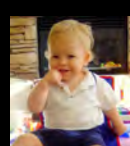
Psych 215L: Language Acquisition

Lecture 2
The Language Learning Mechanism
&
Some Child Language Research Methods

Language Acquisition Timeline

Timeline of Language Development: Year 1



phonology — vocal play — canonical babbling

lexicon — recognition of own name — first word

grammar

Timeline of Language Development: Year 2

phonology — reorganization & consolidation of sound system

lexicon — 50 word productive vocabulary — vocabulary "spurt"

grammar — first word combinations

Timeline of Language Development: Year 3-3.5

phonology

lexicon

500 word productive vocabulary

grammar

increasing length of word combinations

adding grammatical morphemes

complex multi-clause utterances

negatives and questions

Levels of Representation Marr (1982)

Describing vs. Explaining

"…it gradually became clear that something important was missing that was not present in either of the disciplines of neurophysiology or psychophysics. The key observation is that neurophysiology and psychophysics have as their business to *describe* the behavior of cells or of subjects but not to *explain* such behavior….What are the problems in doing it that need explaining, and what level of description should such explanations be sought?" - 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."

On Explaining (Marr 1982)

"But the important point is that if the notion of different types of understanding is taken very seriously, it allows the study of the information-processing basis of perception to be made *rigorous*. It becomes possible, by separating explanations into different levels, to make explicit statements about what is being computed and why and to construct theories stating that what is being computed is optimal in some sense or is guaranteed to function correctly. The ad hoc element is removed…"

On Explaining (Marr 1982)

"But the important point is that if the notion of different types of understanding is taken very seriously, it allows the study of the information-processing basis of perception to be made *rigorous*. It becomes possible, by separating explanations into different levels, to make explicit statements about what is being computed and why and to construct theories stating that what is being computed is optimal in some sense or is guaranteed to function correctly. The ad hoc element is removed…"

Our goal: Substitute "language acquisition" for "perception".

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

(3,4) $\longrightarrow$ 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 being computed

The three levels: An example with the cash register

Computational
What does this device do?
Arithmetic (ex: addition).
Addition: Mapping a pair of numbers to another number.

Algorithmic
What is the input, output, and method of transformation?
Input: arabic numerals (0,1,2,3,4…)
Output: arabic numerals (0,1,2,3,4…)
Method of transformation: rules of addition, where least significant digits are added first and sums over 9 have their next digit carried over to the next column

$$\begin{array}{r} 99 \\ +\ \ 5 \\ \hline \end{array}$$

The three levels: An example with the cash register

Computational
What does this device do?
Arithmetic (ex: addition).
Addition: Mapping a pair of numbers to another number.

Algorithmic
What is the input, output, and method of transformation?
Input: arabic numerals (0,1,2,3,4…)
Output: arabic numerals (0,1,2,3,4…)
Method of transformation: rules of addition, where least significant digits are added first and sums over 9 have their next digit carried over to the next column

$$\begin{array}{r} 99 \\ +\ \ 5 \\ \hline 14 \end{array}$$

The three levels: An example with the cash register

Computational
What does this device do?
Arithmetic (ex: addition).
Addition: Mapping a pair of numbers to another number.

Algorithmic
What is the input, output, and method of transformation?
Input: arabic numerals (0,1,2,3,4…)
Output: arabic numerals (0,1,2,3,4…)
Method of transformation: rules of addition, where least significant digits are added first and sums over 9 have their next digit carried over to the next column

$$\begin{array}{r} 1\ \ \\ 99 \\ +\ \ 5 \\ \hline 4 \end{array}$$

The three levels: An example with the cash register

Computational
What does this device do?
Arithmetic (ex: addition).
Addition: Mapping a pair of numbers to another number.

Algorithmic
What is the input, output, and method of transformation?
Input: arabic numerals (0,1,2,3,4…)
Output: arabic numerals (0,1,2,3,4…)
Method of transformation: rules of addition, where least significant digits are added first and sums over 9 have their next digit carried over to the next column

$$\begin{array}{r} 1\ \ \\ 99 \\ +\ \ 5 \\ \hline 104 \end{array}$$

The three levels: An example with the cash register

Computational

What does this device do?

Arithmetic (ex: addition).

Addition: Mapping a pair of numbers to another number.

Algorithmic

What is the input, output, and method of transformation?

