How Do Children Learn Language?

Aida Nematzadeh
nematzadeh@google.com

Language Acquisition in Children

Children **effortlessly learn** their language from a noisy and ambiguous input.

Language Acquisition in Machines

Understanding language acquisition might help us build AI systems that understand and produce natural languages.

Is Language Learned? How?

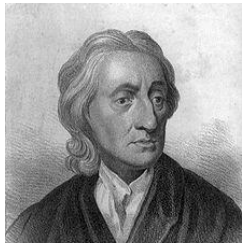
Is Language Learning Effortless?

Learning Mechanisms

Learning about Words

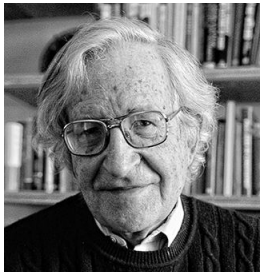
Learning the Structure

Nurture vs Nature



John Locke

empiricism — nativism



Noam Chomsky

Knowledge and reason come from experience.

Language: outcome of how children are **nurtured** (like table manner).

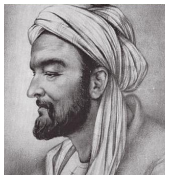
Mind has preexisting structure to interpret experience.

Language: outcome of **nature** -- an innate endowment (like upright posture).

Empiricism vs Nativism

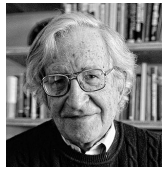
"The human intellect at birth is rather like a **tabula rasa**, a pure potentiality that is actualized through education and comes to know. Knowledge is attained through empirical familiarity with objects in this world from which one abstracts universal concepts."

Avicenna (980-1037 AD)



"**Language learning** is not really something that the child does; it is something that happens to the child placed in an appropriate environment, **much as the child's body grows and matures** in a predetermined way when provided with appropriate nutrition and environmental stimulation."

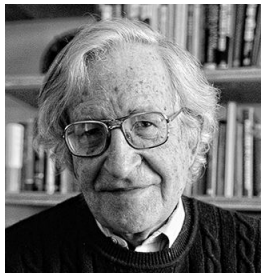
Chomsky (1928-)



Cognitive Revolution

Burrhus Frederic Skinner

behaviorism ⟷ cognitivism



Noam Chomsky

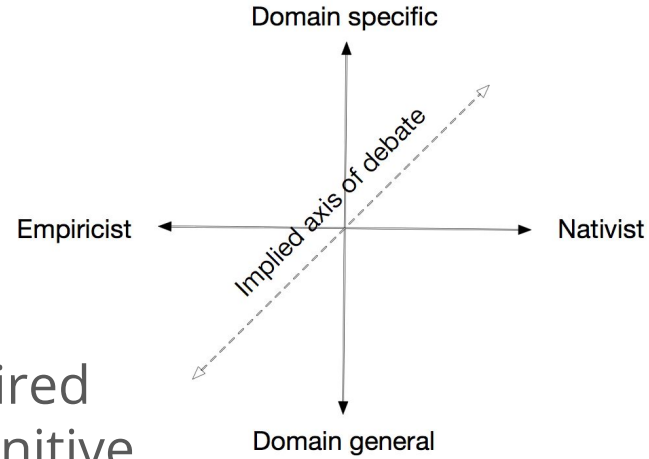
Can explain behavior in terms of things external to mind.

Language ~ verbal behavior

Explaining behavior requires understanding the mind.

Language ~ mental process

Domain-General vs Domain-Specific Learning



Language is acquired rapidly, effortlessly, and without direct instruction.

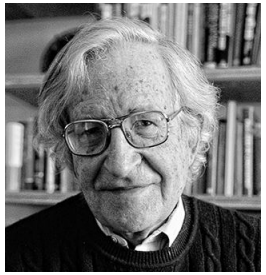
Language is acquired using general cognitive skills like memory, capacity for symbolic representation, and statistical learning.

[Frank *et al*, 2019]

Language for Communication

Michael Tomasello

functionalism ⟷ formalism



Noam Chomsky

Language is shaped by its communicative functions.

Language is acquired through communication (not passive observation).

Language form is independent of its function.

Acquisition of language is not affected by the fact that we use it to communicate.

Discussion: Development vs Learnability

Modeling language development to shed light on its underlying mechanism. cognitive science

Can we learn language (certain linguistic phenomena) from data (in an unsupervised way)?

