Psych 229: Language Acquisition

Lecture 1

Introduction to Language Acquisition

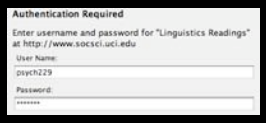
Administrivia

Class web page:

http://www.socsci.uci.edu/~lpearl/courses/psych229_2008win/index.html

Accessible from EEE and my home page, as well. Contains overview, schedule, readings, course assignments, and grading policies.

Important to access readings
user name = psych229
user password = langacq



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.

Jackendoff (1994)

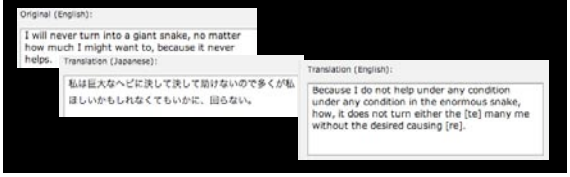
For the moment, the main thing is to appreciate how hard a problem this is. The fact that we can talk (and cats can't) seems so obvious that it hardly bears mention. But just because it's obvious doesn't mean it's easy to explain. Think of another perfectly obvious, well-known phenomenon: the fact that metals turn red when you heat them enough. Why does this happen? It could be otherwise—they might just as well turn green or not change color at all. It's a simple phenomenon, easily observable, but the explanation isn't simple at all. It turns out to involve at the very least the theories of electromagnetic radiation and quantum mechanics, two of the more amazing intellectual advances of the past century. So it is, I want to suggest, with the human ability to use language.

So About That Universal Translator…

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.

Automatic translation attempts (when structural differences strike!)
(using http://www1.worldlingo.com/en/products_services/worldlingo_translator.html)



Kids Do Amazing Things

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

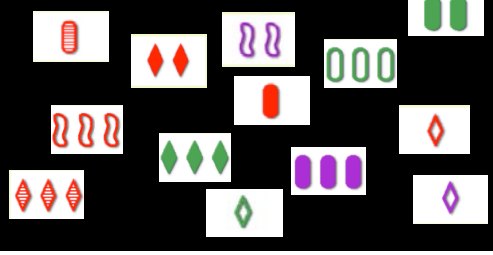
…when kids can't tie their own shoes or even count to 4.

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

"Rules" of language = grammar

A learning analogy: Set

Here are some cards - they have some salient properties associated with them.



A learning analogy: Set

Task: Find Sets.

Here's one:

What generalizations might you make about Sets?

A learning analogy: Set

Task: Find Sets.

Here's another one:

Does this fit your generalization?

A learning analogy: Set

Task: Find Sets.

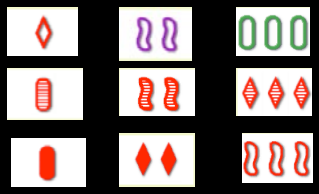
Here's another one:

What about this one?

A learning analogy: Set

Task: Find Sets.

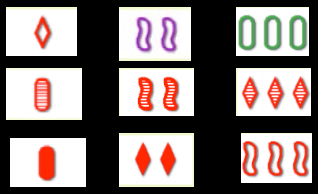
Are these Sets?



A learning analogy: Set

Task: Find Sets.

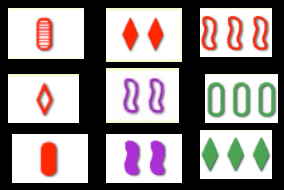
Are these Sets?



A learning analogy: Set

Task: Find Sets.

Here are some more examples:

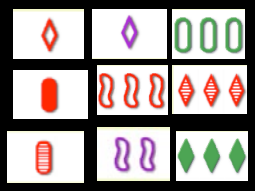


What generalization can you make now?

A learning analogy: Set

Task: Find Sets.

Are these Sets?

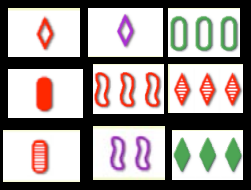


Can you guess the rule of Set?

A learning analogy: Set

Task: Find Sets.

Are these Sets?