Input: arabic numerals (0,1,2,3,4…)

Output: arabic numerals (0,1,2,3,4…)

Method of transformation: rules of addition

Implementational

How can the representation and algorithm be realized physically?

A series of electrical and mechanical components inside the cash register.

The three levels

Marr (1982)

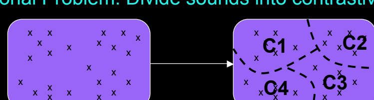
“Although algorithms and mechanisms are empirically more accessible, it is the top level, the level of computational theory, which is critically important from an information-processing point of view. The reason for this is that the nature of the computations that underlie perception depends more upon the computational problems that have to be solved than upon the particular hardware in which their solutions are implemented. To phrase the matter another way, an algorithm is likely to be understood more readily by understanding the nature of the problem being solved than by examining the mechanism (and the hardware) in which it is embodied.”

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

hɨwzəfɹljdəvðəbɪ□gbɹ□dwə□lf

↓

hɨwz əfɹljd əv ðə bɪ□g bɹ□d wə□lf

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: 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 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

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

Input = sounds, syllables, words, phrases, …
Output = sound categories, words, grammatical categories, sentences, …
Method = statistical learning, algebraic learning, prior knowledge about how human languages work, …

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 are available, and should the learner use all of them?
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 are available, and should the learner use all of them?
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.

Experimental Methods: What, When, and Where

Experimental Methods

How do we tell what infants know, or use, or are sensitive to?

Researchers use indirect measurement techniques.

High Amplitude Sucking (HAS)

Infants are awake and in a quietly alert state. They are placed in a comfortable reclined chair and offered a sterilized pacifier that is connected to a pressure transducer and a computer via a piece of rubber tubing. Once the infant has begun sucking, the computer measures the infant's average sucking amplitude (strength of the sucks).

Experimental Methods

How do we tell what infants know, or use, or are sensitive to?

Researchers use indirect measurement techniques.

High Amplitude Sucking (HAS)

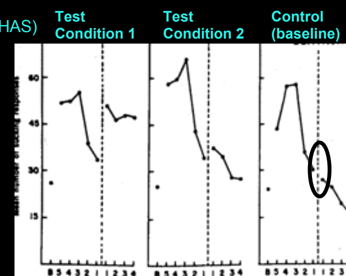
A sound is presented to the infant every time a strong or "high amplitude" suck occurs. Infants quickly learn that their sucking controls the sounds, and they will suck more strongly and more often to hear sounds they like the most. The sucking rate can also be measured to see if an infant notices when new sounds are played.

Experimental Methods

How do we tell what infants know, or use, or are sensitive to?

Researchers use indirect measurement techniques.

High Amplitude Sucking (HAS)

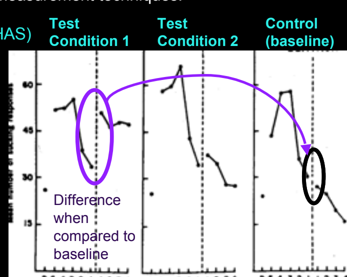


Experimental Methods

How do we tell what infants know, or use, or are sensitive to?

Researchers use indirect measurement techniques.

High Amplitude Sucking (HAS)

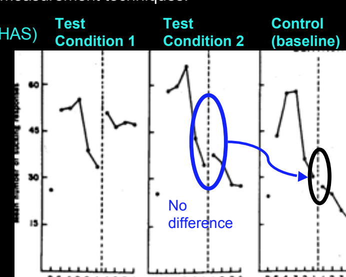


Experimental Methods

How do we tell what infants know, or use, or are sensitive to?

Researchers use indirect measurement techniques.

High Amplitude Sucking (HAS)



Experimental Methods

How do we tell what infants know, or use, or are sensitive to?

Researchers use indirect measurement techniques.

High Amplitude Sucking (HAS)

Infants have sophisticated discrimination abilities, but they don't abstract sounds into categories the way that adults do.

“dɹ” Adult perception “tɹ”

phonemic category phonemic category

Experimental Methods

How do we tell what infants know, or use, or are sensitive to?

Researchers use indirect measurement techniques.