ML/NLP

Nature of Nature

Investigate the innateness/learnability of

- knowledge -- inborn linguistic knowledge?
- computational procedure -- domain-general or domain-specific learning mechanism?
  - Same models for vision and language?

Is Language Learned? How?

Is Language Learning Effortless?

Learning Mechanisms

Learning about Words

Learning the Structure

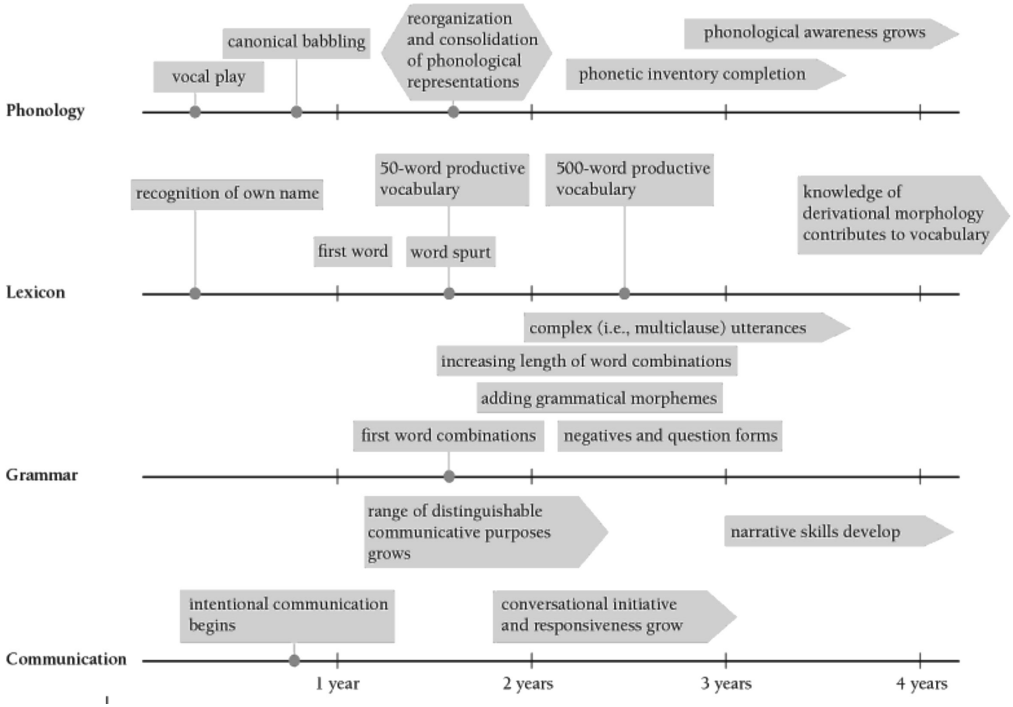
0-12m	12-24m	18-30m	24-48m
prelinguistic communication	single words	telegraphic speech	grammatical development
"bobo"	“mummy” “doggy”	“daddy sleep” “orange juice”	“I want some eggs” “Put it table”

Takes children 5 years (14,600h, 8h/day).

Would take adults 56 years (2920 weeks, 5h/week).

0-12m	12-24m	18-30m	24-48m
prelinguistic communication	single words	telegraphic speech	grammatical development
“bobo”	“mummy” “doggy”	**“daddy sleep”** “orange juice”	“I want some eggs” **“Put it table”**

Children make errors but learn to correct them.


Phonology
vocal play
canonical babbling
reorganization and consolidation of phonological representations
phonetic inventory completion
phonological awareness grows
Lexicon
recognition of own name
first word
word spurt
50-word productive vocabulary
500-word productive vocabulary
knowledge of derivational morphology contributes to vocabulary
Grammar
first word combinations
increasing length of word combinations
adding grammatical morphemes
complex (i.e., multiclause) utterances
negatives and question forms
Communication
intentional communication begins
range of distinguishable communicative purposes grows
conversational initiative and responsiveness grow
narrative skills develop
1 year
2 years
3 years
4 years

[Hoff, 2004]

Discussion

Should AI models make the same mistakes as children?

Should we model all the domains at the same time?

Is Language Learned? How?
Is Language Learning Effortless?
Learning Mechanisms
Learning about Words
Learning the Structure

Babies as Statistical Learners [Saffran *et al*, Science 1996]

8-month-old infants learn within- and between-word transitional probabilities from novel speech.

