
Pearl & Sprouse 2013a, 2013b, 2015

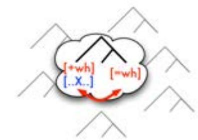


English anaphoric *one*:
Pearl & Mis 2011, 2016

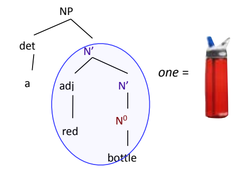


UG proposal refinement: Recent successful forays

Syntactic islands (constraints on *wh*-dependencies):
Pearl & Sprouse 2013a, 2013b, 2015

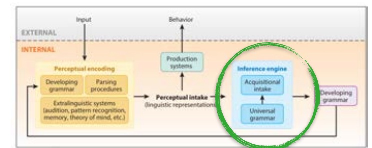


English anaphoric *one*:
Pearl & Mis 2011, 2016



Recurring themes:

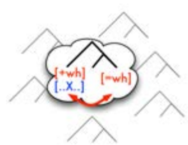
(1) Broadening the set of relevant data in the acquisitional intake



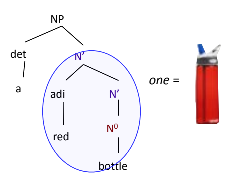
Lidz & Gagliardi 2015

UG proposal refinement: Recent successful forays

Syntactic islands (constraints on *wh*-dependencies):
Pearl & Sprouse 2013a, 2013b, 2015



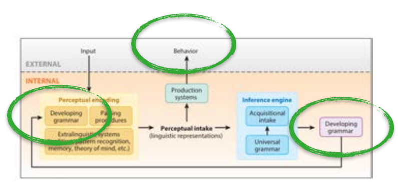
English anaphoric *one*:
Pearl & Mis 2011, 2016



Recurring themes:

(1) Broadening the set of relevant data in the acquisitional intake

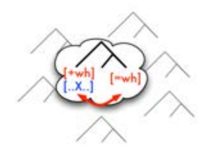
(2) Evaluating output by how useful it is



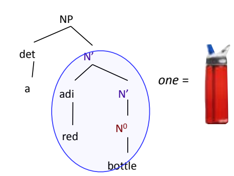
Lidz & Gagliardi 2015

UG proposal refinement: Recent successful forays

Syntactic islands (constraints on *wh*-dependencies):
Pearl & Sprouse 2013a, 2013b, 2015



English anaphoric *one*:
Pearl & Mis 2011, 2016

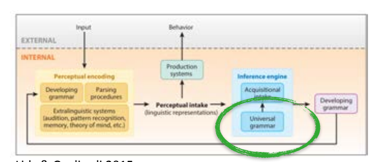


Recurring themes:

(1) Broadening the set of relevant data in the acquisitional intake

(2) Evaluating output by how useful it is

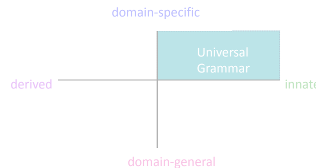
(3) Not necessarily needing the prior knowledge we thought we did



Lidz & Gagliardi 2015

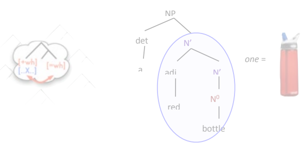
Today's Plan

Investigating Universal Grammar (UG)



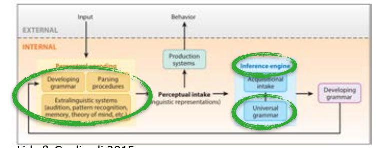
Characterizing learning problems precisely enough to informatively model them

UG modeling forays



Characterizing learning problems

Initial state:



Lidz & Gagliardi 2015

Pearl & Sprouse 2015, Pearl & Mis 2016

Characterizing learning problems

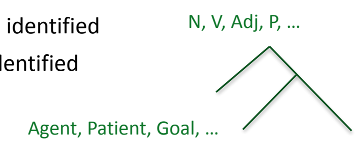
Initial state:

- initial knowledge state

ex: syntactic categories exist and can be identified

ex: phrase structure exists and can be identified

ex: participant roles can be identified



Pearl & Sprouse 2015, Pearl & Mis 2016

Characterizing learning problems

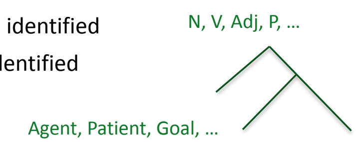
Initial state:

- initial knowledge state

ex: syntactic categories exist and can be identified

ex: phrase structure exists and can be identified

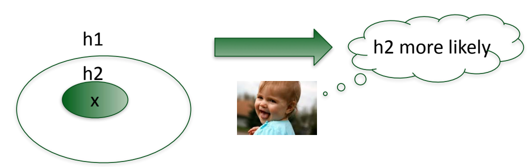
ex: participant roles can be identified



- learning biases & capabilities

ex: frequency information can be tracked

ex: distributional information can be leveraged

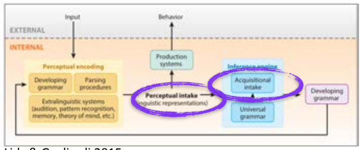


Pearl & Sprouse 2015, Pearl & Mis 2016

Characterizing learning problems

Initial state: initial knowledge state + learning biases & capabilities

Data intake:



Lidz & Gagliardi 2015

Pearl & Sprouse 2015, Pearl & Mis 2016

Characterizing learning problems

Initial state: initial knowledge state + learning biases & capabilities

Data intake:

- encoding + acquisitional intake = data perceived as relevant for learning

(Fodor 1998, Lidz & Gagliardi 2015)

ex: all *wh*-utterances for learning about *wh*-dependencies

ex: all pronoun data when learning about anaphoric *one*

ex: syntactic and conceptual data for learning syntactic knowledge that links with conceptual knowledge

[defined by knowledge & biases/capabilities in the initial state]

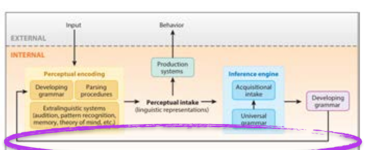
Pearl & Sprouse 2015, Pearl & Mis 2016

Characterizing learning problems

Initial state: initial knowledge state + learning biases & capabilities

Data intake: data perceived as relevant for learning

Learning period:



Lidz & Gagliardi 2015

Pearl & Sprouse 2015, Pearl & Mis 2016

Characterizing learning problems

Initial state: initial knowledge state + learning biases & capabilities

Data intake: data perceived as relevant for learning

Learning period:

- how long children have to reach the target knowledge state (when inference & iteration happen)

ex: 3 years, ~1,000,000 data points

ex: 4 months, ~36,500 data points

Pearl & Sprouse 2015, Pearl & Mis 2016

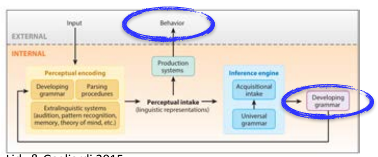
Characterizing learning problems

Initial state: initial knowledge state + learning biases & capabilities

Data intake: data perceived as relevant for learning

Learning period: how long children have to learn

Target state:



Lidz & Gagliardi 2015

Pearl & Sprouse 2015, Pearl & Mis 2016

Characterizing learning problems

Initial state: initial knowledge state + learning biases & capabilities

Data intake: data perceived as relevant for learning

Learning period: how long children have to learn

Target state:

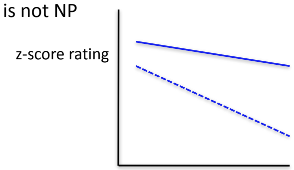
- the knowledge children are trying to attain (as indicated by their behavior)

ex: *Where did Jack think the necklace from __ was too expensive?

ex: *one* is category N' when it is not NP

ex:

looking time preferences



Expectations of argument roles

Pearl & Sprouse 2015, Pearl & Mis 2016

Characterizing learning problems

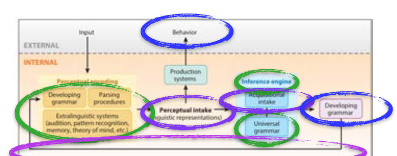
Initial state: initial knowledge state + learning biases & capabilities

Data intake: data perceived as relevant for learning

Learning period: how long children have to learn

Target state: the knowledge children must attain

Once we have all these pieces specified, we should be able to implement an informative model of the learning process.



Lidz & Gagliardi 2015

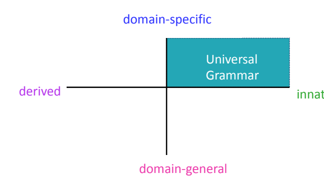
Pearl & Sprouse 2015, Pearl & Mis 2016

Informing UG (+ acquisition theory)

When we identify a successful learning strategy via modeling, this is an existence proof that children could solve that learning problem using the learning biases, knowledge, and capabilities comprising that strategy.

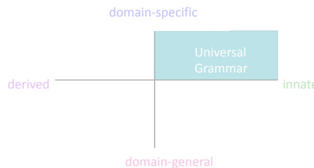
This identifies useful learning strategy components, which we can then examine to see where they might come from.

Knowledge 1
Knowledge 2
Capability 1
Bias 1
Bias 2
Bias 3
...



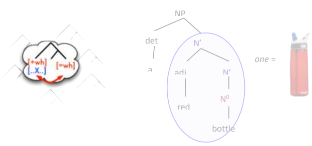
Today's Plan

Investigating Universal Grammar (UG)

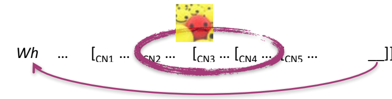


Characterizing learning problems precisely enough to informatively model them

UG modeling forays



Syntactic islands



- **Why**? Central to UG-based syntactic theories.

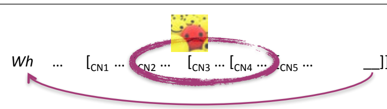
- **What**? Dependencies can exist between two non-adjacent items. They do not appear to be constrained by length (Chomsky 1965, Ross 1967), but rather by whether the dependency crosses certain structures (called "syntactic islands").

What does Jack think __?

What does Jack think that Lily said that Sarah heard that Jareth believed __?

Pearl & Sprouse 2013a, 2013b, 2015

Syntactic islands



- **Why**? Central to UG-based syntactic theories.

- **What**? Dependencies can exist between two non-adjacent items. They do not appear to be constrained by length (Chomsky 1965, Ross 1967), but rather by whether the dependency crosses certain structures (called "syntactic islands").

<u>Some example islands</u>

Complex NP island:
*What did you make [the claim that Jack bought __]?

Subject island:
*What do you think [the joke about __] offended Jack?

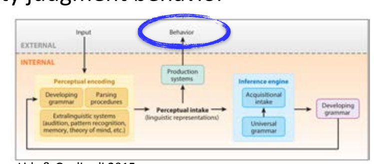
Whether island:
*What do you wonder [whether Jack bought __]?

Adjunct island:
*What do you worry [if Jack buys __]?

Pearl & Sprouse 2013a, 2013b, 2015

Syntactic islands: Acquisition target

Adult knowledge as measured by acceptability judgment behavior



Lidz & Gagliardi 2015

What does Jack think __?

What does Jack think that Lily said that Sarah heard that Jareth believed __?

Complex NP island:
*What did you make [the claim that Jack bought __]?

Subject island:
*What do you think [the joke about __] offended Jack?

Whether island:
*What do you wonder [whether Jack bought __]?

Adjunct island:
*What do you worry [if Jack buys __]?

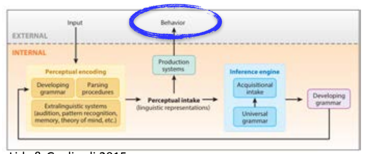
Pearl & Sprouse 2013a, 2013b, 2015

Syntactic islands: Acquisition target

Adult knowledge as measured by acceptability judgment behavior

Sprouse et al. (2012) collected magnitude estimation judgments for four different islands, using a factorial definition that controlled for two salient properties of island-crossing dependencies:

- length of dependency
(matrix vs. embedded)
- presence of an island structure
(non-island vs. island)



Lidz & Gagliardi 2015

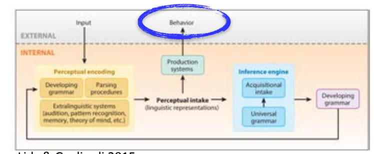
Pearl & Sprouse 2013a, 2013b, 2015

Syntactic islands: Acquisition target

Adult knowledge as measured by acceptability judgment behavior

Sprouse et al. (2012) collected magnitude estimation judgments for four different islands, using a factorial definition that controlled for two salient properties of island-crossing dependencies:

- length of dependency
(matrix vs. embedded)
- presence of an island structure
(non-island vs. island)



Lidz & Gagliardi 2015

Complex NP islands

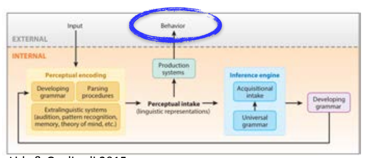
Who __ claimed that Lily forgot the necklace?	matrix | non-island
What did the teacher claim that Lily forgot __?	embedded | non-island
Who __ made the claim that Lily forgot the necklace?	matrix | island
*What did the teacher make the claim that Lily forgot __?	embedded | island

Pearl & Sprouse 2013a, 2013b, 2015

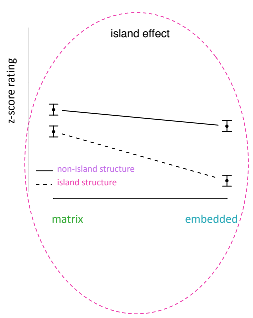
Syntactic islands: Acquisition target

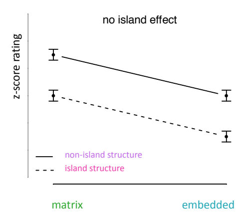
Adult knowledge as measured by acceptability judgment behavior

Syntactic island = superadditive interaction of the two factors (additional unacceptability that arises when the two factors are combined, above and beyond the independent contribution of each factor).



