Psych229: Language Acquisition

Lecture 17
Poverty of the Stimulus

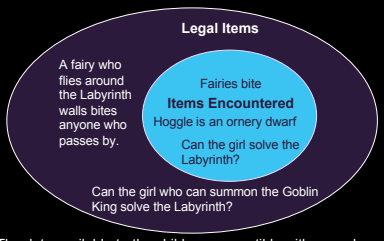
Poverty of the Stimulus

Language

Can be thought of as the set of legal items in the language (sentences, strings, etc.). The child's job: figure out the rules that generate that legal set and don't generate illegal items.

Poverty of the Stimulus

The argument for having innate biases to guide language learning



Idea: The data available to the child are compatible with a number of generalizations. However, children only seem to pick the right ones. Therefore, they must have some other constraints guiding their language learning.

The innate part: The guiding information must be available prior to learning.

Poverty of the Stimulus: A Famous Example

Jareth can alter time.
Can Jareth alter time?

Rule?

Poverty of the Stimulus: A Famous Example

Jareth can alter time.
Can Jareth alter time?

Rule: Move first auxiliary?

Poverty of the Stimulus: A Famous Example

Jareth can alter time.
Can Jareth alter time?

Rule: Move first auxiliary?

Anyone who can wish away their brother might be tempted to do it.
Might anyone who can wish away their brother be tempted to do it?

Poverty of the Stimulus: A Famous Example

Jareth can alter time.
Can Jareth alter time?

Rule: Move first auxiliary?

Rule?

Anyone who can wish away their brother might be tempted to do it.
Might anyone who can wish away their brother be tempted to do it?

Poverty of the Stimulus: A Famous Example

Jareth can alter time.
Can Jareth alter time?

Rule: Move first auxiliary?

Rule: Move second auxiliary?

Anyone who can wish away their brother might be tempted to do it.
Might anyone who can wish away their brother be tempted to do it?

Poverty of the Stimulus: A Famous Example

Jareth can alter time.
Can Jareth alter time?

Rule: Move first auxiliary?

Rule: Move second auxiliary?

Anyone who can wish away their brother might be tempted to do it.
Might anyone who can wish away their brother be tempted to do it?

That anyone who can wish away their brother might be tempted to do it is up for debate.
Is that anyone who can wish away their brother might be tempted to do it up for debate?

Poverty of the Stimulus: A Famous Example

Jareth can alter time.
Can Jareth alter time?

Rule: Move first auxiliary?

Rule: Move second auxiliary?

Anyone who can wish away their brother might be tempted to do it.
Might anyone who can wish away their brother be tempted to do it?

Rule?

That anyone who can wish away their brother might be tempted to do it is up for debate.
Is that anyone who can wish away their brother might be tempted to do it up for debate?

Poverty of the Stimulus: A Famous Example

Jareth can alter time.
Can Jareth alter time?

Rule: Move first auxiliary?

Rule: Move second auxiliary?

Anyone who can wish away their brother might be tempted to do it.
Might anyone who can wish away their brother be tempted to do it?

Rule: Move last auxiliary?

That anyone who can wish away their brother might be tempted to do it is up for debate.
Is that anyone who can wish away their brother might be tempted to do it up for debate?

Poverty of the Stimulus: A Famous Example

Jareth can alter time.
Can Jareth alter time?

Rule: Move first auxiliary?

Rule: Move second auxiliary?

Anyone who can wish away their brother might be tempted to do it.
Might anyone who can wish away their brother be tempted to do it?

Rule: Move last auxiliary?

That anyone who can wish away their brother might be tempted to do it is up for debate.
Is that anyone who can wish away their brother might be tempted to do it up for debate?

Rule???

Someone who is not easily fooled might trick someone who is.
Might someone who is not easily fooled trick someone who is?

Poverty of the Stimulus: A Famous Example

Jareth can alter time.
Can Jareth alter time?

Idea: Look at structure, not just linear order

Anyone who can wish away their brother might be tempted to do it.
Might anyone who can wish away their brother be tempted to do it?

That anyone who can wish away their brother might be tempted to do it is up for debate.
Is that anyone who can wish away their brother might be tempted to do it up for debate?

Someone who is not easily fooled might trick someone who is.
Might someone who is not easily fooled trick someone who is?

Poverty of the Stimulus: A Famous Example

Jareth can alter time.
Can Jareth alter time?

Idea: Look at structure, not just linear order

Anyone might be tempted to do it.
Might anyone be tempted to do it?

Rule: Move main clause auxiliary

That is up for debate.
Is that up for debate?

Someone might trick someone who is.
Might someone trick someone who is?

Poverty of the Stimulus: A Famous Example

Jareth can alter time.
Can Jareth alter time?

Idea: Look at structure, not just linear order

Anyone might be tempted to do it.
Might anyone be tempted to do it?

Rule: Move main clause auxiliary

That is up for debate.
Is that up for debate?