- bidakupadotigolabutupiropadotibidaku

Statistical learning in other domains: phonology, syntax, & words.[Gomez *et al*, 2000; Mintz *et al*, 2002; Smith & Yu, 2008; Romberg & Saffran, 2010]

Statistical learning is **domain-** & **species-** general.

Babies as Rule Learners [Marcus *et al*, Science 1999]

Seven-month-old infants can learn simple “algebra-like” rules.

- “ga ti ti” “li la la” (ABB) or “li la li” “ga la ga” (ABA)

Rule learning is statistical learning? [Christiansen & Curtin, 1999; Seidenberg & Elman, 1999; McClelland & Plaut, 1999]

Babies as Social Learners

Sharing joint attention.

Understanding and sharing intention. [Tomasello *et al,* 2005]

Infants learn about phonetics by listening to native speakers but not their audio/video. [Kuhl *et al,* 2003]

Discussion

What type of learning does each linguistic domain require?

What about other domain-general cognitive mechanisms (*e.g*, memory?) [Nematzadeh *et al*, BAICS 2020]

What modeling frameworks are suitable for each?

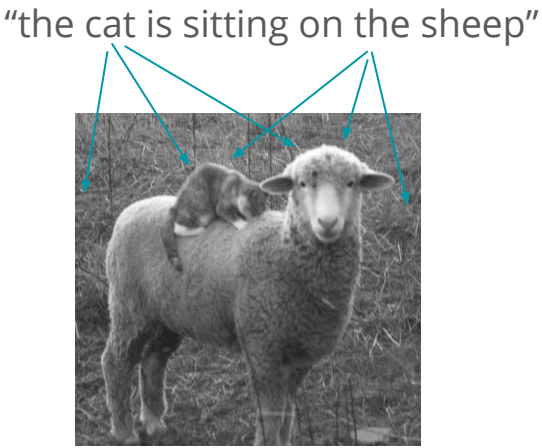
Is Language Learned? How?
Is Language Learning Effortless?
Learning Mechanisms
Learning about Words
Learning the Structure

Word Learning Stages

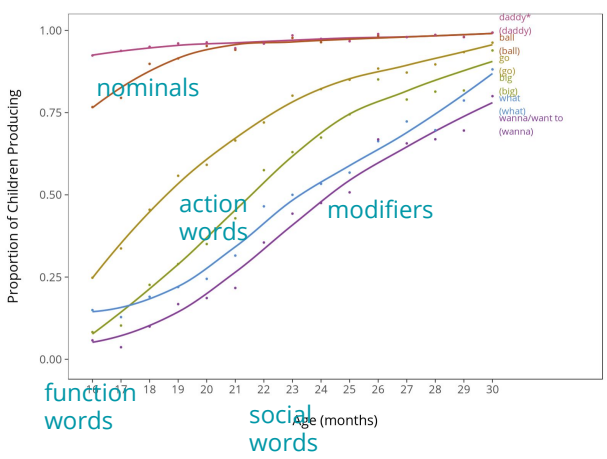
Segmenting speech to words.

[de][kat][iz][sitting][on][de][sheep]

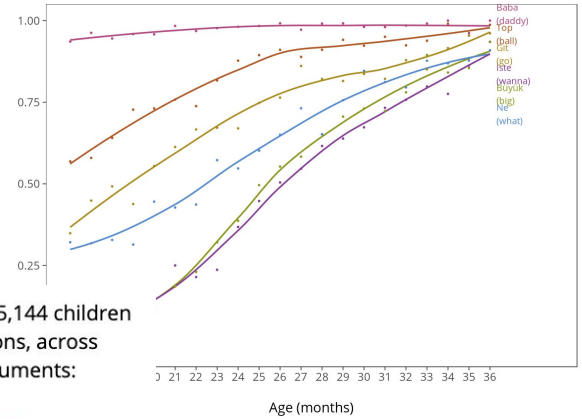
Mapping a meaning to words.



Early Words



English



Wordbank contains data from 75,144 children and 82,983 CDI administrations, across 29 languages and 56 instruments:

http://wordbank.stanford.edu/

Turkish

Word Learning Errors

Underextension: using words in a more restricted fashion; “dog” to refer to spaniels.

Overextension: using words more broadly; all four-legged animals as “doggie”.