Lidz & Gagliardi 2015

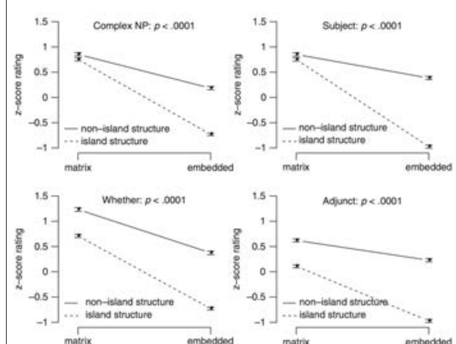


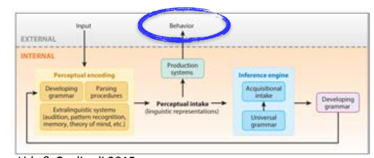


Pearl & Sprouse 2013a, 2013b, 2015

Syntactic islands: Acquisition target

Sprouse et al. (2012): acceptability judgments from 173 adult subjects





Lidz & Gagliardi 2015

Superadditivity present for all islands tested = Knowledge that dependencies cannot cross these island structures is part of adult knowledge about syntactic islands

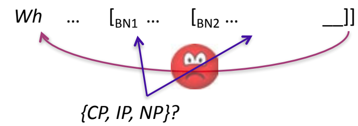
Importance for acquisition: This is one kind of target behavior that we'd like a learner to produce.

Pearl & Sprouse 2013a, 2013b, 2015

Syntactic islands: Representations

Subjacency (Chomsky 1973, Huang 1982, Lasnik & Saito 1984)

(1) A dependency cannot cross two or more bounding nodes.



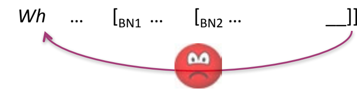
Bounding nodes are language-specific

(CP, IP, and/or NP – must learn which ones are relevant for language)

Syntactic islands: Representations

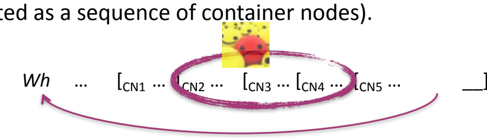
Subjacency (Chomsky 1973, Huang 1982, Lasnik & Saito 1984)

(1) A dependency cannot cross two or more bounding nodes.



Subjacency-ish (Pearl & Sprouse 2013a, 2013b, 2015)

(2) A dependency cannot cross a very low probability region of structure (represented as a sequence of container nodes).



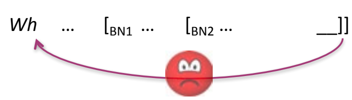
Container node: phrase structure node that contains dependency

[$_{CP}$ *What* *do* [$_{IP}$ *you* [$_{VP}$ *like* __ [$_{PP}$ *in this picture?*]]]]

Syntactic islands: Representations

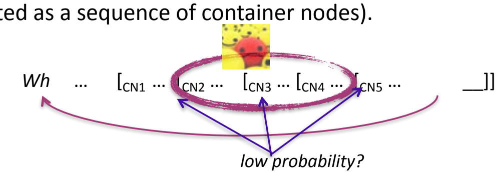
Subjacency (Chomsky 1973, Huang 1982, Lasnik & Saito 1984)

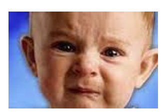
(1) A dependency cannot cross two or more bounding nodes.



Subjacency-ish (Pearl & Sprouse 2013a, 2013b, 2015)

(2) A dependency cannot cross a very low probability region of structure (represented as a sequence of container nodes).





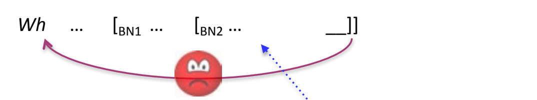
Low probability regions are language-specific

(defined by sequences of container nodes that must be learned)

Syntactic islands: Representations

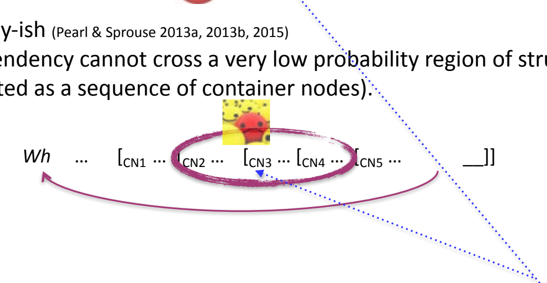
Subjacency (Chomsky 1973, Huang 1982, Lasnik & Saito 1984)

(1) A dependency cannot cross two or more bounding nodes.



Subjacency-ish (Pearl & Sprouse 2013a, 2013b, 2015)

(2) A dependency cannot cross a very low probability region of structure (represented as a sequence of container nodes).



In common: Both rely on local structure anomalies (at some level)

Syntactic islands: Representations

Subjacency (Chomsky 1973, Huang 1982, Lasnik & Saito 1984)

(1) A dependency cannot cross two or more bounding nodes.

Wh … [$_{BN1}$ … [$_{BN2}$ … __]]

(i) Dependencies defined over bounding nodes — track those
(ii) Bounding node = ?
(iii) 2+ bounding nodes =

Subjacency-ish (Pearl & Sprouse 2013a, 2013b, 2015)

(2) A dependency cannot cross a very low probability region of structure (represented as a sequence of container nodes).

Wh … [$_{CN1}$ … [$_{CN2}$ … [$_{CN3}$ … [$_{CN4}$ … [$_{CN5}$ … __]]

(i) Dependencies defined over container node structure — track that already
(ii) Container node = ?
(iii) low probability =

Different: Amount of language-specific knowledge built in just for islands

Pearl & Sprouse 2013a, 2013b, 2015

Syntactic islands: Representations

Subjacency (Chomsky 1973, Huang 1982, Lasnik & Saito 1984)

(1) A dependency cannot cross two or more bounding nodes.

Wh … [$_{BN1}$ … [$_{BN2}$ … __]]

Subjacency-ish (Pearl & Sprouse 2013a, 2013b, 2015)

(2) A dependency cannot cross a very low probability region of structure (represented as a sequence of container nodes).

Wh … [$_{CN1}$ … [$_{CN2}$ … [$_{CN3}$ … [$_{CN4}$ … [$_{CN5}$ … __]]

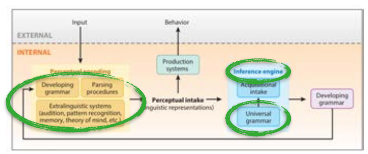
Pearl & Sprouse: Focused on evaluating this one

Pearl & Sprouse 2013a, 2013b, 2015

Syntactic islands: Subjacency-ish

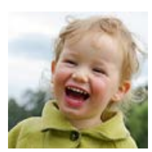
Subjacency-ish implementation:
A dependency cannot cross a very low probability region of structure (represented as a sequence of container nodes).

Wh … [$_{CN1}$ … [$_{CN2}$ … [$_{CN3}$ … [$_{CN4}$ … [$_{CN5}$ … __]]



Lidz & Gagliardi 2015

Initial state:
(i) Dependencies defined over container node structure
(ii) Container nodes recognized
(iii) Track probability of short container node sequences (trigrams)



Pearl & Sprouse 2013a, 2013b, 2015

Subjacency-ish: Initial state implementation

Because *wh*-dependencies are perceived as sequences of container nodes, local pieces of dependency structure can be characterized by container node trigrams.

[$_{CP}$ Who did [$_{IP}$ she [$_{VP}$ think [$_{CP}$ [$_{IP}$ [$_{NP}$ the gift] [$_{VP}$ was [$_{PP}$ from __]]]]]]]?

IP VP CP$_{null}$ IP VP PP

begin-IP-VP-CP$_{null}$-IP-VP-PP-*end* =

begin-IP-VP
IP-VP-CP$_{null}$
VP-CP$_{null}$-IP
CP$_{null}$-IP-VP
IP-VP-PP
VP-PP-*end*

Pearl & Sprouse 2013a, 2013b, 2015

Subjacency-ish: Developing knowledge

A child learns about the frequency of container node trigrams…

[$_{CP}$ Who did [$_{IP}$ she [$_{VP}$ think [$_{CP}$ [$_{IP}$ [$_{NP}$ the gift] [$_{VP}$ was [$_{PP}$ from __]]]]]]]]?

IP VP CP$_{null}$ IP VP PP

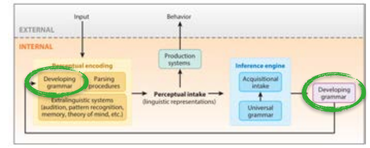
begin-IP-VP-CP$_{null}$-IP-VP-PP-*end* =

begin-IP-VP
IP-VP-CP$_{null}$
VP-CP$_{null}$-IP
CP$_{null}$-IP-VP
IP-VP-PP
VP-PP-*end*

begin-IP-VP ***+1***
IP-VP-CP$_{null}$ ***+1***
…

Pearl & Sprouse 2013a, 2013b, 2015

Subjacency-ish: Developing knowledge



Lidz & Gagliardi 2015

…and at the end of the learning period has a sense of the probability of any given container node trigram, based on its relative frequency.

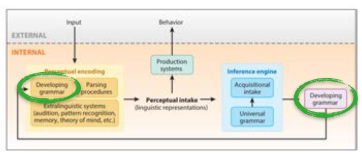
☺ *begin*-IP-VP
IP-VP-CP$_{null}$
begin-IP-*end*

☹ IP-VP-CP$_{if}$
IP-NP-PP

Pearl & Sprouse 2013a, 2013b, 2015

Subjacency-ish: Developing knowledge

Any *wh*-dependency can then have a probability, based on the product of the smoothed probabilities of its trigrams.



Lidz & Gagliardi 2015

Who did she think the gift was from __?

begin-IP-VP-CP$_{null}$-IP-VP-PP-*end* ≈ $\prod$ p(trigram)

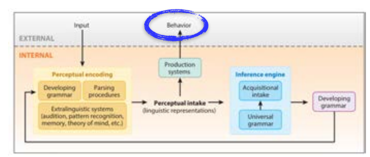
Probability(*begin*-IP-VP-CP$_{null}$-IP-VP-PP-*end*)
= p(*begin*-IP-VP)
p(IP-VP-CP$_{null}$)
p(VP-CP$_{null}$-IP)
p(CP$_{null}$-IP-VP)
p(IP-VP-PP)
p(VP-PP-*end*)

Pearl & Sprouse 2013a, 2013b, 2015

Subjacency-ish: Developing knowledge

This allows the modeled learner to generate judgments about the grammaticality of any dependency.

Higher probability dependencies are more grammatical while lower probability dependencies are less grammatical.



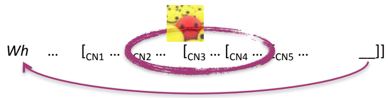
Lidz & Gagliardi 2015

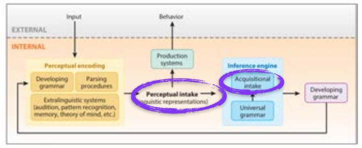
begin-IP-VP-CP$_{null}$-IP-VP-PP-*end* = ☺
begin-IP-VP-CP$_{if}$-IP-VP-*end* = ☹

Pearl & Sprouse 2013a, 2013b, 2015

Syntactic islands: Subjacency-ish

Subjacency-ish input & intake:
A dependency cannot cross a very low probability region of structure (represented as a sequence of container nodes).





Lidz & Gagliardi 2015

Data intake: defined by initial state =

wh-dependencies in child-directed speech, as characterized by container nodes

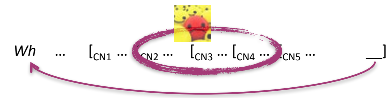
But which *wh*-dependencies? Just the ones being evaluated in the target state?

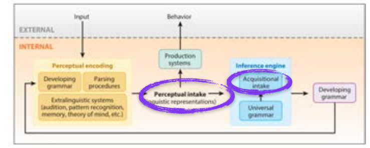
Who __ claimed that Lily forgot the necklace?	matrix | non-island
What did the teacher claim that Lily forgot __?	embedded | non-island
Who __ made the claim that Lily forgot the necklace?	matrix | island
*What did the teacher make the claim that Lily forgot __?	embedded | island

Pearl & Sprouse 2013a, 2013b, 2015

Syntactic islands: Subjacency-ish

Subjacency-ish input & intake:
A dependency cannot cross a very low probability region of structure (represented as a sequence of container nodes).



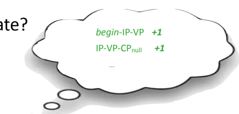


Lidz & Gagliardi 2015

Data intake: defined by initial state =

wh-dependencies in child-directed speech, as characterized by container nodes

But which *wh*-dependencies? Just the ones being evaluated in the target state?

