Psych 245A:
Computational Models of Language Learning

Lecture 1
Introduction to the Linguist's Viewpoint

Knowledge of Language

It's so natural for us to produce and comprehend language that we often don't think about what an accomplishment this is.

Or how we learned language in the first place.

About Language

Language is a complex system of knowledge: includes sound structure, word structure, sentence structure, mapping from sentence structure to meaning, unspoken rules of conversation…

Languages can differ significantly on how they instantiate this knowledge.

And despite all this complexity, children of all languages acquire the necessary knowledge to speak their native language.

Don't goblins like children?

Goblins like children.

goblins

goblin (plural) = goblin + s

gob lins

g a b l ɪ n z

Jackendoff (1994)

"For the moment, the main thing is to appreciate how hard a problem this is. The fact that we can talk (and cats can't) seems so obvious that it hardly bears mention. But just because it's obvious doesn't mean its easy to explain."

Kids Do Amazing Things

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

…when kids can't tie their own shoes or even count to 4.

What kids are doing: extracting patterns and making generalizations from noisy data sets without explicit instruction.

"Rules" of language = grammar

Knowledge of Language & Hidden Rules

Some examples from language:

You know that…

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

Knowledge of Language & Hidden Rules

Some examples from language:

You know that…

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

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

Knowledge of Language & Hidden Rules

Some examples from language:

You know that…

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

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

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

Knowledge of Language & Hidden Rules

Some examples from language:

You know that…

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

cats: 's' = /s/

goblins: 's' = /z/

Why rules?

"The expressive variety of language use implies that a language user's brain contains unconscious grammatical principles" - Jackendoff (1994)

Example: Most sentences we have never seen or used before, but we can still understand them.

Question: Can speakers simply memorize all the possible sentences of a language the way they learn vocabulary of their language? Not if there are an infinite number of them.

Possible objections to
a mental rule set

"Why should I believe I store a set of rules unconsciously in my mind? I just understand sentences because they make sense."

Possible objections to
a mental rule set

"Why should I believe I store a set of rules unconsciously in my mind? I just understand sentences because they make sense."

But why do some sentences make sense and others don't?

Hoggle has two jewels.
*Two Hoggle jewels has.

Possible objections to a mental rule set

"Why should I believe I store a set of rules unconsciously in my mind? I just understand sentences because they make sense."

But why do some sentences make sense and others don't?

Hoggle has two jewels.
*Two Hoggle jewels has.

Why can we recognize patterns even when some of the words are unknown?

'Twas brillig, and the slithy toves
did gyre and gimble in the wabe...

Possible objections to an unconscious rule set

"When I talk, the talk just comes out - I'm not consulting any rule set."

Possible objections to an unconscious rule set

"When I talk, the talk just comes out - I'm not consulting any rule set."

Analogy: wiggling your fingers

When you want to wiggle your fingers, you "just wiggle them".

But your finger-wiggling intention was turned into commands sent by your brain to your muscles, and you're never conscious of the process unless something interferes with it. Nonetheless, there *is* a process, even if you're not aware of it.

Children

Children's Mistakes

(From Martin Braine)

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

Children's Mistakes

(From Martin Braine)

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

An important point for learning: Corrective feedback not always heeded.

Children's Mistakes

From Edward Klima & Ursula Bellugi
Wh-questions

Stage 1
What book name?
Why you smiling?
What soldier marching?

Stage 2
What he can ride in?
Which way they should go?
Why kitty can't stand up?

Stage 3
Where will you go?
Why can't kitty see?
Why don't you know?

Children's Mistakes

From Edward Klima & Ursula Bellugi
Use of negative elements (not, n't)

Stage 1
No the sun shining.
No a boy bed.
No sit there.

Stage 2
He no bite you.
I no want envelope.
I no taste them.

Stage 3
I didn't did it.
You didn't caught me.

Children's Mistakes

From Edward Klima & Ursula Bellugi

Use of past tense verbs
(U-shaped curve of performance)

Stage 1
walked
played
came
went

Stage 2
walked
played
comed
goed
holded

Stage 3
walked
played
camed
wented

Stage 4
walked
played
came
went
held

Children's Mistakes

From Edward Klima & Ursula Bellugi

Use of past tense verbs
(U-shaped curve of performance)

Stage 1
walked
played
came
went

Stage 2
walked
played
comed
goed
holded

Stage 3
walked
played
camed
wented

Stage 4
walked
played
came
went
held

Main points

Children don't just imitate what they've heard - they're trying to figure out the patterns of their native language.