- “cat”: cat, cat’s usual location on the top of TV when absent. [Rescorla, 1980]

Context-bound Words

Used only in one context: saying “duck” **only** when hitting the toy to the bathtub. [Barrett, 1986]

Are parts of language games.

Function-specific understanding -- different from adults’ mental representations of words.

Cross-situational Learning

People (as young as 12-month-old infants) are sensitive to the statistical regularities across situations. [Pinker 1989; Yu & Smith 2007; Smith & Yu, 2008]

A zant

Look at the zant!

Biases that Guide Word Learning

The input is noisy and ambiguous: many possible mappings/hypotheses for word meanings.

People learn word meanings from a few exposures.

Learned/innate biases might facilitate learning.

Biases that Guide Word Learning

mutual exclusivity bias
[Markman & Wachtel, 1988]
taxonomic bias
[Markman & Hutchinson, 1984; Markman, 1989]
basic-level bias
[Rosch *et al*, 1976; Markman, 1991]

whole-object bias [Markman, 1991]
shape bias [Smith & Jones, 1988]

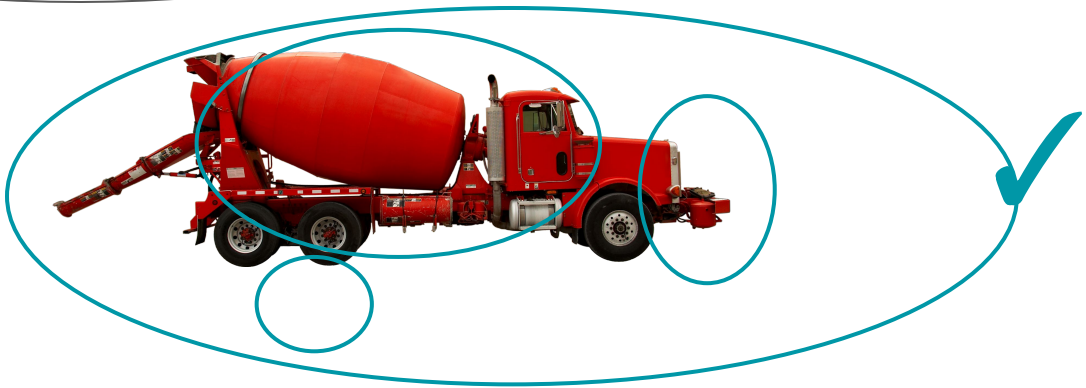
attention
[Samuelson & smith, 1998;
Yu *et al*, 2017]

social-pragmatic biases
communicative intentions
[Bloom, 2000; Tomasello, 2001]
following eye gaze
[Baldwin, 1993]

syntax
[Brown, 1957;
Gelman & Markman, 1985]
noun bias
[Gentner, 1982]

The Whole-Object Bias [Markman, 1991]

What is dax?



Learn word labels for the whole object.

The Mutual Exclusivity Bias [Markman & Wachtel, 1988]

What is dax?

18-old-month children exhibit the bias. [Markman *et al*, 2003]

familiar object

unfamiliar object ✓

Limit the number of possible word labels for a familiar object.

The Basic-Level Bias

Zant → Golden Retriever?

Zant → dog (any dog breed)?

Zant → animal?

Cross-situational statistics are **consistent** with all.

Why dog? A bias that focuses generalization to the **basic-level** (cognitively natural) categories.

Syntactic Bootstrapping

Language structure supports learning new verbs.

[Gleitman, 1990; Fisher et al, 1994]

[Naigles, 1990]

"The rabbit is gorping the duck." or

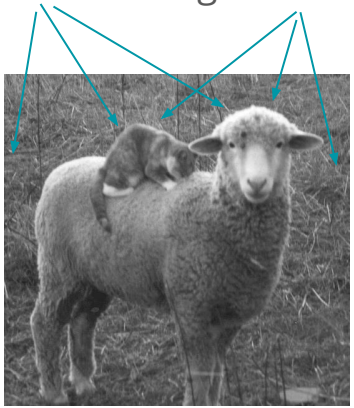
"The rabbit and the duck are gorping."

"where is gorping now?"