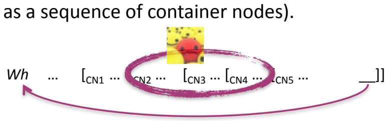


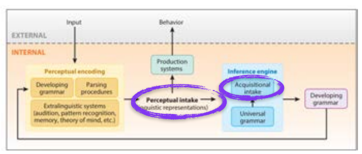
No! Any *wh*-dependency has relevant information about container node trigrams used to determine the grammaticality of *wh*-dependencies in general.

Pearl & Sprouse 2013a, 2013b, 2015

Syntactic islands: Subjacency-ish

Subjacency-ish input & intake:
A dependency cannot cross a very low probability region of structure (represented as a sequence of container nodes).





Lidz & Gagliardi 2015

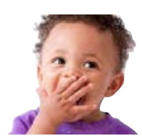
Data intake: defined by initial state =

all *wh*-dependencies in child-directed speech, as characterized by container nodes

(Brown-Adam, Brown-Eve, Suppes, Valian) from CHILDES:
101,838 utterances containing 20,923 *wh*-dependencies

76.7%	*What did you see __?*
12.8%	*What __ happened?*
5.6%	*What did she want to do __?*
2.5%	*What did she read from __?*
1.1%	*What did she think he said __?*

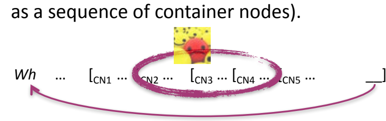
…

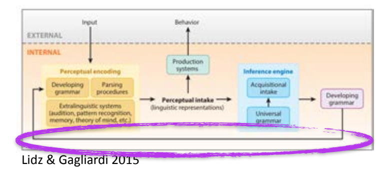


Pearl & Sprouse 2013a, 2013b, 2015

Syntactic islands: Subjacency-ish

Subjacency-ish input & intake:
A dependency cannot cross a very low probability region of structure (represented as a sequence of container nodes).





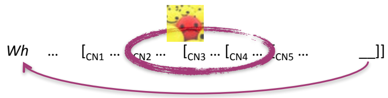
Lidz & Gagliardi 2015

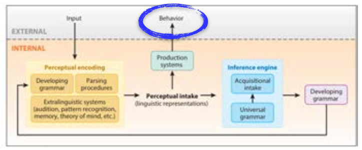
Learning period: defined by empirical estimates from Hart & Risley (1995) (~3 years of data)
=~ 200,000 *wh*-dependency data points

Pearl & Sprouse 2013a, 2013b, 2015

Syntactic islands: Subjacency-ish

Subjacency-ish input & intake:
A dependency cannot cross a very low probability region of structure (represented as a sequence of container nodes).



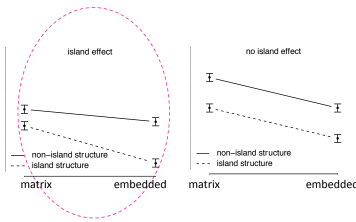


Lidz & Gagliardi 2015

Target state: Behavioral evidence of syntactic islands knowledge

Non-parallel lines indicate superadditivity, which indicates knowledge of islands.

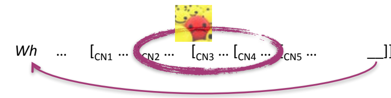
But how do we get acceptability judgment equivalents?

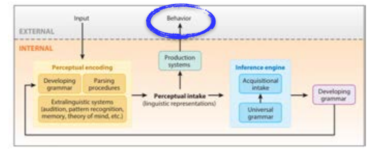


Pearl & Sprouse 2013a, 2013b, 2015

Syntactic islands: Subjacency-ish

Subjacency-ish input & intake:
A dependency cannot cross a very low probability region of structure (represented as a sequence of container nodes).

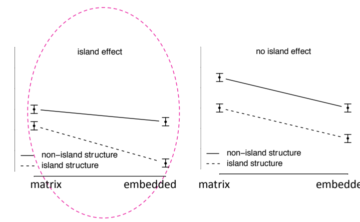




Lidz & Gagliardi 2015

Target state: Behavioral evidence of syntactic islands knowledge

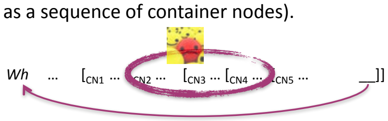
For each set of island stimuli from Sprouse et al. (2012), we generate grammaticality preferences for the modeled learner based on the dependency's perceived probability and use this as a stand-in for acceptability.

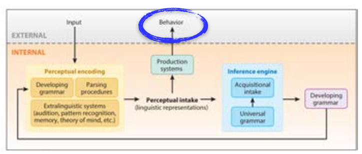


Pearl & Sprouse 2013a, 2013b, 2015

Syntactic islands: Subjacency-ish

Subjacency-ish input & intake:
A dependency cannot cross a very low probability region of structure (represented as a sequence of container nodes).

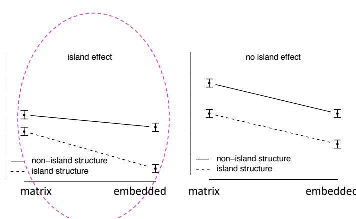




Lidz & Gagliardi 2015

Target state: Behavioral evidence of syntactic islands knowledge

	matrix	embedded
non-island	 *Who __ claimed that Lily forgot the necklace?*	 *What did the teacher claim that Lily forgot __?*
island	 *Who __ made the claim that Lily forgot the necklace?*	 **What did the teacher make the claim that Lily forgot __?*



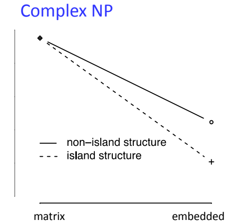
Pearl & Sprouse 2013a, 2013b, 2015

Subjacency-ish: Success!

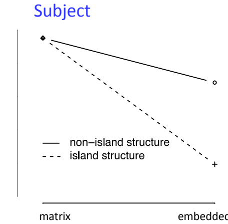
Superadditivity observed for all four islands — the qualitative behavior suggests that this learner has knowledge of these syntactic islands.

The Subjacency-ish representation that relies on container node trigram probabilities can solve this learning problem.

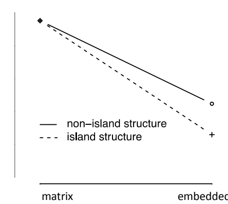
Complex NP



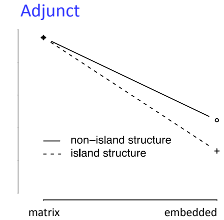
Subject



Whether



Adjunct



Pearl & Sprouse 2013a, 2013b, 2015

Subjacency-ish: Take away

Wh … $[_{CN1}$ … $[_{CN2}$ … $[_{CN3}$ … $[_{CN4}$ … $[_{CN5}$ … __]]

Representation validation
If dependencies are represented as container node sequences, acquisition works well for these four syntactic islands.

Pearl & Sprouse 2013a, 2013b, 2015

Subjacency-ish vs. Subjacency: What's in UG?

Subjacency-ish

Wh … $[_{CN1}$ … $[_{CN2}$ … $[_{CN3}$ … $[_{CN4}$ … $[_{CN5}$ … __]]

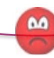
Fewer pieces of knowledge necessarily in UG + empirically-motivated alternative proposal for one component.

UG = innate + domain-specific

	Innate	Derived	Domain-specific	Domain-general
Attend to container nodes of a particular kind	?	?	*	
Low probability items are dispreferred	*			*

Subjacency

Wh … $[_{BN1}$ … $[_{BN2}$ … __]]



	Innate	Derived	Domain-specific	Domain-general
Attend to bounding nodes (BNs)	*		*	
Dependencies crossing 2+ BNs are not allowed	*		*	

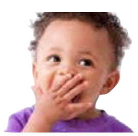
Pearl & Sprouse 2013a, 2013b, 2015

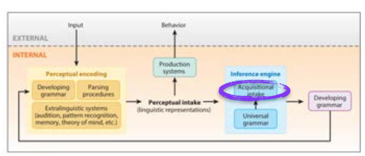
Recurring themes: Syntactic islands

Informing theories of representation & acquisition

Recurring themes, as seen in syntactic island acquisition:

(1) Broadening the set of relevant data in the acquisitional intake to include all *wh*-dependencies





Lidz & Gagliardi 2015

Pearl & Sprouse 2013a, 2013b, 2015

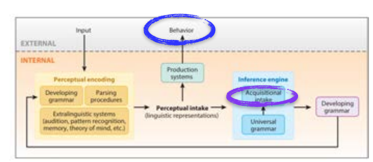
Recurring themes: Syntactic islands

Informing theories of representation & acquisition

Recurring themes, as seen in syntactic island acquisition:

(1) Broadening the set of relevant data in the acquisitional intake to include all *wh*-dependencies

(2) Evaluating output by how useful it is for generating acceptability judgment behavior



Lidz & Gagliardi 2015

Pearl & Sprouse 2013a, 2013b, 2015

Recurring themes: Syntactic islands

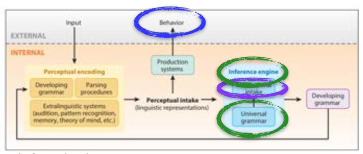
Informing theories of representation & acquisition

Recurring themes, as seen in syntactic island acquisition:

(1) Broadening the set of relevant data in the acquisitional intake to include all *wh*-dependencies

(2) Evaluating output by how useful it is for generating acceptability judgment behavior

(3) Not necessarily needing the prior knowledge we thought we did in UG: container nodes rather than bounding nodes, no domain-specific constraint on length



Lidz & Gagliardi 2015

Pearl & Sprouse 2013a, 2013b, 2015

Open questions

This learning strategy relying on the Subjacency-ish representation for *wh*-dependencies makes some developmental predictions – can we verify these experimentally?

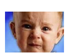
"*that*-trace" effect prediction:
Children initially disprefer all dependencies containing *that*, even ones adults allow, due to the infrequency of container node trigrams with CP_{that} in child-directed speech

Pearl & Sprouse 2013a, 2013b, 2015

Open questions

This learning strategy relying on the Subjacency-ish representation for *wh*-dependencies makes some developmental predictions – can we verify these experimentally?

"*that*-trace" effect prediction:
Children initially disprefer all dependencies containing *that*, even ones adults allow, due to the infrequency of container node trigrams with CP_{that} in child-directed speech

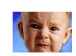
Subject extraction
*Who do you think ***that*** __ read the book? 
Who do you think __ read the book? 

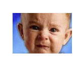
Pearl & Sprouse 2013a, 2013b, 2015

Open questions

This learning strategy relying on the Subjacency-ish representation for *wh*-dependencies makes some developmental predictions – can we verify these experimentally?

"*that*-trace" effect prediction:
Children initially disprefer all dependencies containing *that*, even ones adults allow, due to the infrequency of container node trigrams with CP_{that} in child-directed speech

Subject extraction
*Who do you think ***that*** __ read the book? 
Who do you think __ read the book? 

Object extraction
What do you think ***that*** he read __ ? 
What do you think he read __ ? 

Pearl & Sprouse 2013a, 2013b, 2015

Open questions

How does this learning strategy for *wh*-dependencies measure up cross-linguistically?

Island effects vary.
Ex: Italian does not have a subject island effect when the *wh*-dependency is part of a relative clause, though it does when the *wh*-dependency is part of a question. (Sprouse et al. in press)

Would the input naturally lead the Subjacency-ish learner to this distinction?

Pearl & Sprouse 2013a, 2013b, 2015

Open questions

Can we extend this learning strategy to create an integrated theory of syntactic acquisition?

Related phenomena: The distribution of gaps

Parasitic gaps: Dependencies that span an island (and so should be ungrammatical) but which are somehow rescued by another dependency in the utterance.

*Which book did you laugh [before reading __]? Adjunct island
Which book did you judge __$_{true}$ [before reading __$_{parasitic}$]?

Pearl & Sprouse 2013a, 2013b, 2015

Open questions

Can we extend this learning strategy to create an integrated theory of syntactic acquisition?

Related phenomena: The distribution of gaps

Across-the-board (ATB) extraction: Similar situation.

Which book did you [[read __] and [then review __]]?
dependency for both gaps: IP-VP-VP

*Which book did you [[read the paper] and [then review __]]? Coordinate structure island
dependency for gap: IP-VP-VP

*Which book did you [[read __] and [then review the paper]]?
dependency for gap: IP-VP-VP

Pearl & Sprouse 2013a, 2013b, 2015

Open questions

Can we extend this learning strategy to create an integrated theory of syntactic acquisition?

Semi-related phenomena: Binding dependencies

There don't appear to be the same restrictions on binding dependencies that there are on *wh*-dependencies.

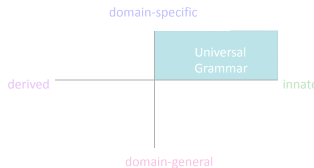
The boy thought the joke about himself was really funny.

*Who did the boy think [the joke about __] was really funny? Subject island

Pearl & Sprouse 2013a, 2013b, 2015

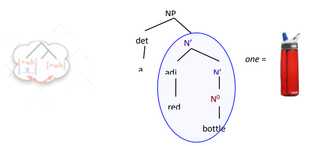
Today's Plan

Investigating Universal Grammar (UG)



Characterizing learning problems precisely enough to informatively model them

UG modeling forays



English anaphoric *one*

- **Why**? A traditional poverty-of-the-stimulus problem used to motivate specific proposals for the contents of UG.
- **What**? Look - a red bottle!

Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*

- **Why**? A traditional poverty-of-the-stimulus problem used to motivate specific proposals for the contents of UG.
- **What**? Look - a red bottle! Do you see another *one*?

Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*

- **Why**? A traditional poverty-of-the-stimulus problem used to motivate specific proposals for the contents of UG.
- **What**? Look - a red bottle! Do you see another *one*?

Process of interpretation: First determine the linguistic antecedent of *one* (what expression *one* is referring to) based on its syntactic category.
→ antecedent of *one* = "red bottle"

Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*

- **Why**? A traditional poverty-of-the-stimulus problem used to motivate specific proposals for the contents of UG.
- **What**? Look - a red bottle! Do you see another *one*?

Process of interpretation: 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

Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*

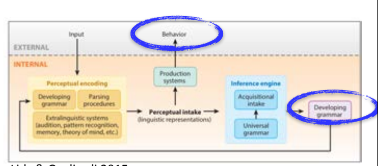
- **Why**? A traditional poverty-of-the-stimulus problem used to motivate specific proposals for the contents of UG.
- **What**? Look - a red bottle! Do you see another *one*?

Two steps:
(1) Identify linguistic antecedent (based on *one*'s syntactic category)
(2) Identify referent (based on linguistic antecedent)

Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Acquisition target

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

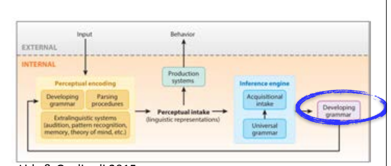


Lidz & Gagliardi 2015

Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Acquisition target

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

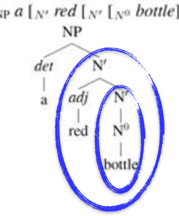


Lidz & Gagliardi 2015

Adult Knowledge

Standard linguistic theory (Chomsky 1970, Jackendoff 1977) has posited that *one* in these kinds of utterances is a syntactic category smaller than an entire noun phrase (NP), but larger than just a noun (N^0). This category has been called N', and includes strings like "bottle" and "red bottle".

$[_{NP}\ a\ [_{N'}\ red\ [_{N'}\ [_{N^0}\ bottle]]]]$

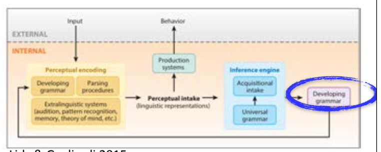


Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Acquisition target

Look - a red bottle!

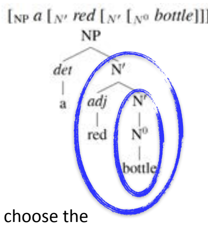
Do you see another *one*?



Lidz & Gagliardi 2015

Adult Knowledge

Because *one* is thought to be this same category (N′), available adult interpretations for *one* include both



"Do you see another bottle?"

and

"Do you see another red bottle?"

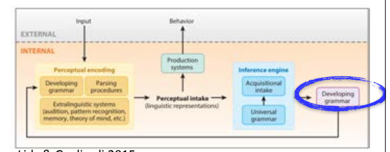
Additional preferences allow adults to choose the appropriate interpretation from these options in context.

Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Acquisition target

Look - a red bottle!

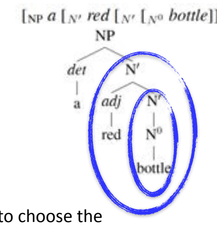
Do you see another *one*?



Lidz & Gagliardi 2015

Adult Knowledge

Because *one* is thought to be this same category (N′), available adult interpretations for *one* include both



"Do you see another bottle?"

and

"Do you see another red bottle?"

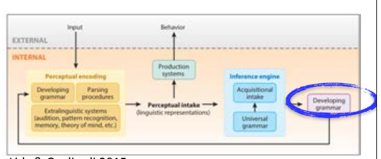
Additional preferences allow adults to choose the appropriate interpretation from these options in context.

Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Acquisition target

Look - a red bottle!

Do you see another *one*?



Lidz & Gagliardi 2015

Adult Knowledge

Syntactic category of *one* in this utterance = N′

Referent of *one* can be the object that contains the property in the modifier (RED BOTTLE)

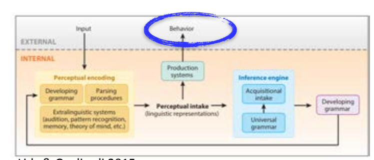
Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Acquisition target

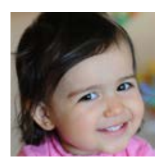
Child knowledge as measured by looking time behavior

Look - a red bottle!

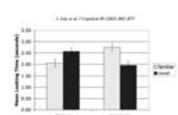
Now look…



Lidz & Gagliardi 2015



Child behavior at 18 months: Lidz et al. 2003



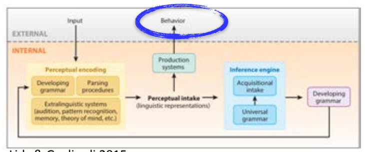
Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Acquisition target

Child knowledge as measured by looking time behavior

Look - a red bottle!

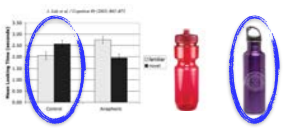
Now look…



Lidz & Gagliardi 2015

Child behavior at 18 months: Lidz et al. 2003

Control/Noun:
“What do you see now?”
“Do you see another bottle?”
Baseline novelty preference
Average probability of looking to same color bottle: 0.459



Prefer to look at novel bottle.

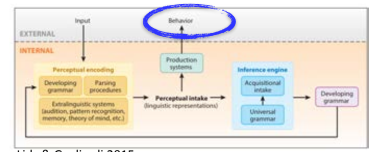
Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Acquisition target

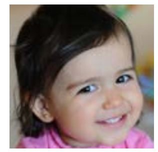
Child knowledge as measured by looking time behavior

Look - a red bottle!

Now look…



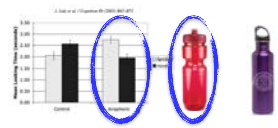
Lidz & Gagliardi 2015



Child behavior at 18 months: Lidz et al. 2003

Control/Noun:
“What do you see now?”
“Do you see another bottle?”
Prefer to look at novel bottle.
(0.459 to same color)

Anaphoric/Adjective-Noun:
“Do you see another one?”
“Do you see another red bottle?”
Adjusted familiarity preference
Average probability of looking to same color bottle: 0.587



Prefer to look at same color bottle.

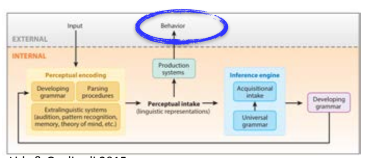
Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Acquisition target

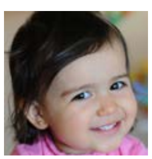
Child knowledge as measured by looking time behavior

Look - a red bottle!

Now look…



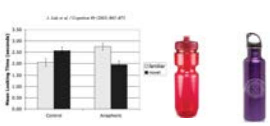
Lidz & Gagliardi 2015



Child behavior at 18 months: Lidz et al. 2003

Control/Noun:
“What do you see now?”
“Do you see another bottle?”
Prefer to look at novel bottle.
(0.459 to same color)

Anaphoric/Adjective-Noun:
“Do you see another one?”
“Do you see another red bottle?”
Prefer to look at same color bottle.
(0.587 to same color)



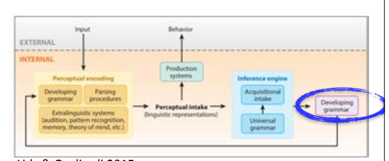
Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Acquisition target

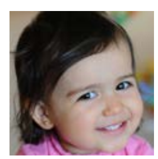
Child knowledge as measured by looking time behavior

Look - a red bottle!

Now look…



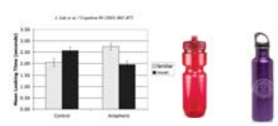
Lidz & Gagliardi 2015



Child behavior at 18 months: Lidz et al. 2003

Control/Noun:
“What do you see now?”
“Do you see another bottle?”
Prefer to look at novel bottle.
(0.459 to same color)

Anaphoric/Adjective-Noun:
“Do you see another one?”
“Do you see another red bottle?”
Prefer to look at same color bottle.
(0.587 to same color)



Developed knowledge according to Lidz et al. 2003:
18-month-olds interpret *one*’s antecedent as “red bottle” (an N′) and its referent as the RED BOTTLE.

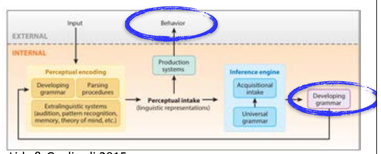
Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Acquisition target

Target state for acquisition: knowledge and behavior

Look - a red bottle!

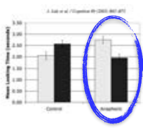
Now look…



Lidz & Gagliardi 2015

Child behavior at 18 months: Lidz et al. 2003

Control/Noun:	Anaphoric/Adjective-Noun:
"What do you see now?"	**"Do you see another one?"**
"Do you see another bottle?"	**"Do you see another red bottle?"**
Prefer to look at novel bottle (0.459 to same color)	Prefer to look at same color bottle (0.587 to same color)

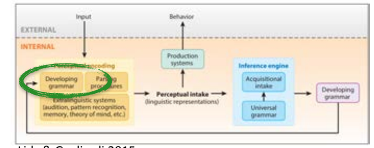


Developed knowledge according to Lidz et al. 2003:
18-month-olds interpret *one*'s antecedent as "red bottle" (an N') and its referent as the RED BOTTLE.

Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Representations

Proposed solutions for necessary knowledge & learning biases



Lidz & Gagliardi 2015

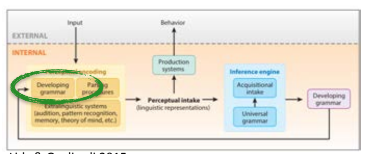
Things in common:

- ✦ Syntactic categories exist (particularly NP, N', and N^0), and can be recognized.

Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Representations

Proposed solutions for necessary knowledge & learning biases



Lidz & Gagliardi 2015

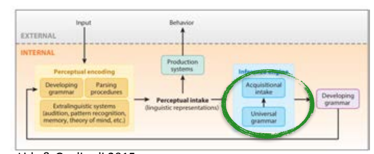
Things in common:

- ✦ Syntactic categories exist (particularly NP, N', and N^0), and can be recognized.
- ✦ Anaphoric elements like *one* take linguistic antecedents of the same category.

Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Representations

Proposed solutions for necessary knowledge & learning biases



Lidz & Gagliardi 2015

Things that differ:

- ✦ Which input is considered relevant from the perceptual intake = acquisitional intake

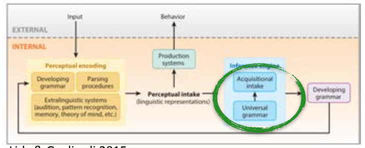
Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Representations

Proposed solutions for necessary knowledge & learning biases

Things that differ:

✦ Which input is considered relevant from the perceptual intake = acquisitional intake



Lidz & Gagliardi 2015

Baker (1978): One that won't work = DirUnamb

Only utterances directly using *one* are relevant for learning about anaphoric *one*.

Only utterances where *one*'s antecedent is unambiguous are relevant.

DirUnamb: specific combination of utterance and situation

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

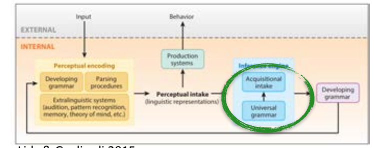
Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Representations

Proposed solutions for necessary knowledge & learning biases

Things that differ:

✦ Which input is considered relevant from the perceptual intake = acquisitional intake



Lidz & Gagliardi 2015

Baker (1978): One that won't work = DirUnamb

Only utterances directly using *one* are relevant for learning about anaphoric *one*.

Only utterances where *one*'s antecedent is unambiguous are relevant.

Why won't it work? The direct unambiguous data are too sparse. There's nothing to learn from.

Pearl & Mis 2011, 2016 affirmation:

0 examples in the 17,521 utterances in the Brown-Eve corpus (Brown 1973) from CHILDES.

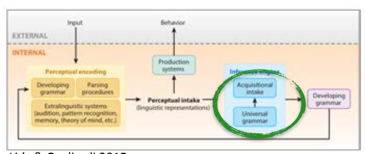
Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Representations

Proposed solutions for necessary knowledge & learning biases

Things that differ:

✦ Which input is considered relevant from the perceptual intake = acquisitional intake



Lidz & Gagliardi 2015

Baker (1978): One that could work = DirUnamb + N'

Only utterances directly using *one* are relevant for learning about anaphoric *one*.

Only utterances where *one*'s antecedent is unambiguous are relevant.

Children already know that *one* can't be N^0, so it must be N'. UG knowledge

This solves the problem of *one*'s syntactic category.

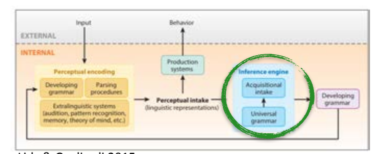
Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Representations

Proposed solutions for necessary knowledge & learning biases

Things that differ:

✦ Which input is considered relevant from the perceptual intake = acquisitional intake



Lidz & Gagliardi 2015

Pearl & Lidz 2009: One that doesn't work = DirEO

Only utterances directly using *one* are relevant for learning about anaphoric *one*.