High Amplitude Sucking (HAS)

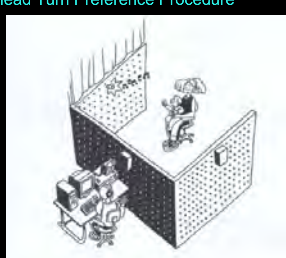
Infants have sophisticated discrimination abilities, but they don't abstract sounds into categories the way that adults do.

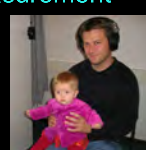
Infant perception

“dɹ 1” “dɹ 2” “tɹ 1” “tɹ 2”

Another useful indirect measurement

Head Turn Preference Procedure

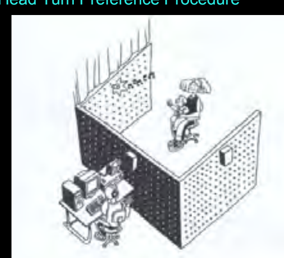


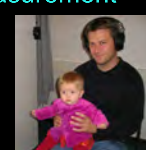


Infant sits on caretaker's lap. The wall in front of the infant has a green light mounted in the center of it. The walls on the sides of the infant have red lights mounted in the center of them, and there are speakers hidden behind the red lights.

Another useful indirect measurement

Head Turn Preference Procedure

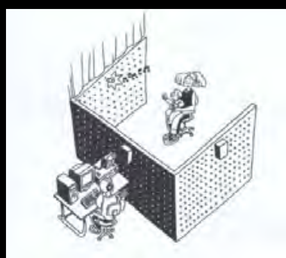




Sounds are played from the two speakers mounted at eye-level to the left and right of the infant. The sounds start when the infant looks towards the blinking side light, and end when the infant looks away for more than two seconds.

Another useful indirect measurement

Head Turn Preference Procedure



Thus, the infant essentially controls how long he or she hears the sounds. Differential preference for one type of sound over the other is used as evidence that infants can detect a difference between the types of sounds.

Note on infant attention:

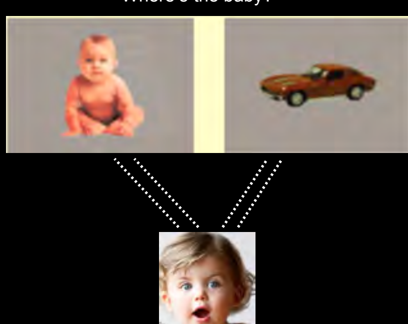
Familiarity vs. Novelty Effects

For procedures that involve measuring where children prefer to look (such as head turn preference), sometimes children seem to have a “familiarity preference” where they prefer to look at something similar to what they habituated to. Other times, children seem to have a “novelty” preference where they prefer to look at something different to what they habituated to.

Kidd, Piantadosi, & Aslin (2010) provide some evidence that this may have to do with the informational content of the test stimulus. There may be a “Goldilocks” effect where children prefer to look at stimuli that are neither to boring nor too surprising, but are instead “just right” for learning, given the child’s current knowledge state.

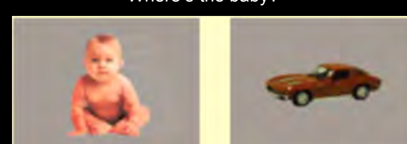
Eyetracking: measures fixations on target picture

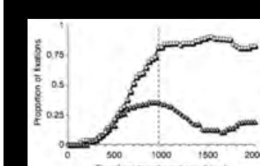
“Where’s the baby?”



Eyetracking: measures fixations on target picture

“Where’s the baby?”





“Where’s the **baby**?”

“Where’s the **vaby**?”

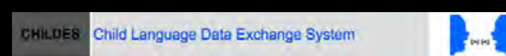
Computational Methods: How

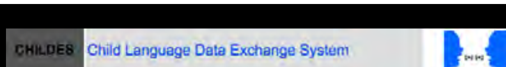
Computational Methods

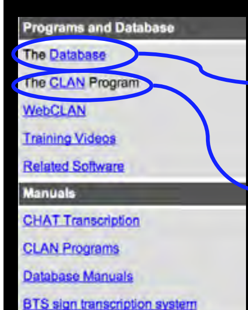
Control over the entire learning mechanism:
- what hypotheses the (digital) child considers
- what data the child learns from
- how the child updates beliefs in different hypotheses

Ground with empirical data available
- want to make this as realistic as possible (ex: use actual data distributions, cognitively plausible update procedures)
- a good source of empirical data: CHILDES database

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



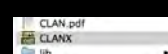




Download annotated transcripts from the database.