The patterns they produce during learning are often stripped-down versions of the adult pattern, but they make mistakes that cannot be attributed directly to the input.

Standard Theory

The Standard Theory, according to Chomsky
(summarized by Seidenberg (1997))

Big Questions of Language Acquisition:

What constitutes knowledge of language?

How is this knowledge acquired?

How is this knowledge used?

Standard Theory

Knowledge of language, according to Chomsky

Knowledge of language = grammar

Grammar = complex set of rules and constraints that gives speakers intuitions that some sentences belong in the language while others do not

Competence Hypothesis: Grammar is separate from "performance factors", like dysfluencies (she said…um..wrote that), errors (I bringed it), memory capacity (The boy that the dog that the cat chased bit ran home.), and statistical properties of language (frequency of transitive (Sarah ate the peach) vs. intransitive use (Sarah ate))

"I think we are forced to conclude that…probabilistic models give no particular insight into some of the basic problems of syntactic structure" - Chomsky, 1957

Standard Theory

Properties of language, according to Chomsky

Grammar is generative: it can be used to produce and comprehend an infinite number of sentences

Grammar involves abstract structures: information that speakers unconsciously used is not overtly available in the observable data

Grammar is modular: there are separate components with different types of representations governed by different principles

Grammar is domain-specific: language exhibits properties not seen in other areas of cognition, so it cannot be the product of our general ability to think and learn

Standard Theory

Language acquisition, according to Chomsky

How does a child acquire a grammar that has those properties (generative, involving abstract structures, modular, domain-specific)?

Poverty of the stimulus problem

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.

Illegal Items

Legal items

Fairies bite

Hoggle is an ornery dwarf

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

Hoggle a dwarf ornery is

Bite fairies

Can the girl who can summon the Goblin King can solve the Labyrinth?
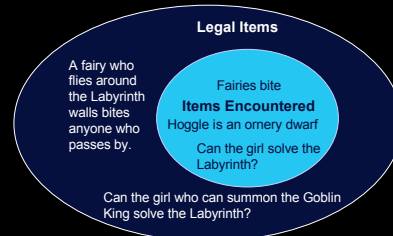
Poverty of the Stimulus

The Logic of Poverty of the Stimulus (The Logical Problem of Language Acquisition)

1) Suppose there is some data.

2) Suppose there is an incorrect hypothesis compatible with the data.

3) Suppose children behave as if they never entertain the incorrect hypothesis.

Conclusion: Children possess innate knowledge ruling out the incorrect hypothesis from the hypothesis space considered.

Poverty of the Stimulus

The argument for having innate biases to guide language learning

Legal Items

A fairy who flies around the Labyrinth walls bites anyone who passes by.

Fairies bite
Items Encountered
Hoggle is an ornery dwarf

Can the girl solve the Labyrinth?

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

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: More Details

Jareth can alter time.
Can Jareth alter time?

Rule?

Poverty of the Stimulus: More Details

Jareth can alter time.
Can Jareth alter time?

Rule: Move first auxiliary?

Poverty of the Stimulus: More Details

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: More Details

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: More Details

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: More Details

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: More Details

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: More Details

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: More Details

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 could trick someone who is.
Could someone who is not easily fooled trick someone who is?

Poverty of the Stimulus: More Details

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 could trick someone who is.
Could someone who is not easily fooled trick someone who is?

Poverty of the Stimulus: More Details

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 could trick someone who is.
Could someone trick someone who is?

Poverty of the Stimulus: More Details

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

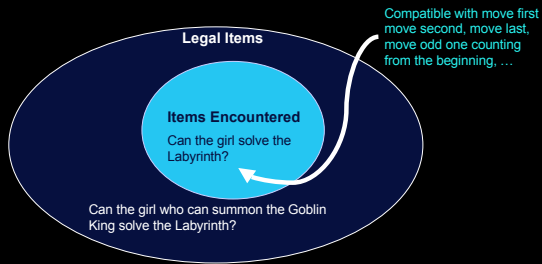That is up for debate.
Is that up for debate?

Learning bias: try structure-dependent rules

Someone could trick someone who is.
Could 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.