Use probabilistic inference to leverage ambiguous information about *one.*

All ambiguous data are relevant (Equal Opportunity).

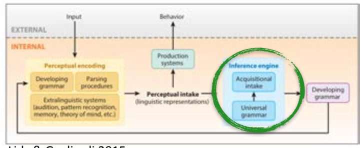
Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Representations

Proposed solutions for necessary knowledge & learning biases

Things that differ:

✦ Which input is considered relevant from the perceptual intake = acquisitional intake



Lidz & Gagliardi 2015

Pearl & Lidz 2009: One that doesn't work = DirEO

Only utterances directly using *one* are relevant for learning about anaphoric *one*.

Use probabilistic inference to leverage ambiguous information about *one*.

All ambiguous data are relevant (Equal Opportunity).

DirRefSynAmb: Ambiguous about whether antecedent is "bottle" (N^0, N') or "red bottle" (N').

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

0.66% of utterances containing a pronoun in Brown-Eve corpus

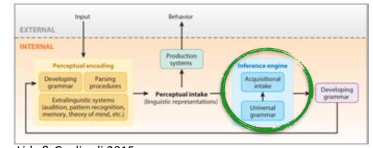
Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Representations

Proposed solutions for necessary knowledge & learning biases

Things that differ:

✦ Which input is considered relevant from the perceptual intake = acquisitional intake



Lidz & Gagliardi 2015

Pearl & Lidz 2009: One that doesn't work = DirEO

Only utterances directly using *one* are relevant for learning about anaphoric *one*.

Use probabilistic inference to leverage ambiguous information about *one*.

All ambiguous data are relevant (Equal Opportunity).

DirSynAmb: Ambiguous about antecedent category (*bottle* = N^0, N').

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

7.52% of utterances containing a pronoun in Brown-Eve corpus

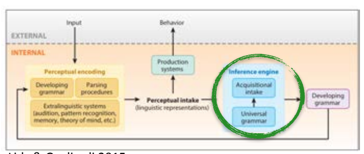
Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Representations

Proposed solutions for necessary knowledge & learning biases

Things that differ:

✦ Which input is considered relevant from the perceptual intake = acquisitional intake



Lidz & Gagliardi 2015

Pearl & Lidz 2009: One that doesn't work = DirEO

Only utterances directly using *one* are relevant for learning about anaphoric *one*.

Use probabilistic inference to leverage ambiguous information about *one*.

All ambiguous data are relevant (Equal Opportunity).

DirSynAmb: Ambiguous about antecedent category (*bottle* = N^0, N').

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

Turn out to be harmful to learning - they cause the learner to think *one's* category should be N^0.

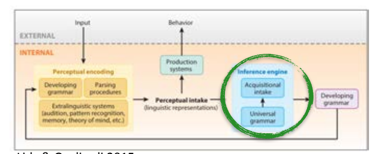
Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Representations

Proposed solutions for necessary knowledge & learning biases

Things that differ:

✦ Which input is considered relevant from the perceptual intake = acquisitional intake



Lidz & Gagliardi 2015

Pearl & Lidz 2009, Regier & Gahl 2004: One that does work for target knowledge = DirFiltered

Only utterances directly using *one* are relevant for learning about anaphoric *one*.

Use probabilistic inference to leverage ambiguous information about *one*.

Filter out the harmful DirSynAmb data.

DirSynAmb: Ambiguous about antecedent category (*bottle* = N^0, N').

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

Turn out to be harmful to learning - they cause the learner to think *one's* category should be N^0.

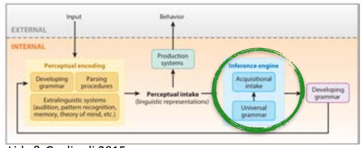
Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Representations

Proposed solutions for necessary knowledge & learning biases

Things that differ:

✦ Which input is considered relevant from the perceptual intake = acquisitional intake



Lidz & Gagliardi 2015

Pearl & Mis 2011, 2016: One that could work = IndirPro

O~~nl~~y utterances directly using *one* are relevant for learning about anaphoric *one*.

Use probabilistic inference to leverage ambiguous information about *one*.

Utterances using other pronouns anaphorically are relevant for learning about anaphoric *one*.
This is indirect evidence coming from other pronouns.

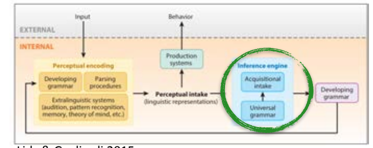
Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Representations

Proposed solutions for necessary knowledge & learning biases

Things that differ:

✦ Which input is considered relevant from the perceptual intake = acquisitional intake



Lidz & Gagliardi 2015

Pearl & Mis 2011, 2016: One that could work = IndirPro

O~~nl~~y utterances directly using *one* are relevant for learning about anaphoric *one*.

Use probabilistic inference to leverage ambiguous information about *one*.

Utterances using other pronouns anaphorically are relevant for learning about anaphoric *one*.
This is indirect evidence coming from other pronouns.

IndirUnamb: Relevant because indicates whether antecedent includes the mentioned property (it always does here), which is helpful when choosing between different interpretation options in other contexts.

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

a red bottle

8.42% of utterances containing a pronoun in Brown-Eve corpus

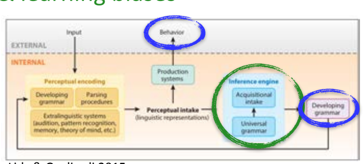
Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Representations

Proposed solutions for necessary knowledge & learning biases

Things that differ:

✦ Which input is considered relevant from the perceptual intake = acquisitional intake



Lidz & Gagliardi 2015

Learning proposal comparisons

Successful?

	Unamb	*one*≠N^0	ProbInf	–DirSynAmb	+OtherPro	Representations	Behavior
DirUnamb	✓					?	?

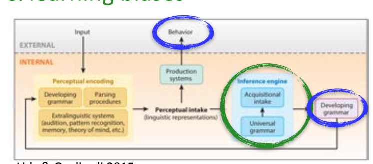
Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Representations

Proposed solutions for necessary knowledge & learning biases

Things that differ:

✦ Which input is considered relevant from the perceptual intake = acquisitional intake



Lidz & Gagliardi 2015

Learning proposal comparisons

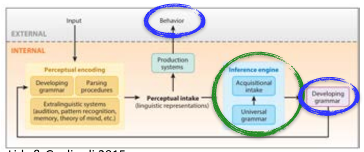
Successful?

	Unamb	*one*≠N^0	ProbInf	–DirSynAmb	+OtherPro	Representations	Behavior
DirUnamb	✓					?	?
DirUnamb + N′	✓	✓				?	?

Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Representations

Proposed solutions for necessary knowledge & learning biases



Lidz & Gagliardi 2015

Things that differ:

✦ Which input is considered relevant from the perceptual intake = acquisitional intake

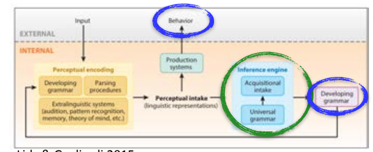
Learning proposal comparisons

	Unamb	*one*≠N^0	ProbInf	–DirSynAmb	+OtherPro	Successful? Representations	Behavior
DirUnamb	✓					?	?
DirUnamb + N′	✓	✓				?	?
DirFiltered			✓	✓		✓	?

Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Representations

Proposed solutions for necessary knowledge & learning biases



Lidz & Gagliardi 2015

Things that differ:

✦ Which input is considered relevant from the perceptual intake = acquisitional intake

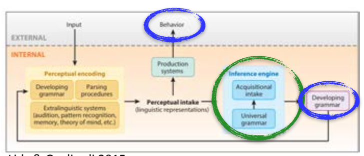
Learning proposal comparisons

	Unamb	*one*≠N^0	ProbInf	–DirSynAmb	+OtherPro	Successful? Representations	Behavior
DirUnamb	✓					?	?
DirUnamb + N′	✓	✓				?	?
DirFiltered			✓	✓		✓	?
DirEO			✓			✗	?

Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Representations

Proposed solutions for necessary knowledge & learning biases



Lidz & Gagliardi 2015

Things that differ:

✦ Which input is considered relevant from the perceptual intake = acquisitional intake

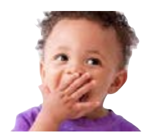
Learning proposal comparisons

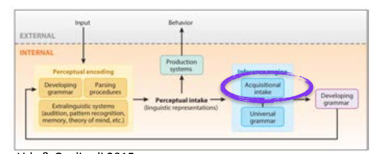
	Unamb	*one*≠N^0	ProbInf	–DirSynAmb	+OtherPro	Successful? Representations	Behavior
DirUnamb	✓					?	?
DirUnamb + N′	✓	✓				?	?
DirFiltered			✓	✓		✓	?
DirEO			✓			✗	?
IndirPro			✓		✓	?	?

Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Data intake

Data intake: The data relevant for learning





Lidz & Gagliardi 2015

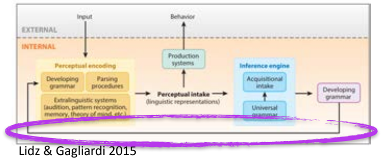
Data potentially in the acquisitional intake

Data type	Example	Learning strategies using these data
DirUnamb	*Look– a red bottle! There isn't another one here, though.*	DirUnamb, DirUnamb + N′, DirFiltered, DirEO, IndirPro
DirRefSynAmb	*Look– a red bottle! Oh, look– another one!*	DirFiltered, DirEO, IndirPro
DirSynAmb	*Look– a bottle! Oh, look– another one!*	DirEO, IndirPro
IndirUnamb	*Look a red bottle! I want it/one.*	IndirPro

Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Learning period

Learning period: How long children have to learn = how much data



Lidz & Gagliardi 2015

✦ Syntactic categories exist (particularly NP, N′, and N^0), and can be recognized.

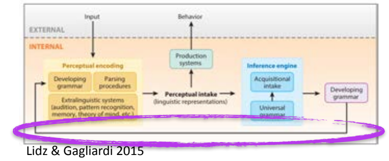
Before this learning process can begin, children need to know something about syntactic categories. Experimental data from Booth & Waxman (2003) suggests they recognize linguistic markers of categories like Noun and Adjective around 14 months.

Beginning: 14 months

Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Learning period

Learning period: How long children have to learn = how much data



Lidz & Gagliardi 2015

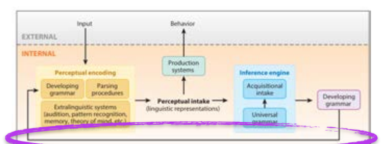
The experimental data from Lidz et al. (2003) suggest they should reach the knowledge state that generates that observable behavior by 18 months.

Beginning: 14 months
End: 18 months

Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Learning period

Learning period: How long children have to learn = how much data



Lidz & Gagliardi 2015

Using empirical estimates from Hart & Risley (1995), we can estimate this as approximately 36,500 data points containing an anaphoric pronoun.

Beginning: 14 months
End: 18 months

= 4 months' worth of data

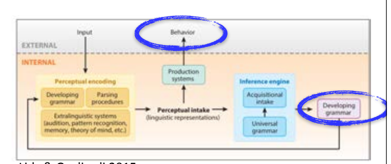
Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Target state

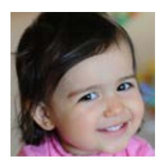
Target state: knowledge and behavior

Look - a red bottle!

Now look…

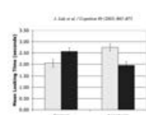


Lidz & Gagliardi 2015



Child behavior at 18 months: Lidz et al. 2003

Control/Noun:	Anaphoric/Adjective-Noun:
"What do you see now?"	**"Do you see another one?"**
"Do you see another bottle?"	**"Do you see another red bottle?"**
Prefer to look at novel bottle (0.459 to same color)	Prefer to look at same color bottle (0.587 to same color)

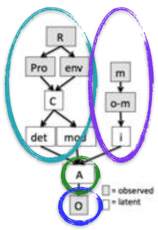


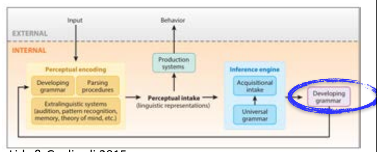
Developed knowledge according to Lidz et al. 2003: 18-month-olds interpret *one*'s antecedent as "red bottle" (an N′) and its referent as the RED BOTTLE.

Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Learning process

Update & iteration of developing grammar





Lidz & Gagliardi 2015

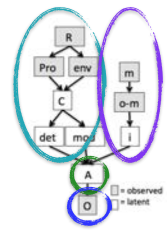
Model of understanding a referential expression involving an anaphoric pronoun, which includes both syntactic information and referential information when determining the antecedent which then picks out the referent.

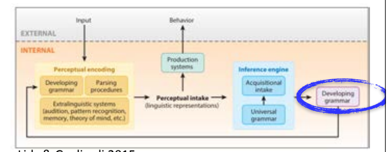
Developed knowledge according to Lidz et al. 2003: 18-month-olds interpret *one*'s antecedent as "red bottle" (an N') and its referent as the RED BOTTLE.

Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Learning process

Update & iteration of developing grammar





Lidz & Gagliardi 2015

$p_{N'}$ = probability that *one*'s category is N' (vs. N^0)

p_{incl} = probability that *one*'s antecedent includes the mentioned modifier (e.g., *red*) vs. not

d_x = probability that data point indicates this

$$p_x = \frac{\alpha + d_x}{\alpha + \beta + D_x}, \alpha = \beta = 1$$

D_x = 1 for every data point encountered

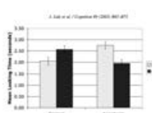
Model of understanding a referential expression involving an anaphoric pronoun, which includes both syntactic information and referential information when determining the antecedent which then picks out the referent.

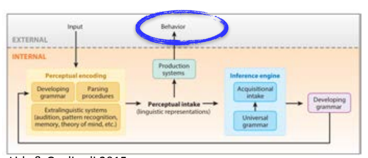
Developed knowledge according to Lidz et al. 2003: 18-month-olds interpret *one*'s antecedent as "red bottle" (an N') and its referent as the RED BOTTLE.

Pearl & Mis 2011, Pearl & Mis 2016

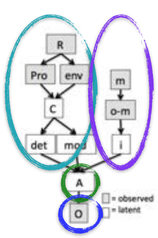
English anaphoric *one*: Learning process

Update & iteration of developing grammar





Lidz & Gagliardi 2015



Control/Noun:
"What do you see now?"
"Do you see another bottle?"
Prefer to look at novel bottle.
(0.459 to same color)

Anaphoric/Adjective-Noun:
"Do you see another one?"
"Do you see another red bottle?"
Prefer to look at same color bottle.
(0.587 to same color)

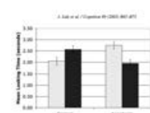
Model of understanding a referential expression involving an anaphoric pronoun, which includes both syntactic information and referential information when determining the antecedent which then picks out the referent.

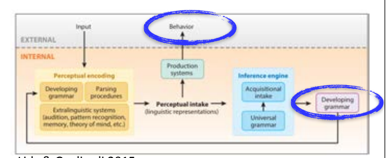
p_{beh} = probability of producing target behavior (looking to same color bottle)

Pearl & Mis 2011, Pearl & Mis 2016

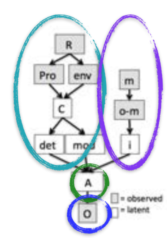
English anaphoric *one*: Learning process

Update & iteration of developing grammar





Lidz & Gagliardi 2015



Control/Noun:
"What do you see now?"
"Do you see another bottle?"
Prefer to look at novel bottle.
(0.459 to same color)

Anaphoric/Adjective-Noun:
"Do you see another one?"
"Do you see another red bottle?"
Prefer to look at same color bottle.
(0.587 to same color)

Model of understanding a referential expression involving an anaphoric pronoun, which includes both syntactic information and referential information when determining the antecedent which then picks out the referent.

$p_{rep|beh}$ = probability of having target representation (antecedent = "red bottle") when producing target behavior (looking to same color bottle)

Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Learning results

Averages over 1000 simulations, standard deviations in parentheses.

Note: Target p_{beh} = 0.587, all other target p = 1.000

$p_{N'}$	
p_{incl}	
p_{beh}	
$p_{rep\|beh}$	

Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Learning results

Averages over 1000 simulations, standard deviations in parentheses.

Note: Target p_{beh} = 0.587, all other target p = 1.000

	DirUnamb
$p_{N'}$	0.500 (<0.01)
p_{incl}	0.500 (<0.01)
p_{beh}	0.475 (<0.01)
$p_{rep\|beh}$	0.158 (<0.01)

A learner who only looks at direct unambiguous data has no data to learn from, so it learns nothing. (Poverty of the stimulus.)

It's at chance for having the target syntactic and referential knowledge necessary to choose the correct antecedent.

It doesn't generate the observed toddler looking preference, and it's unlikely to have the target representation if it looks at the familiar bottle.

Pearl & Mis 2011, Pearl & Mis 2016

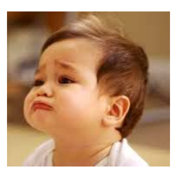
English anaphoric *one*: Learning results

Averages over 1000 simulations, standard deviations in parentheses.

Note: Target p_{beh} = 0.587, all other target p = 1.000

	DirUnamb
$p_{N'}$	0.500 (<0.01)
p_{incl}	0.500 (<0.01)
p_{beh}	0.475 (<0.01)
$p_{rep\|beh}$	0.158 (<0.01)

Implication: Something else is needed.
(Baker (1978)'s original observation)



Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Learning results

Averages over 1000 simulations, standard deviations in parentheses.

Note: Target p_{beh} = 0.587, all other target p = 1.000

	DirUnamb	DirUnamb + N'
$p_{N'}$	0.500 (<0.01)	1.000
p_{incl}	0.500 (<0.01)	
p_{beh}	0.475 (<0.01)	
$p_{rep\|beh}$	0.158 (<0.01)	

What if the learner also knows that *one* is category N'? (Baker 1978)

Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Learning results

Averages over 1000 simulations, standard deviations in parentheses.

Note: Target p_{beh} = 0.587, all other target p = 1.000

	DirUnamb	DirUnamb + N'
$p_{N'}$	0.500 (<0.01)	1.000
p_{incl}	0.500 (<0.01)	0.500 (<0.01)
p_{beh}	0.475 (<0.01)	0.492 (<0.01)
$p_{rep\|beh}$	0.158 (<0.01)	0.306 (<0.01)

This learner still has no data to learn from, so it learns nothing about the correct referential knowledge necessary to choose the correct antecedent.

This lack of referential knowledge causes it not to generate the observed toddler looking preference in context, and even if it happens to look at the familiar bottle, to be unlikely to have the target representation when doing so.

Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Learning results

Averages over 1000 simulations, standard deviations in parentheses.

Note: Target p_{beh} = 0.587, all other target p = 1.000

	DirUnamb	DirUnamb + N'
$p_{N'}$	0.500 (<0.01)	1.000
p_{incl}	0.500 (<0.01)	0.500 (<0.01)
p_{beh}	0.475 (<0.01)	0.492 (<0.01)
$p_{rep\|beh}$	0.158 (<0.01)	0.306 (<0.01)

Implication: Knowing *one* is category N' isn't sufficient to generate target behavior if only direct unambiguous data are relevant.

Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Learning results

Averages over 1000 simulations, standard deviations in parentheses.

Note: Target p_{beh} = 0.587, all other target p = 1.000

	DirUnamb	DirUnamb + N'	DirFiltered
$p_{N'}$	0.500 (<0.01)	1.000	0.991 (<0.01)
p_{incl}	0.500 (<0.01)	0.500 (<0.01)	0.963 (<0.01)
p_{beh}	0.475 (<0.01)	0.492 (<0.01)	0.574 (<0.01)
$p_{rep\|beh}$	0.158 (<0.01)	0.306 (<0.01)	0.918 (<0.01)

The DirFiltered learner (Regier & Gahl 2004, Pearl & Lidz 2009) believes *one* is N' when it is smaller than NP and a mentioned property should be included in the antecedent, as found previously.

It's also close to generating the observed toddler looking preference, and is likely to have the target representation when looking at the familiar bottle.

Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Learning results

Averages over 1000 simulations, standard deviations in parentheses.

Note: Target p_{beh} = 0.587, all other target p = 1.000

	DirUnamb	DirUnamb + N'	DirFiltered
$p_{N'}$	0.500 (<0.01)	1.000	0.991 (<0.01)
p_{incl}	0.500 (<0.01)	0.500 (<0.01)	0.963 (<0.01)
p_{beh}	0.475 (<0.01)	0.492 (<0.01)	0.574 (<0.01)
$p_{rep\|beh}$	0.158 (<0.01)	0.306 (<0.01)	0.918 (<0.01)

Implication: This new finding suggests this is a pretty successful learning strategy for matching the available behavioral data.

Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Learning results

Averages over 1000 simulations, standard deviations in parentheses.

Note: Target p_{beh} = 0.587, all other target p = 1.000

	DirUnamb	DirUnamb + N'	DirFiltered	DirEO
$p_{N'}$	0.500 (<0.01)	1.000	0.991 (<0.01)	0.246 (0.03)
p_{incl}	0.500 (<0.01)	0.500 (<0.01)	0.963 (<0.01)	0.379 (0.05)
p_{beh}	0.475 (<0.01)	0.492 (<0.01)	0.574 (<0.01)	0.464 (<0.01)
$p_{rep\|beh}$	0.158 (<0.01)	0.306 (<0.01)	0.918 (<0.01)	0.050 (0.01)

The DirEO learner (explored by Pearl & Lidz 2009) prefers *one* to be N^0 when it is smaller than NP, and does not believe the mentioned property should be included in the antecedent. Neither of these is the target knowledge.

This causes the learner not to generate the observed toddler looking preference, and not to have the target representation if it looks at the familiar bottle.

Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Learning results

Averages over 1000 simulations, standard deviations in parentheses.

Note: Target p_{beh} = 0.587, all other target p = 1.000

	DirUnamb	DirUnamb + N'	DirFiltered	DirEO
$p_{N'}$	0.500 (<0.01)	1.000	0.991 (<0.01)	0.246 (0.03)
p_{incl}	0.500 (<0.01)	0.500 (<0.01)	0.963 (<0.01)	0.379 (0.05)
p_{beh}	0.475 (<0.01)	0.492 (<0.01)	0.574 (<0.01)	0.464 (<0.01)
$p_{rep\|beh}$	0.158 (<0.01)	0.306 (<0.01)	0.918 (<0.01)	0.050 (0.01)

Implication: This new finding suggests this isn't a good learning strategy for matching the available behavioral data.

Pearl & Mis 2011, Pearl & Mis 2016

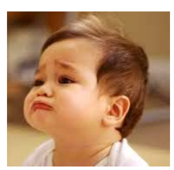
English anaphoric *one*: Learning results

Averages over 1000 simulations, standard deviations in parentheses.

Note: Target p_{beh} = 0.587, all other target p = 1.000

	DirUnamb	DirUnamb + N'	DirFiltered	DirEO	IndirPro
$p_{N'}$	0.500 (<0.01)	1.000	0.991 (<0.01)	0.246 (0.03)	0.368 (0.04)
p_{incl}	0.500 (<0.01)	0.500 (<0.01)	0.963 (<0.01)	0.379 (0.05)	1.000 (<0.01)
p_{beh}	0.475 (<0.01)	0.492 (<0.01)	0.574 (<0.01)	0.464 (<0.01)	
$p_{rep\|beh}$	0.158 (<0.01)	0.306 (<0.01)	0.918 (<0.01)	0.050 (0.01)	

The IndirPro learner robustly decides the antecedent should include the mentioned property. However, this learner has a moderate dispreference for believing *one* is N' when it is smaller than NP. This isn't the target representation, w.r.t syntactic category.



Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Learning results

Averages over 1000 simulations, standard deviations in parentheses.

Note: Target p_{beh} = 0.587, all other target p = 1.000

	DirUnamb	DirUnamb + N'	DirFiltered	DirEO	IndirPro
$p_{N'}$	0.500 (<0.01)	1.000	0.991 (<0.01)	0.246 (0.03)	0.368 (0.04)
p_{incl}	0.500 (<0.01)	0.500 (<0.01)	0.963 (<0.01)	0.379 (0.05)	1.000 (<0.01)
p_{beh}	0.475 (<0.01)	0.492 (<0.01)	0.574 (<0.01)	0.464 (<0.01)	0.587 (<0.01)
$p_{rep\|beh}$	0.158 (<0.01)	0.306 (<0.01)	0.918 (<0.01)	0.050 (0.01)	0.998 (<0.01)

However…this learner still generates the observed toddler looking preference perfectly, and has the target representation when looking at the familiar bottle.

Why? The learner believes very strongly that the mentioned property must be included in the antecedent.

Only one antecedent allows this: $[_{N'}$ red$[_{N'}[_{N0}$ bottle$]]]$

Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Learning results

Averages over 1000 simulations, standard deviations in parentheses.

Note: Target p_{beh} = 0.587, all other target p = 1.000

	DirUnamb	DirUnamb + N'	DirFiltered	DirEO	IndirPro
$p_{N'}$	0.500 (<0.01)	1.000	0.991 (<0.01)	0.246 (0.03)	0.368 (0.04)
p_{incl}	0.500 (<0.01)	0.500 (<0.01)	0.963 (<0.01)	0.379 (0.05)	1.000 (<0.01)
p_{beh}	0.475 (<0.01)	0.492 (<0.01)	0.574 (<0.01)	0.464 (<0.01)	0.587 (<0.01)
$p_{rep\|beh}$	0.158 (<0.01)	0.306 (<0.01)	0.918 (<0.01)	0.050 (0.01)	0.998 (<0.01)

So, because the antecedent includes the mentioned property, it and the pronoun referring to it (*one*) must be N' *in this context* - even if the learner believes *one* is not N' in general.

Only one antecedent allows this: $[_{N'}$ red$[_{N'}[_{N0}$ bottle$]]]$

Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Learning results

Averages over 1000 simulations, standard deviations in parentheses.