Learning bias: try structure-dependent rules

Someone might trick someone who is.
Might someone trick someone who is?

Poverty of the Stimulus: Data

Induction Problem: Logical Problem of Language Acquisition

Children don't usually get access to all the data we just saw by the time they have the correct generalization (move main clause auxiliary). They learn from a subset of the legal items in the language. And *still* they seem to converge on the right generalizations…without trying out the wrong ones.

Pinker (2004)

Clarifying the Logical Problem of Language Acquisition

It is not the belief that the input is too inconsistent to acquire language. (Obviously not, because kids do acquire language.)

It is a question of *how* children make the right generalizations from the data available.

For a learner to do this, there must be prior constraints that are being obeyed.
Connectionists: features defining units and topology of neural net

Chomskyans: categories, operations, principles (priors over grammars)

Emergentists: cues, items, competition, indirect negative evidence

Pinker (2004)

Clarifying the Logical Problem of Language Acquisition

It is not the belief that there is no negative evidence, indirect or otherwise. That's just a description of the data available.

However, it is important to document that exact nature of the data available. How often do highly informative (unambiguous) data appear? How often do less informative (ambiguous) data appear?

It is not a belief that children don't learn.

Ex: For yes/no question formation, children must learn *which* structure-dependent rule is appropriate.

Bias guiding learning (argued to be innate): Don't try to posit structure-independent rules.

Pinker (2004)

Clarifying the Logical Problem of Language Acquisition

It is not saying that there is no role for probabilistic learning.

Probabilistic learning (like Bayesian learning) is a method for updating beliefs about the hypothesis space, given the available data. But the child needs to *have* a defined hypothesis space.

Innate/prior bias: What hypotheses should the child consider?
Ex: Structure-dependent rules for question formation

Innate/prior bias: How should the child use the data available?
Ex: Use only highly informative data, ignore noisy data

It is not saying there is no role for generalization.

Instead: why do children generalize along some dimensions (past tense +ed), and not others?

An example where kids don't generalize

Crain & McKee (1985)

While he danced around the throne room, Jareth smiled.
(he = Jareth)

Jareth smiled while he danced around the throne room.
(he = Jareth)

An example where kids don't generalize

Crain & McKee (1985)

While he danced around the throne room, Jareth smiled.
(he = Jareth)

Jareth smiled while he danced around the throne room.
(he = Jareth)

While Jareth danced around the throne room, he smiled.
(he = Jareth)

He smiled while Jareth danced around the throne room.
(he ≠ Jareth)

Idea: Constraint on Interpretation with pronouns

Gerken (2006): Making Generalizations - Experimental Evidence

Generalizations from artificial language data

Previous work in artificial languages: when children are familiarized in the laboratory for a short period of time, they can extract generalizations (Chambers et al. 2003, Gerken 2004, Gómez 2002, Gómez & Gerken 1999, Gómez & Lakusta 2004, Marcus et al. 1999, Maye et al. 2002, Maye & Weiss 2003, Saffran & Thiessen 2003, Saffran et al 1996)

What signals generalizations: previous work

Gómez 2002: 18-month olds only track & generalize non-adjacent dependencies (AXB, CXD) when the intervening item is highly variable.

Gerken, Wilson, & Lewis 2005: 17-month olds can generalize Russian noun inflectional pattern only if a subset of the data also have additional cues (markings for gender)

Gerken (2006): Making Generalizations - Experimental Evidence

But what happens when multiple generalizations are possible - specifically, one that is less general and one that is more general?

Option 1: Children can make both generalizations.
Option 2: Children can't make either generalization.
Option 3: Children generalize one way or the other, based on the available data.

Discussion:
How does this relate to the logical problem of language acquisition? Is choosing between a less-general and more-general generalization a reasonable depiction of the problem?

Gerken (2006): Making Generalizations - Experimental Evidence

Data & generalizations (Marcus et al. 1999): AAB pattern

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

Infants: trained on AAB (or ABA) pattern, learned AAB (or ABA).
Note that pattern also consistent with "ends with {CV}", ex: di.

Gerken (2006): Making Generalizations - Experimental Evidence

Data & generalizations: Gerken 2006

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

More specific generalization: "ends in di"
More general generalization: "AAB pattern"

Gerken (2006): Making Generalizations - Experimental Evidence

Experiment 1 (Head Turn Preference Procedure)
If given data are consistent with more specific generalization and more general generalization, do they pick the more general generalization?

Control: If given data are consistent with more general generalization only, do they pick the more general generalization?

leledi…wiwidi…jijidi…dededi

leledi…wiwije…jijili…dedewe

Stimuli: 2 minutes of 3 syllable words from table.

Test condition words: novel AAB/ABA pattern words
Ex: kokoba in AAB condition (novel syllables: ko, ba)

Subjects: 9-month olds, 16 in more specific & 16 in more general condition

Gerken (2006): Making Generalizations - Experimental Evidence

Experiment 1 Results

	Exp. 1 Diagonal condition	Exp. 1 Column condition
Consistent	13.51 (1.82)	10.74 (1.56)
Inconsistent	10.14 (1.39)	10.18 (1.60)
Difference in sec. and significance	3.37 ($p<0.05$)	0.56 (n.s.)