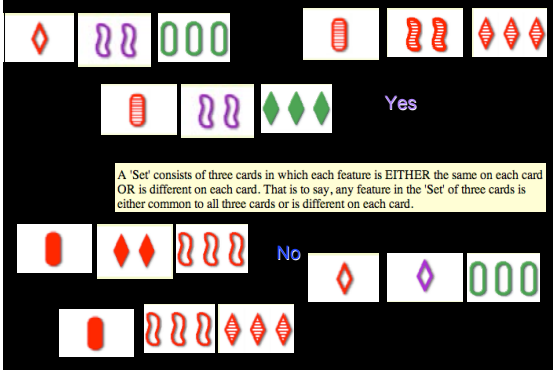
No

No

Yes

Can you guess the rule of Set?

The Grammar of Set



Yes

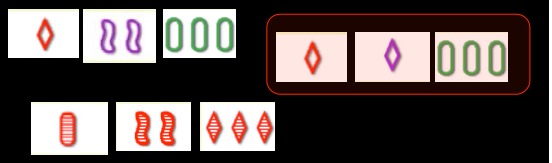
A 'Set' consists of three cards in which each feature is EITHER the same on each card OR is different on each card. That is to say, any feature in the 'Set' of three cards is either common to all three cards or is different on each card.

No

Back to Kids & Language

Children infer rules with this amount of complexity (and more!) from examples of language. And sometimes, even when there's noise.

Noise Analogy: All these are Sets.



Knowledge of Language & Hidden Rules

Some examples from language:

You know that…

…*strep* is a possible word of English, while *stlep* isn't.

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

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

…In "Hoggle has a ripe peach, and Sarah has one, too," one = 'ripe peach'

…In 'cats', the 's' sounds like 's'; in 'dogs', the 's' sounds like 'z'

…If the nonsense word 'pa tih keh' became used in English, it is much more likely to be pronounced "PA tih keh" than "pa tih KEH"

Chomsky's Arguments

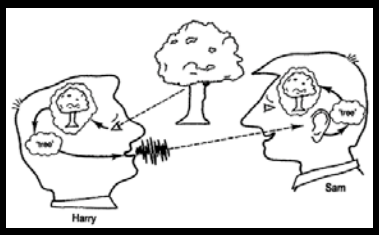
First laid out in late 1950s and early 1960s

The Argument for Mental Grammar:
The expressive variety of language use implies that a language user's brain contains a set of unconscious grammatical principles.

The Argument for Innate Knowledge:
The way children learn to talk implies that the human brain contains a genetically determined specialization for language.

These two arguments lead to conclusion that learning language (English, French, Japanese, Zulu, Mohawk, …) is a complex interaction of nature and nurture

The argument for mental grammar



Harry tells Sam about a tree - this is a fairly involved process.

The argument for mental grammar

Other things Harry might say:

a There's a bird in the tree.
b A bird was in the tree yesterday.
c Are there any birds in that tree?
d A bird might be in the tree.
e Birds like that tree.
f That tree looks like a bird.

These show off the expressive variety of language. (This differs from animal communication.)

The argument for mental grammar

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

- most sentences we have never seen or used before, but we can still understand them
- Can speakers simply memorize all the possible sentences of a language the way they learn vocabulary of their language?

Linguistic Infinity

The number of sentences we are capable of using is just too large to store them individually. Let me run up the number in some rather stupid ways, just as a sample. Consider this series of sentences, all of which are perfectly comprehensible.

(2) Amy ate two peanuts.
Amy ate three peanuts.
Amy ate four peanuts.
. . .
Amy ate forty-three million, five hundred nine peanuts.
. . .
. . .

There are as many sentences in this series as there are nameable integers. The biggest number name listed in my Webster's Collegiate is a vigintillion (10^{63} in US/French usage; 10^{120} in British/German usage). With all the numbers up to this at our disposal, we can create more sentences in this series than there are elementary particles in the universe.

Linguistic Infinity

Here's another way to make lots of sentences. There are at least some tens of thousands of nouns in English. Let's be conservative and say we know ten thousand (10^4). Now let's construct all the sentences we can by putting in different nouns for X and Y in "An X is not a Y." Here are some of them.

(3) A numeral is not a numbskull.
A numeral is not a nun.
A numeral is not a nunnery.
. . .
A numbskull is not a numeral.
A numbskull is not a nun.
A numbskull is not a nunnery.
. . .
A nun is not a nursery.
. . .
An oboe is not an octopus.
. . .

These are all completely absurd, but they *are* sentences of English nevertheless. There will be something like $10^4 \times 10^4$ of them = 10^8.