Compatible with move first move second, move last, move odd one counting from the beginning, …

Legal Items

Items Encountered
Can the girl solve the Labyrinth?

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

Poverty of the Stimulus: Real Data

Child Input

Very frequent
Is Hoggle t_{is} running away from Jareth?

Very infrequent, if ever
Can someone who can solve the Labyrinth t_{can} show someone who can't how?

Hypotheses for frequent data type
Structure-independent (linear)
Front first auxiliary, Front last auxiliary, …

Structure-independent (hierarchical)
Front the first auxiliary following the first noun phrase, Front the first auxiliary preceding a verb phrase, …

Structure-independent (creative)
Front the auxiliary closest to a noun, Front the auxiliary that is an odd-numbered position, …

Poverty of the Stimulus: Real Data

The Real Rule

Front the auxiliary following the subject noun phrase in the main clause.

But the unbiased child has to rule out all the other options, even ones that are simpler to compute. (For instance: front first auxiliary is much easier to compute.) We would expect to see errors of this type.

Is the dwarf who t_{is} talking to Jareth is going to give Sarah the peach?

Real Children

But kids don't seem to make this error (Crain & Nakayama, 1987).

Implication: They've already ruled out that hypothesis, even though they've likely not seen much data (if any at all) incompatible with it. This is due to an innate bias to look for structure-dependent rules.

Poverty of the Stimulus: Real Data

A larger point about data-driven learning

Problem: "…wild statistical disparities between what is presented to children and how children actually learn" - Legate & Yang (2002)

Example: Subject-drop (lots of "data", late generalization)
"Solves the Labyrinth" instead of "She solves the Labyrinth"
Almost all English sentences contain a subject, but children don't get it till 3.

Example: Verb-Raising in French (little "data", early generalization)
"She eats not the peach"
Only 7% of French sentences show this, but children acquire it by 1.5 years.

The point: Children come with innate biases that allow them to use data in specific ways to update their hypotheses.

Some Clarifications

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

Standard Theory

Language acquisition, according to Chomsky

How does a child acquire a grammar that has those properties (generative, involving abstract structures, modular, domain-specific)?

Poverty of the stimulus problem: Available data insufficient to determine all these properties of the grammar. Therefore, children must bring innate knowledge to the language learning problem that guides them to the correct instantiation of grammar.

Available data properties leading to this inductive problem:
  noisy (degenerate): sometimes there are incorrect examples in the input
  variable: no child's input is the same as another's, but all converge
  no reliable negative evidence: no labeled examples of what's not in the language
  no positive evidence for some generalizations: yet children still converge on them

Specificity of Innate Knowledge

"Innate capacities may take the form of biases or sensitivities toward particular types of information inherent in environmental events such as language, rather than a priori knowledge of grammar itself." - Seidenberg (1997)

Levels of Representation
Marr (1982)

On Explaining (Marr 1982)

"…[need] a clear understanding of what is to be computed, how it is to be done, the physical assumptions on which the method is based, and some kind of analysis of the algorithms that are capable of carrying it out."

"This was what was missing - the analysis of the problem as an information-processing task. Such analysis does not usurp an understanding at the other levels - of neurons or of computer programs - but it is a necessary complement to them, since without it there can be no real understanding of the function of all those neurons."

The three levels

Computational
What is the goal of the computation? What is the logic of the strategy by which is can be carried out?

Algorithmic
How can this computational theory be implemented? What is the representation for the input and output, and what is the algorithm for the transformation?

Implementational
How can the representation and algorithm be realized physically?

The three levels:
An example with the cash register

Computational
What does this device do?
Arithmetic.

Task: Master theory of addition.

The three levels:
An example with the cash register

Computational
What does this device do?
Arithmetic.

Task: Master theory of addition.

Algorithmic (Addition)
Addition: Mapping of a pair of numbers to another number.

$$(3,4) \longrightarrow 7 \qquad \text{(often written } (3+4=7))$$
Properties: (3+4) = (4+3) [commutative], (3+4)+5 = 3+(4+5) [associative], (3+0) = 3 [identity element], (3+ -3) = 0 [inverse element]

True no matter how numbers are represented: this is what is being computed

The three levels:
An example with the cash register

Computational
  What does this device do?
      Arithmetic.

  Task: Master theory of addition.

Algorithmic (Addition)
  Addition: Mapping of a pair of numbers to another number.