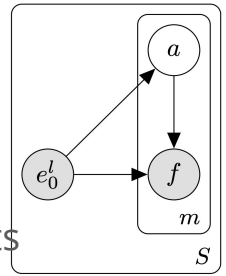
Modeling Word Learning

Solving the translation problem: mapping words to observations. [Siskind, 1996; Yu & Ballard, 2007; Frank *et al*, 2009; Fazly *et al*, 2010; Nematzadeh *et al*, 2015]

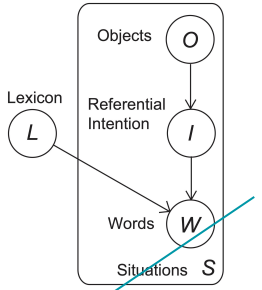
"the cat is sitting on the sheep"



objects



words



[Frank *et al*, 2009]

How do we model intent?

Word Learning Challenges (Children / NLP)

Learning of:

- Verbs -- relations among entities; event boundary
- Abstract words (freedom vs. cooperation)
- Low frequency words

Shortcomings of modeling words as vectors in a Euclidean space [Nematzadeh et, 2017; Griffiths et al., 2007]

Is Language Learned? How?
Is Language Learning Effortless?
Learning Mechanisms
Learning about Words
Learning the Structure

Language is Productive

We have the capacity to produce and understand an infinite number of new sentences.

Two productive systems:

- Syntax: sentence structure; ordering of words.
- Morphology: structure of words & word parts.

Syntax: Level of Abstraction

“Rita drinks milk.”

- Sentence → Rita + drinks + milk (not productive)
- Sentence → agent of action + action + theme

“Rita resembles Ray.”

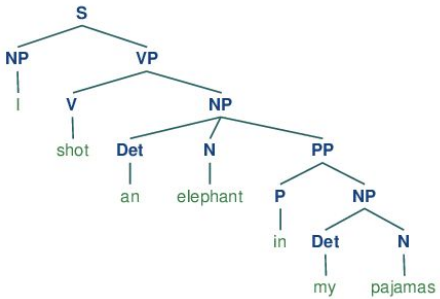
- Sentence → noun + verb + noun

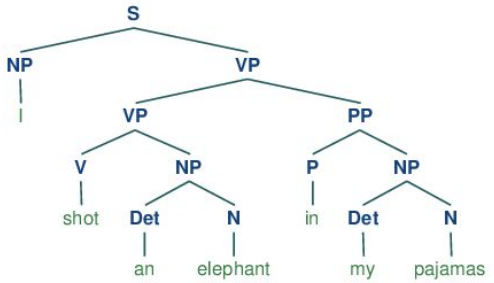
What is origin of the variables and the rules?

Syntax: Type of Structure

Sentences have hierarchical structure.

- *“The (clever) cat cried (a river).”*
- S → NP + VP, NP → (det) + (adj) + N, VP → V + NP





Syntax: Type of Structure

Sentences have hierarchical structure.

- *“The (clever) cat cried (a river).”*
- S → NP + VP, NP → (det) + (adj) + N, VP → V + NP

Is human language use hierarchical? [Frank *et al*, 2012]

Morphology

Adds grammatical information to words.

- Plural s in English

Children learn morphology earlier when language is morphologically rich. [Peters, 1995]

Easy morphemes to learn: frequent, fixed form and relative position to stem, clear function.

Do Children Know Grammatical Rules?

Early word combinations are systematic.

- “my teddy” (possessor + possessed)
- “daddy sit” (actor + action)

Overgeneralization errors:

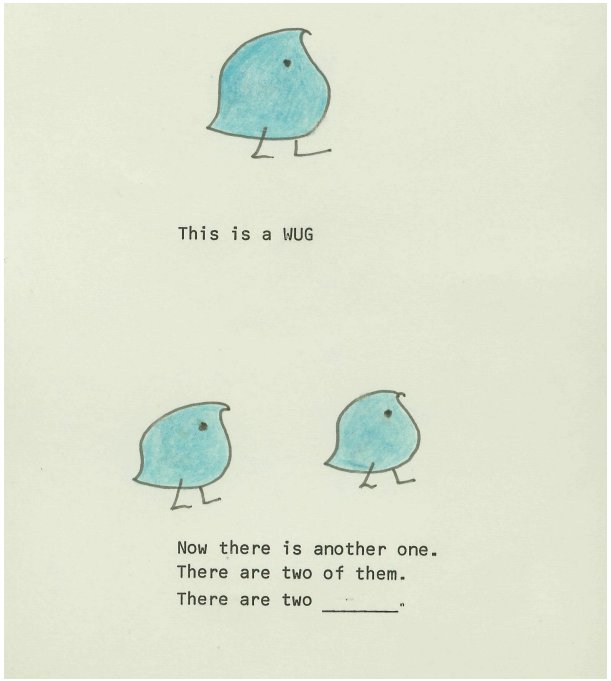
- “I am a good boy, amn’t I” (syntax)
- “toothes”; “breaked” (morphology)