Linguistic Infinity

Now let's put pairs of these sentences together with "since," like this:

(4) Since a numeral is not a numbskull, a numbskull is not a nun.
Since a numeral is not a numbskull, a numbskull is not a nunnery.
Since a numeral is not a numbskull, a numbskull is not a nuptial.
. . .
Since a numeral is not a nursery, a numbskull is not a nun.
. . .
Since an oboe is not an octopus, a numeral is not a numbskull.
. . .

And so on it goes, giving us $10^8 \times 10^4 = 10^{16}$ absolutely ridiculous sentences. Given that there are on the order of ten billion (10^{10}) neurons in the entire human brain, this divides out to 10^6, or one million sentences per neuron. Thus it would be impossible for us to store them all in our brains, in the unlikely event that we should ever want to use or understand any of them. But still, you did just understand a sampling of them. And these lists are only a minute proportion of the sentences you can understand. What lists include the sentences of this paragraph, for instance?

Linguistic Infinity

In short, we can't possibly keep in memory all the sentences we are likely to encounter or want to use—not to mention all the unlikely ones such as the sentences in (2)–(4). On the other hand, we are apparently ready to encounter them—we seem to know what the possibilities are.

The way the brain seems to achieve expressive variety is to store not whole sentences, but rather words and their meanings, plus *patterns* into which words can be placed.

Patterns:

Amy ate *n* peanuts.

An X is not a Y.

Since an X is not a Y, a Z is not a W.

Linguistic Infinity

But even using these kinds of fixed patterns isn't quite good enough. Consider the list of sentences in (5).

(5) *a* Bill thinks that Beth is a genius.
b Sue suspects that Bill thinks that Beth is a genius.
c Charlie said that Sue suspects that Bill thinks that Beth is a genius.
d Jean knows that Charlie said that Sue suspects that Bill thinks that Beth is a genius.
. . .

This sequence can be extended as long as we have the patience—that is, it is effectively infinite. (To be more precise, there is no longest sentence in this sequence, because we can always add one more.) As a result, we can't specify a single pattern for this list,the way we could for the lists sampled in (2)–(4). Rather, each sentence has to come from a different pattern, and the patterns get longer and longer. (6) shows the first three of these patterns; the term "Verbs" stands for one of the words "thinks," "suspects," "knows," and so forth.

(6) X Verbs that Y is a Z.
W Verbs that X Verbs that Y is a Z.
T Verbs that W Verbs that X Verbs that Y is a Z.
. . .

Linguistic Infinity

Pattern: X Verbs that [sentence].

This shows recursion because "X Verbs that [sentence]" is itself a sentence.

Sentence --> X Verbs that Sentence

Sentence --> Hoggle thinks that [Sentence]
--> Hoggle thinks that [Sarah has Jareth's attention].
--> Hoggle thinks that [Ludo knows that [Sarah has Jareth's attention]].
--> Hoggle thinks that [Ludo knows that [Didymus suspects that [Sarah has Jareth's attention]]].

Two more examples

(8) *a* Ben's father is a linguist.
b Ben's father's older brother is a linguist.
c Ben's father's older brother's best friend is a linguist.
d Ben's father's older brother's best friend's former lover is a linguist.
. . .

(9) *a* This is the house that Jack built.
b This is the refrigerator that sits in the house that Jack built.
c This is the cheese that fell out of the refrigerator that sits in the house that Jack built.
d This is the mold that grew on the cheese that fell out of the refrigerator that sits in the house that Jack built.
. . .

Noun-Phrase
--> Noun-Phrase's Noun

Sentence
--> This is Noun-Phrase

Noun-Phrase
--> Noun-Phrase that Sentence

The argument for mental grammar

In short, in order for us to be able to speak and understand novel sentences, we have to store in our heads not just the words of our language but also the patterns of sentences possible in our language. These patterns, in turn, describe not just patterns of *words* but also patterns of *patterns*. Linguists refer to these patterns as the *rules* of language stored in memory; they refer to the complete collection of rules as the *mental grammar* of the language, or *grammar* for short.

Note: some people object to this, and believe humans don't abstract this much…or at least don't do it for a lot of things. Instead, there's a more "item-based" approach that is sensitive to the frequency of usage an individual lexical items or constructions.