Consistent: familiarized to AAB (leledi), heard AAB (kokoba)
Inconsistent: familiarized to ABA (ledile), heard AAB (kokoba)

Diagonal (more general generalization): familiarity preference for abstract pattern
Infants extracted the more general generalization.

Column (more specific generalization): no preference for abstract pattern
Infants did not extract the more general generalization.

What does this mean exactly? What were infants doing when they only heard more specific generalization data?

Gerken (2006): Making Generalizations - Experimental Evidence

Experiment 2 (Head Turn Preference Procedure)
If given data are consistent with more specific generalization, what precisely are they doing? Is there any pattern extraction at all?

leledi…wiwidi…jijidi…dededi

Stimuli: 2 minutes of 3 syllable words from "-di" data.

Test condition words: novel AAdi /AdiA pattern words
Ex: kokodi in AAB condition (novel syllable: ko)

Subjects: 16 9-month olds

Gerken (2006): Making Generalizations - Experimental Evidence

Experiment 2 Results

	Exp. 2
Consistent	9.33 (1.18)
Inconsistent	6.25 (0.64)
Difference in sec. and significance	3.08($p<0.05$)

Consistent: familiarized to AAB (leledi), heard AAB (kokodi)
Inconsistent: familiarized to ABA (ledile), heard AAB (kodiko)

Column (more specific generalization): familiarity preference for abstract pattern with similar final syllable "di"
Infants extracted the more specific generalization (AAdi) consistent with the data.

Gerken (2006): Making Generalizations - Experimental Evidence

Overall results
Infants make a more specific generalization (AAdi) when the data is consistent with both the more specific and the more general one (AAB).

When the data are consistent only with the more general generalization, children do make that generalization.

Suggests a conservative learning approach (no unnecessary abstraction).

leledi…wiwidi…jijidi…dededi

AAdi → memedi kokodi nanadi…

AAB → memedi kokodi nanadi…
memeli kokoli nanali…
memewe kokowe nanawe…
…

Gerken (2006): Making Generalizations - Experimental Evidence

How to formalize "conservative learning approach"

Similar in spirit to Subset Principle (Manzini & Wexler 1987), which guides the child to pick the generalization that generates the smallest language.

leledi…wiwidi…jijidi…dededi

Smaller language

AAdi → memedi kokodi nanadi…

AAB → memedi kokodi nanadi…
memeli kokoli nanali…
memewe kokowe nanawe…
…

Can be formalized mathematically by Size Principle in Bayesian learning (Tenenbaum & Griffiths 2001) with two hypotheses in a subset-superset relationship.

The point: subset data points are ambiguous

Side note: Only works if learner already has hypotheses in mind to consider

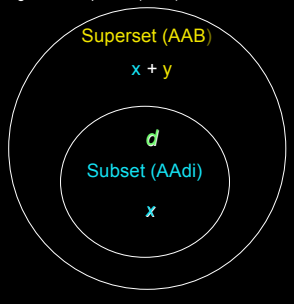
Size Principle

Two ways to describe size principle logic:
One way: Likelihood of given ambiguous data point *d* (leledi)

Suppose the learner encounters an ambiguous data point ***d***

Let the number of examples covered by subset be ***x***.

Let the number of examples covered by superset be ***x*** + ***y***.



Size Principle

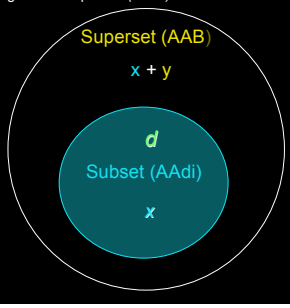
Two ways to describe size principle logic:
One way: Likelihood of given ambiguous data point *d* (leledi)

The likelihood that *d* was produced from **subset** is **1/*x***

The likelihood that *d* was produced from **superset** is **1/(*x*+*y*)**

Since x+y > x, 1/x+y < 1/x.

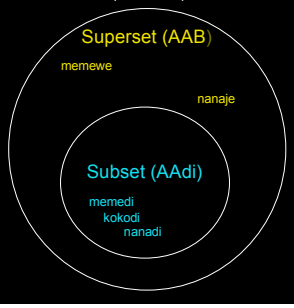
So, subset has a higher probability of having produced *d*. Thus, **subset is favored** when encountering ambiguous data.



Size Principle

Two ways to describe size principle logic:
Another way: Learner's expectation of set of data points in input

If superset were correct, learner should encounter some **unambiguous data points for superset**, which cannot be accounted for by the subset.



Size Principle

Two ways to describe size principle logic:
Another way: Learner's expectation of set of data points in input

If the learner keeps not encountering data compatible only with the superset, the subset becomes more likely to be the hypothesis generating the data.