Note: Target p_{beh} = 0.587, all other target p = 1.000

	DirUnamb	DirUnamb + N'	DirFiltered	DirEO	IndirPro
$p_{N'}$	0.500 (<0.01)	1.000	0.991 (<0.01)	0.246 (0.03)	0.368 (0.04)
p_{incl}	0.500 (<0.01)	0.500 (<0.01)	0.963 (<0.01)	0.379 (0.05)	1.000 (<0.01)
p_{beh}	0.475 (<0.01)	0.492 (<0.01)	0.574 (<0.01)	0.464 (<0.01)	0.587 (<0.01)
$p_{rep\|beh}$	0.158 (<0.01)	0.306 (<0.01)	0.918 (<0.01)	0.050 (0.01)	0.998 (<0.01)

Implication: A learner viewing other pronoun data as relevant can generate target behavior without necessarily reaching the target knowledge state – instead, this learner has a context-sensitive representation (depending on whether a property was mentioned).

Pearl & Mis 2011, Pearl & Mis 2016

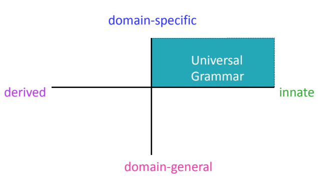
English anaphoric *one*: Learning results

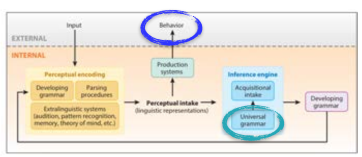
Averages over 1000 simulations, standard deviations in parentheses.

Note: Target p_{beh} = 0.587, all other target p = 1.000

	DirUnamb	DirUnamb + N'	DirFiltered	DirEO	IndirPro
$p_{N'}$	0.500 (<0.01)	1.000	0.991 (<0.01)	0.246 (0.03)	0.368 (0.04)
p_{incl}	0.500 (<0.01)	0.500 (<0.01)	0.963 (<0.01)	0.379 (0.05)	1.000 (<0.01)
p_{beh}	0.475 (<0.01)	0.492 (<0.01)	0.574 (<0.01)	0.464 (<0.01)	0.587 (<0.01)
$p_{rep\|beh}$	0.158 (<0.01)	0.306 (<0.01)	0.918 (<0.01)	0.050 (0.01)	0.998 (<0.01)

Let's look at the strategies that worked and see what the implications are for Universal Grammar, as compared to the original UG proposal by Baker that didn't work.

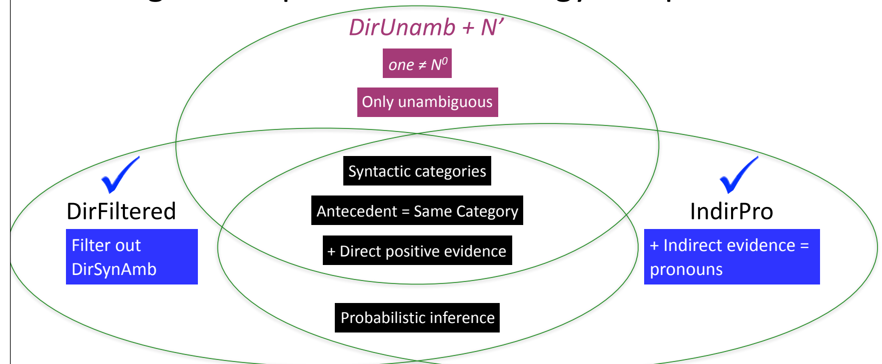


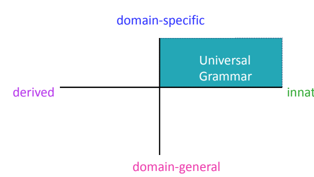


Lidz & Gagliardi 2015

Pearl & Mis 2011, Pearl & Mis 2016

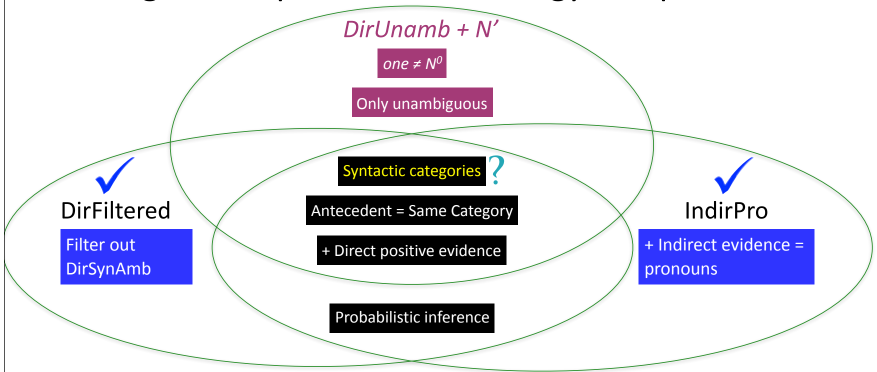
English anaphoric *one*: Strategy components

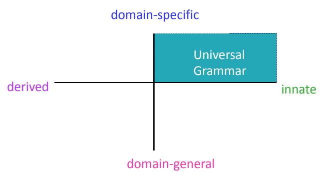




Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Strategy components





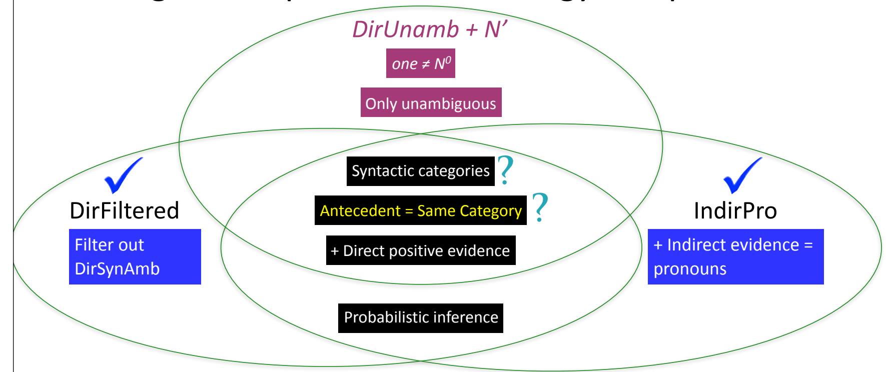
Things in common:

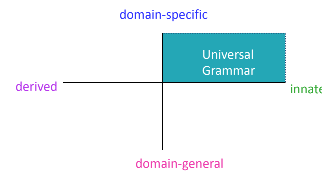
It may be possible to derive the domain-specific knowledge of the specific syntactic categories needed using distributional clustering techniques over words…but that remains to be shown.

Some innate knowledge may be necessary (UG).

Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Strategy components





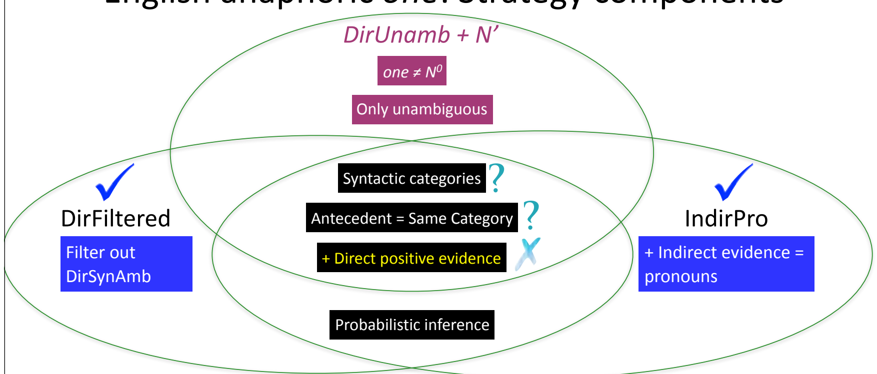
Things in common:

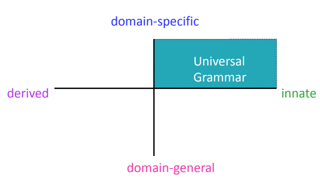
It may be possible to derive the domain-specific knowledge that anaphoric antecedents are the same category by observing the category of antecedents that are unambiguous…but that remains to be shown.

Some innate knowledge may be necessary (UG).

Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Strategy components



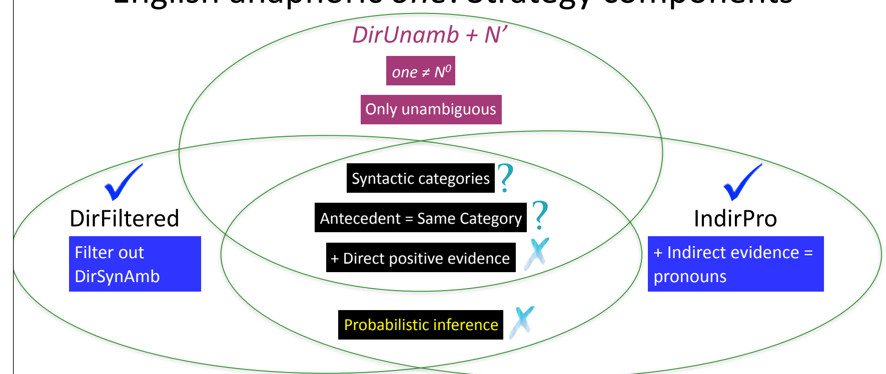


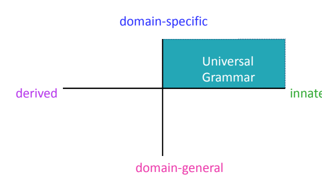
Things in common:

It seems likely that the preference to consider direct positive evidence relevant is innate and domain-general.

Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Strategy components





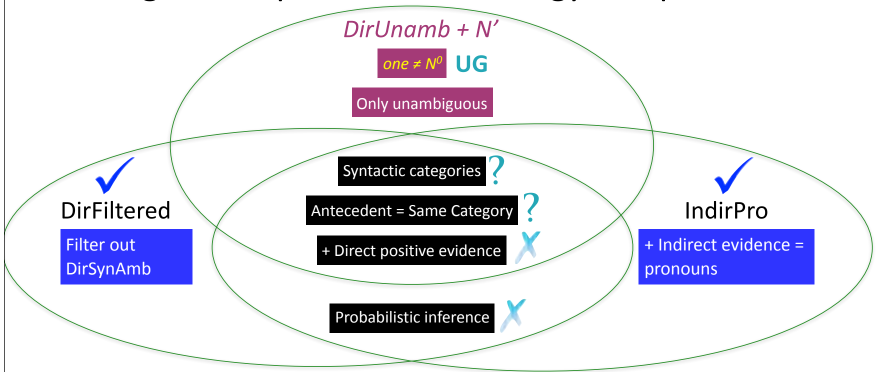
Things in common:

Similarly, the preference to use probabilistic inference to leverage the information in ambiguous data seems likely to be innate and domain-general.

While this is a new strategy component, it's unlikely to be part of UG.

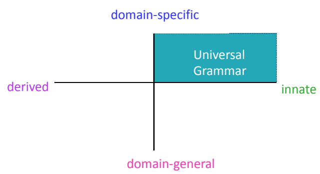
Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Strategy components



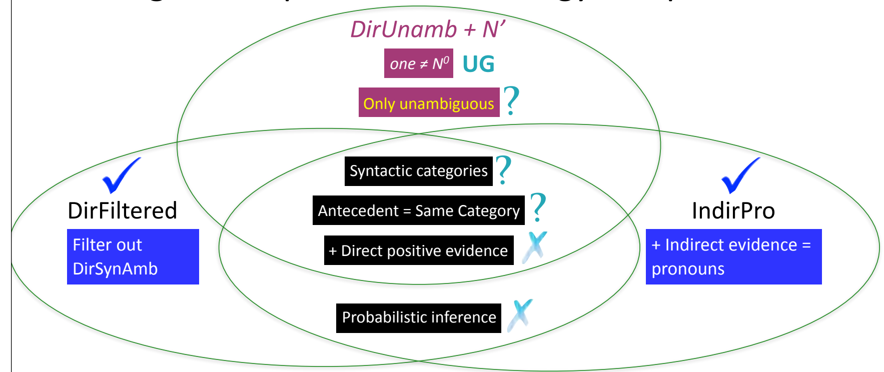
Old unsuccessful proposal:

The domain-specific knowledge that *one* is not category N^0 was thought to be innate and so part of UG.



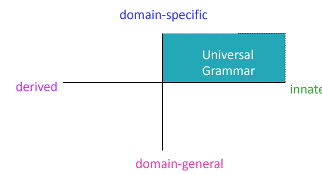
Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Strategy components



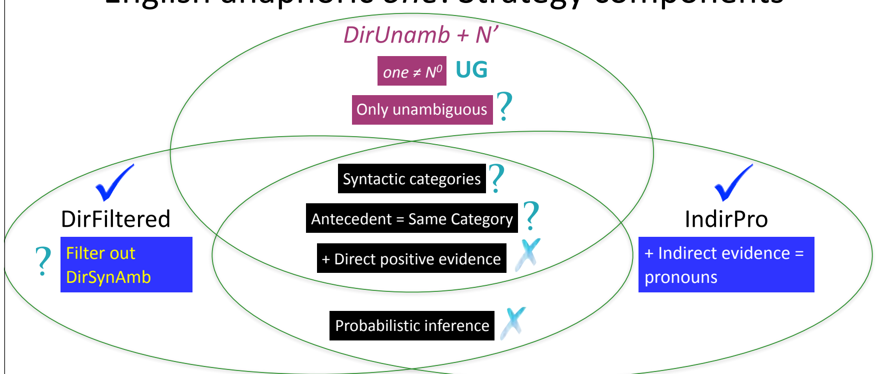
Old unsuccessful proposal:

The preference to rely only on unambiguous evidence might be innate, but could well be domain-general and so not part of UG.