Download the program to search these transcripts, and its manual.

Plug for the spring



To learn how to use the freely available Computerized Language ANalysis tool available at http://childes.psy.cmu.edu/clan, check out Psych247M in the spring, a computational methods course for language (acquisition) research.

Back to modeling

Gauges of modeling success & contributions to science

Formal sufficiency: does the model learn what it's supposed to learn when it's supposed to learn it from the data it's supposed to learn it from?

Developmental compatibility: Does it learn in a psychologically plausible way? Is this something children could feasibly do?

Explanatory power: what's the crucial part of the model that makes it work? How does this impact the larger language acquisition story?

Sample learning models

Phoneme acquisition (Vallabha et al . 2007): learning contrastive sounds from raw acoustic data

Word segmentation (Gambell & Yang 2006): learning to identify words in fluent speech from streams of syllables

Categorization (Mintz 2003): learning to identify what category a word is (noun, verb) from segmented speech

Sample learning models

Morphology (Rumelhart & McClelland 1986, Yang 2002, Albright & Hayes 2002, Yang 2005): learning to identify past tense affixes from speech segmented into phonemes/syllables/words

Learning the interpretation of referential elements (Foraker et al. 2007, 2009, Pearl & Lidz 2009): learning to identify syntactic category and semantic referent of *one* from segmented speech and referents in the world

Syntactic acquisition (Reali & Christiansen 2005, Kam et al. 2008, Pearl & Weinberg 2007): learning to identify correct word order (rules) from speech segmented into words

Stress (Pearl 2008): learning to identify correct stress patterns (and rules behind them) from words with stress contours

General Modeling Process

(1) Decide what kind of learner the model represents (ex: normally developing 6-month-old child learning first language)

(2) Decide what data the child learns from (ex: Bernstein corpus from CHILDES) and how the child processes that data (ex: data divided into syllables)

(3) Decide what hypotheses the child has (ex: what the words are) and what information is being tracked in the input (ex: transitional probability between syllables)

(4) Decide how belief in different hypotheses is updated (ex: based on transitional probability minima between syllables)

General Modeling Process

(5) Decide what the measure of success is

- precision and recall (ex: finding the right words in a word segmentation task)

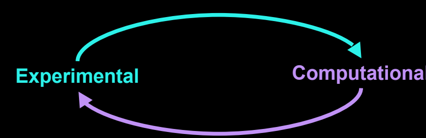
- matching an observed performance trajectory (ex: English past tense acquisition often has a U-shaped curve)

- achieving a certain knowledge state by the end of the learning period (ex: knowing there are 4 vowel categories at the end of a phoneme identification task)

- making correct generalizations (ex: preferring a correctly formed sentence over an incorrectly formed one)

The Aim

Dovetailing between experimental and computational methods, each feeding into the other to increase general understanding of language acquisition.



Extra slides

Looking at children's brains

ERPs: Event-related brain potentials, gauged via electrode caps. The location of ERPs associated with different mental activities is taken as a clue to the area of the brain responsible for those activities.

Good: non-invasive, relatively undemanding on the subject, provide precise timing on brain events

Bad: poor information on exact location of ERP since just monitoring the scalp

Looking at children's brains

Brain-imaging techniques: gauge what part of the brain is active as subjects perform certain tasks

PET scans: Positron emission topography scans

- subjects inhale low-level radioactive gas or injected with glucose tagged with radioactive substance
- experimenters can see which parts of the brain are using more glucose (requiring the most energy)

fMRI scans: functional magnetic resonance imaging

- subjects have to be very still inside MRI machine, which is expensive to operate
- experimenters can see which parts of the brain are getting more blood flow or consuming more oxygen

Looking at children's brains

Brain-imaging techniques: gauge what part of the brain is active as subjects perform certain tasks

MEG: Magnetoencephalography

- subjects have to be very still
- experimenters can see which parts of the brain are active

Looking at children's brains

Brain-imaging techniques: gauge what part of the brain is active as subjects perform certain tasks

Optical Topography: Near-infrared spectroscopy (NIRS)

- transmission of light through the tissues of the brain is affected by hemoglobin concentration changes, which can be detected