            (3,4) ⟶ 7   (often written (3+4=7))
Properties: (3+4) = (4+3) [commutative], (3+4)+5 = 3+(4+5) [associative],
(3+0) = 3 [identity element], (3+ -3) = 0 [inverse element]

True no matter how numbers are represented: this is what is being computed

Implementational
  How does cash register implement this? A series of mechanical and
electronic components.

Mapping the Framework:
Algorithmic Theory of Language Learning

Goal: Understanding the "how" of language learning

First, we need a computational-level description of the learning problem.

Computational Problem: Divide sounds into contrastive categories
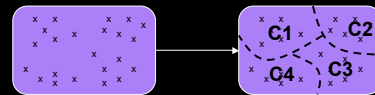
---

Mapping the Framework:
Algorithmic Theory of Language Learning

Goal: Understanding the "how" of language learning

First, we need a computational-level description of the learning problem.

Computational Problem: Divide spoken speech into words

        húwzəfréjdəvðəbÍgbæˈdwəˈlf

     húwz   əfréjd əv ðə   bÍg   bæˈd wəˈlf
     who's afraid of the big   bad   wolf

Mapping the Framework:
Algorithmic Theory of Language Learning

Goal: Understanding the "how" of language learning

First, we need a computational-level description of the learning problem.

Computational Problem: Map word forms to speaker-invariant forms

fwiends

friends

friends

"friends"

Mapping the Framework:
Algorithmic Theory of Language Learning

Goal: Understanding the "how" of language learning

First, we need a computational-level description of the learning problem.

Computational Problem: Identify grammatical categories

"This is a DAX."

DAX = noun

Mapping the Framework:
Algorithmic Theory of Language Learning

Goal: Understanding the "how" of language learning

First, we need a computational-level description of the learning problem.

Computational Problem: Identify word affixes that signal meaning.

What do you have to change about the verb to signal the past tense in English? (There are both regular and irregular patterns.)

blink~blinked   confide~confided
blɪŋk blɪŋkt   kənfajd kənfajdəd

drink~drank
drɪŋk drejŋk

Mapping the Framework:
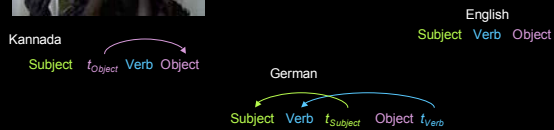## Algorithmic Theory of Language Learning

Goal: Understanding the "how" of language learning

First, we need a computational-level description of the learning problem.

Computational Problem: Identify the rules of word order for sentences.

Jareth   juggles   crystals
Subject   Verb   Object

English
Subject Verb Object

Kannada
Subject t_{Object} Verb Object

German
Subject Verb $t_{Subject}$ Object t_{Verb}

Mapping the Framework:
Algorithmic Theory of Language Learning

Goal: Understanding the "how" of language learning

First, we need a computational-level description of the learning problem.

Second, we need to be able to identify the algorithmic-level description:

Input = sounds, syllables, words, phrases, …
Output = sound categories, words, words with affixes, grammatical categories, …

Process the can take us from input to output: statistical learning, algebraic learning,…?

Considerations: input available to child, psychological plausibility of learning algorithm, hypotheses child considers

Framework for language learning
(algorithmic-level)

What are the hypotheses available (for generating the output from the input)?
  Ex: general word order patterns

Input: words (adjective and noun)
Output: ordered pair

  Adjective before noun (ex: English)
  red apple

  Noun before adjective (ex: Spanish)
  manzana roja
  apple   red

Framework for language learning
(algorithmic-level)

What are the hypotheses available (for generating the output from the input)?
  Ex: general word order patterns

What data is available, and should the learner use all of it?
  Ex: exceptions to general word order patterns

  Ignore special use of adjective before noun in Spanish
  Special use: If the adjective is naturally associated with the noun:
  la blanca nieve
  the white snow

  Why not usual order? Snow is naturally white

Framework for language learning
(algorithmic-level)

What are the hypotheses available (for generating the output from the input)?
  Ex: general word order patterns

What data is available, and should the learner use all of it?
  Ex: exceptions to general word order patterns

How will the learner update beliefs in the competing hypotheses?
  Ex: shifting belief in what the regular word order of adjectives and nouns should be

This usually will involve some kind of probabilistic updating function.