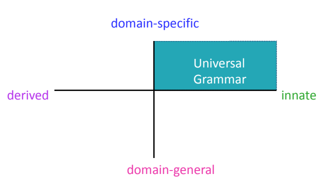
Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Strategy components



Successful DirFiltered proposal

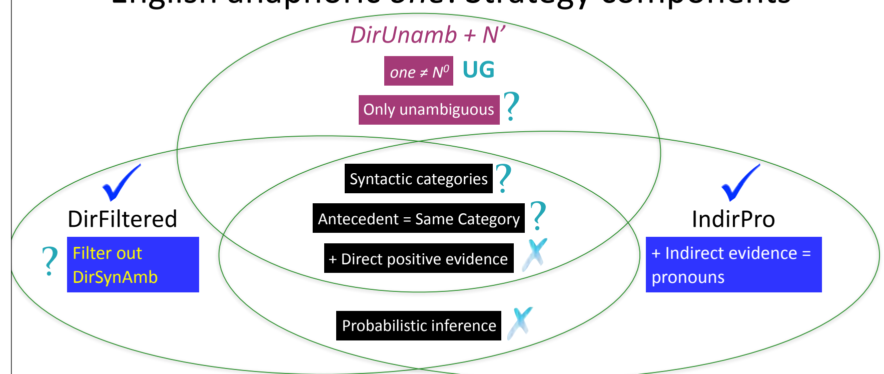
The domain-specific preference to filter out data where only the syntactic category is uncertain (while the referent is clear) may be innate and so part of UG, or it may be derived from an innate, domain-general preference to learn in cases of uncertainty (Pearl & Lidz 2009).



DirSynAmb: Ambiguous about antecedent category (*bottle* = N^0, N').
"Look – a bottle! Oh, look – another *one*!"

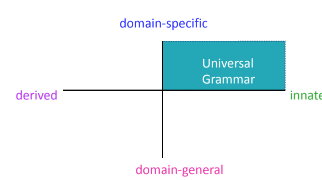
Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Strategy components



Successful DirFiltered proposal

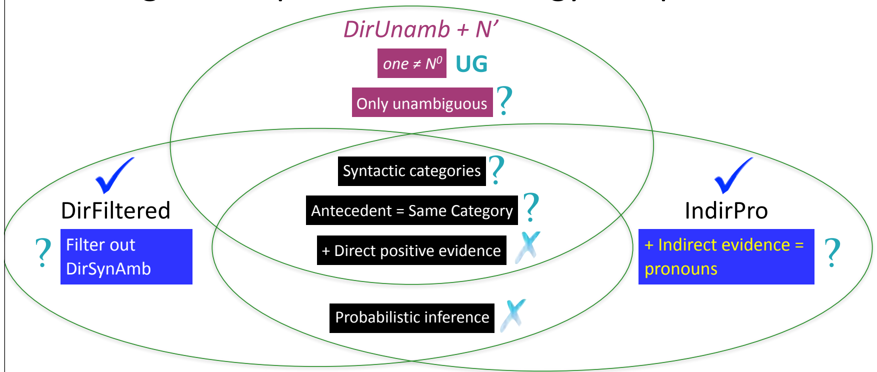
For the domain of language, uncertainty in communication would be what matters. Utterances where only the syntactic category is uncertain may be "good enough" for communication purposes since the referent is clear. So, children are unconcerned about improving linguistic knowledge about these utterances and ignore them.



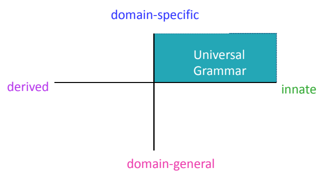
DirSynAmb: Ambiguous about antecedent category (*bottle* = N^0, N').
"Look – a bottle! Oh, look – another *one*!"

Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Strategy components



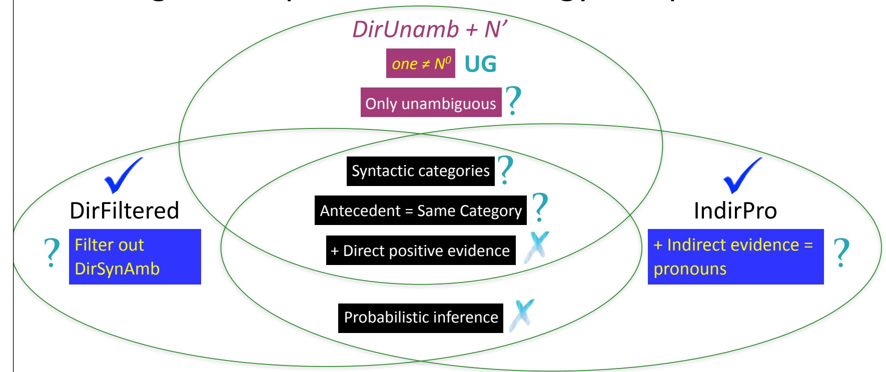
Successful IndirPro proposal



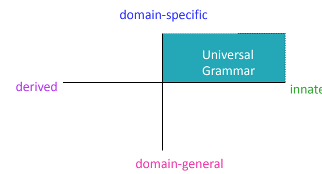
The domain-specific knowledge to consider other pronouns relevant may be innate and so part of UG or it may derive from an overhypothesis (Kemp et al 2007) the learner forms about the similarity of *one* with other anaphoric pronouns in terms of their distribution.

Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Strategy components



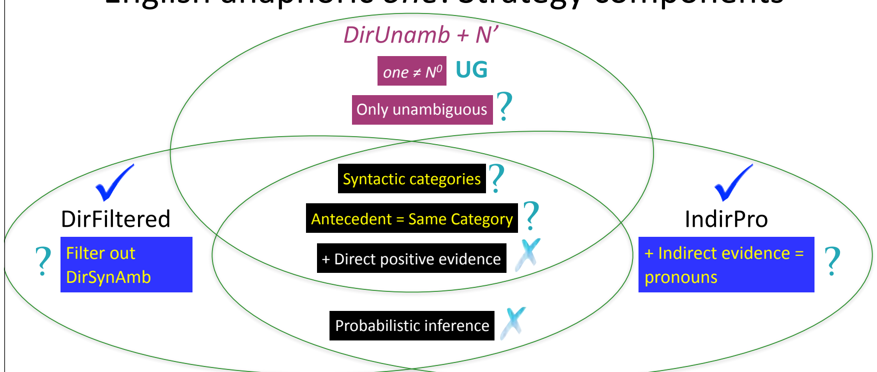
Both successful proposals



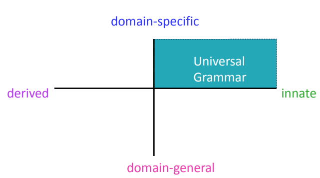
The new components required may not necessarily need to be built into UG. However, if they are, they are less-specific knowledge than the previous proposal supposed (which didn't actually capture children's behavior anyway).

Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Strategy components



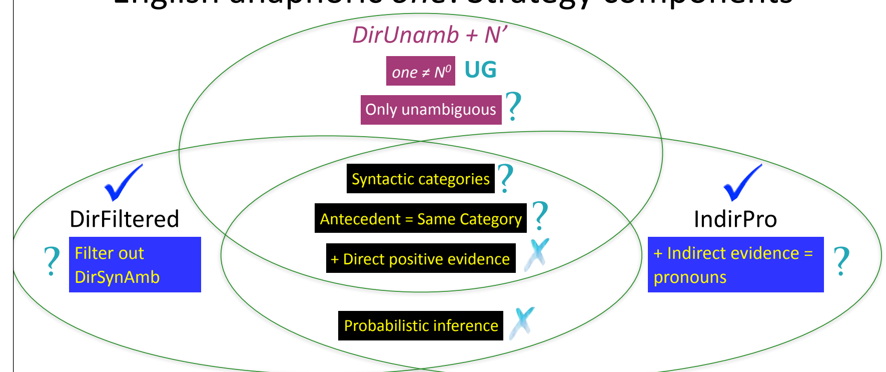
Some open questions



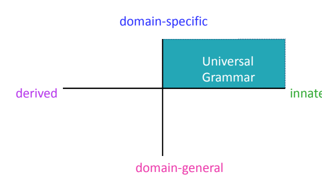
For each component that may be derivable from the input, can we create a learner than can actually derive that component from the available linguistic information? And if so, what are the learning components required to do so?

Pearl & Mis 2011, Pearl & Mis 2016

English anaphoric *one*: Strategy components



Some open questions



How general-purpose are these learning components? Are the components we find useful for making syntactic generalizations about anaphoric *one* useful for making other syntactic generalizations? What about other linguistic generalizations? Or other non-linguistic generalizations?

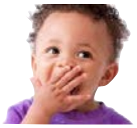
Pearl & Mis 2011, Pearl & Mis 2016

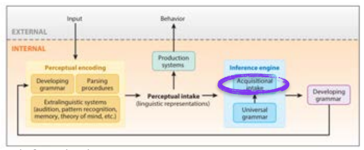
Recurring themes: English anaphoric *one*

Informing theories of representation & acquisition

Recurring themes:

(1) Broadening the set of relevant data in the acquisitional intake to include all pronouns





Lidz & Gagliardi 2015

Pearl & Mis 2011, Pearl & Mis 2016

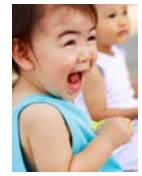
Recurring themes: English anaphoric *one*

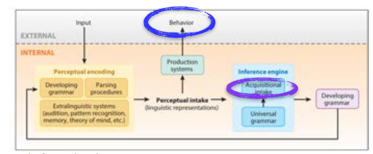
Informing theories of representation & acquisition

Recurring themes:

(1) Broadening the set of relevant data in the acquisitional intake to include all pronouns

(2) Evaluating output by how useful it is for generating toddler looking time behavior





Lidz & Gagliardi 2015

Pearl & Mis 2011, Pearl & Mis 2016

Recurring themes: English anaphoric *one*

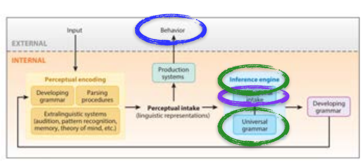
Informing theories of representation & acquisition

Recurring themes:

(1) Broadening the set of relevant data in the acquisitional intake to include all pronouns

(2) Evaluating output by how useful it is for generating toddler looking time behavior

(3) Not necessarily needing the prior knowledge we thought we did in UG: "good enough" derived data filter or derived overhypothesis about pronouns rather than specific knowledge about syntactic category



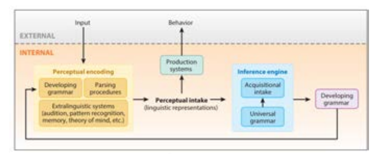
Lidz & Gagliardi 2015

Pearl & Mis 2011, Pearl & Mis 2016

Big picture: Understanding how children acquire syntactic knowledge

If we precisely define the components of any acquisition task by drawing on the insights from different methodologies, we can make progress on how children solve that acquisition task.

In particular, we can understand the nature of children's language acquisition toolkit — what fundamental building blocks they use are, and what is (or is not) part of Universal Grammar.



Lidz & Gagliardi 2015

Computational methods

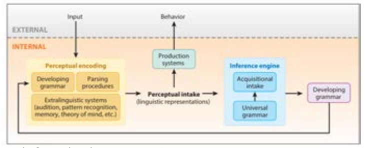
Experimental methods

Theoretical methods

Big picture: Understanding how children acquire syntactic knowledge

If we precisely define the components of any acquisition task by drawing on the insights from different methodologies, we can make progress on how children solve that acquisition task.

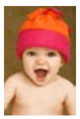
In particular, we can understand the nature of children's language acquisition toolkit — what fundamental building blocks they use are, and what is (or is not) part of Universal Grammar.



Lidz & Gagliardi 2015

Computational methods

Experimental methods



Theoretical methods

This technique is a useful tool — so let's use it to inform our theories of representation and acquisition!

Thank you!

Jon Sprouse Benjamin Mis

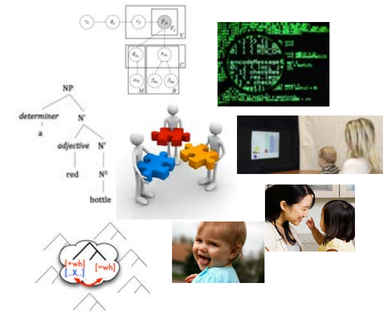
Greg Carlson LouAnn Gerken Jeff Lidz

Computational Models of Language Learning seminar, UC Irvine 2010

Audiences at: CogSci 2011, UChicago 2011 workshops on Language, Cognition, and Computation & Language, Variation, and Change, Input & Syntactic Acquisition Workshop 2012, UMaryland Mayfest 2012, New York University Linguistics colloquium 2012, Stanford Cognition & Language Workshop 2013, GALANA 2015

Computation of Language Laboratory
UC Irvine

Lisa S. Pearl
Associate Professor
Department of Cognitive Sciences
SBSG 2314
University of California, Irvine
lpearl@uci.edu



This work was supported in part by NSF grants BCS-0843896 and BCS-1347028.