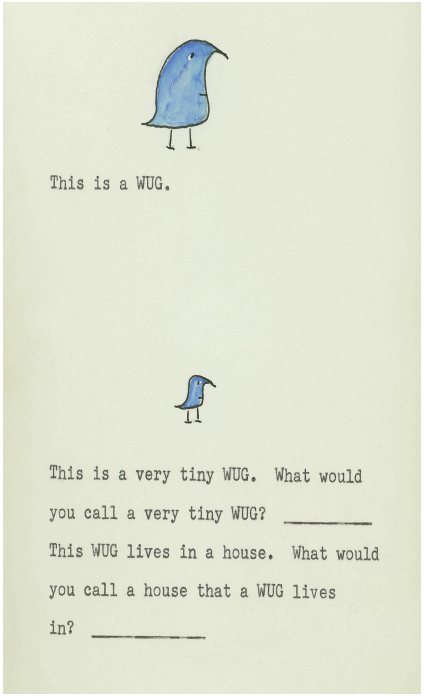
Do Children Know Syntactic Rules?

4-year old children can use novel verbs heard in one sentence structure in others. [Pinker *et al*,1987; Gropen *et al*, 1991]

"The pig is pilking the horse" → "The horse is being pilked by the pig"

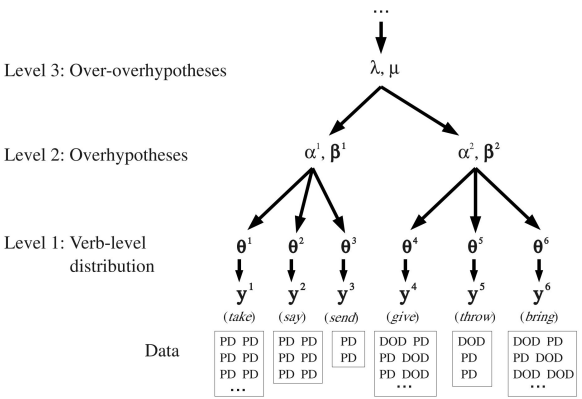
Do Children Know Morphological Rules? [Berko, 1958]



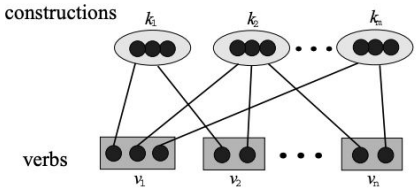


Modeling Structure

Learning abstractions through hierarchical representations. [Alishahi & Stevenson, 2008; Perfors *et al*, 2009; Barak *et al*, 2013]



Debbie gave a pretzel to Dean (PD)
Debbie gave Dean a pretzel (DOD)



[Perfors *et al*, 2009]

[Alishahi & Stevenson, 2008]

Discussion: Nature of Nature

Abstract knowledge

- guides our generalization (priors/inductive biases/constraints)
- results in a productive system (rules/structure)

What are the origins of our abstract knowledge?

Can it be learned from experience?

Generalization to Test Linguistic Knowledge

Children's knowledge of language is examined by generalization tasks:

- Mapping novel words to new/familiar objects.
- Using a new verb in "unheard" structures.
- Applying morphological rules to new words.

Can AI models pass these generalization tasks?

Generalization as Zero-shot Evaluations

Results on fine-tuned task is confounded with dataset size and experimental set-up. [Yogatama et al., 2019]

Probing classifiers can be misleading. [Hewitt and Liang, 2019; Voita and Titov, 2020]

Zero-shot evaluation:

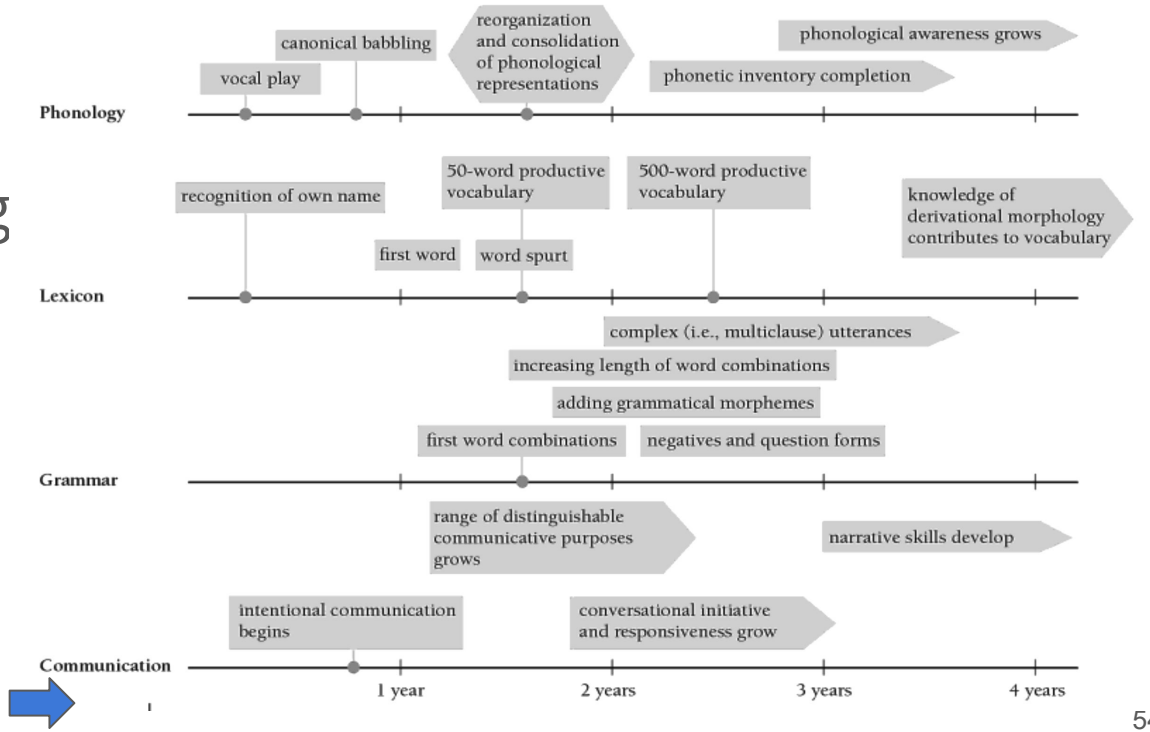
- teach the model what the task is

What Are the Challenges?

Noise in large-scale real-world datasets.

Unsupervised learning of structure.

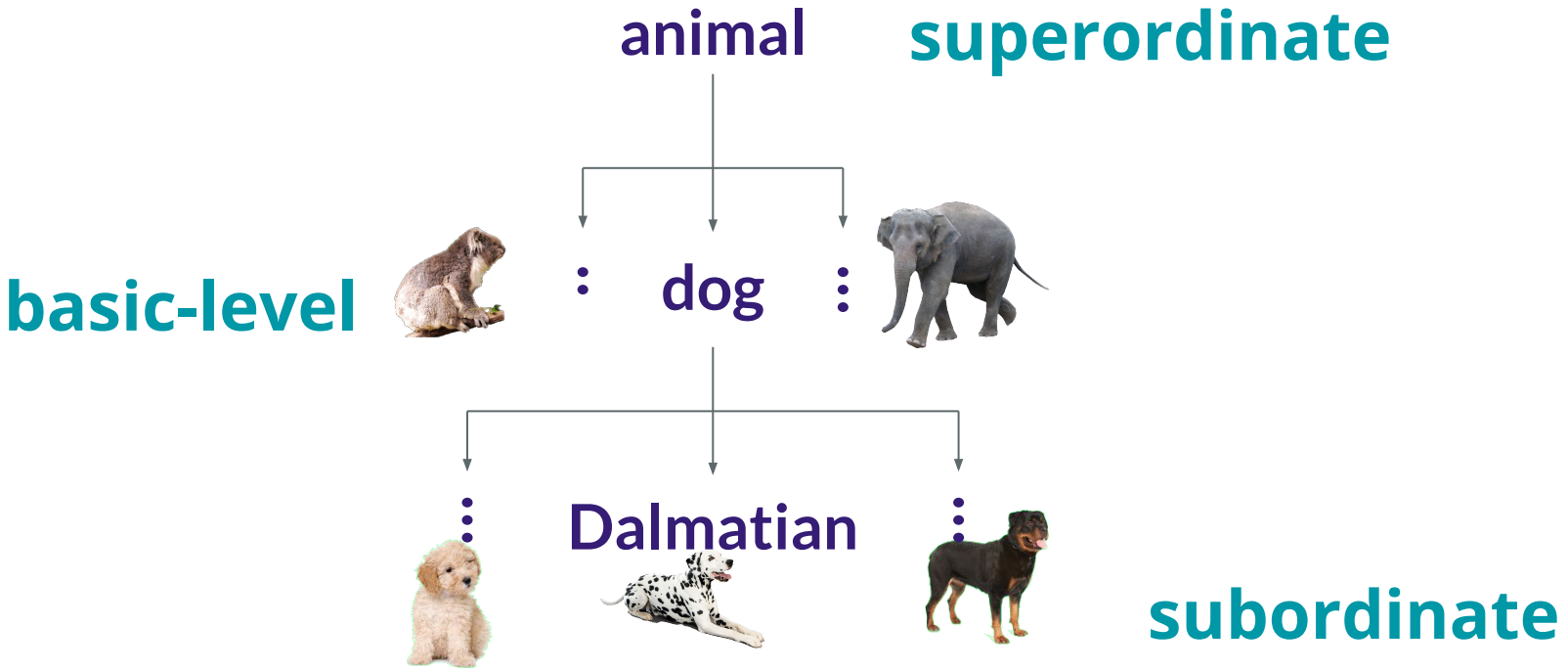
Modeling / annotating intentions.



[Hoff, 2004]

Basic-level Generalization

Different Levels of a Taxonomy



Generalization in People

[Xu & Tenenbaum, Psych Rev 2007]

How to generalize words from a few examples?

Train (3 sub)

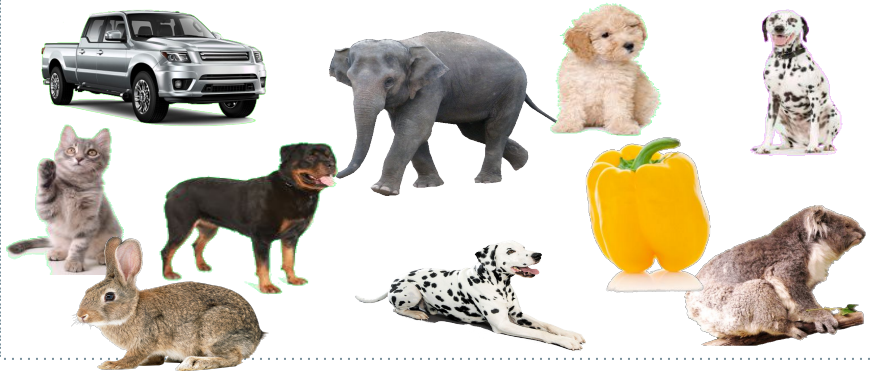
trial 1 *This is a dax.*

trial 2 *Here is a dax.*

trial 3 *A dax.*

Test

Pick everything that is a dax



Generalization in People

[Xu & Tenenbaum, Psych Rev 2007]

How to generalize words from a few examples?

Train (3 basic)

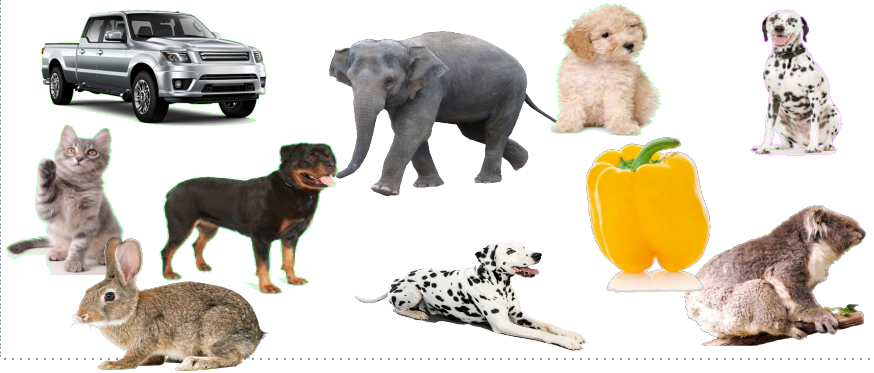
trial 1 *This is a dax.*

trial 2 *Here is a dax.*

trial 3 *A dax.*

Test

Pick everything that is a dax



Generalization in People

[Xu & Tenenbaum, Psych Rev 2007]

How to generalize words from a few examples?

Train (1 sub)

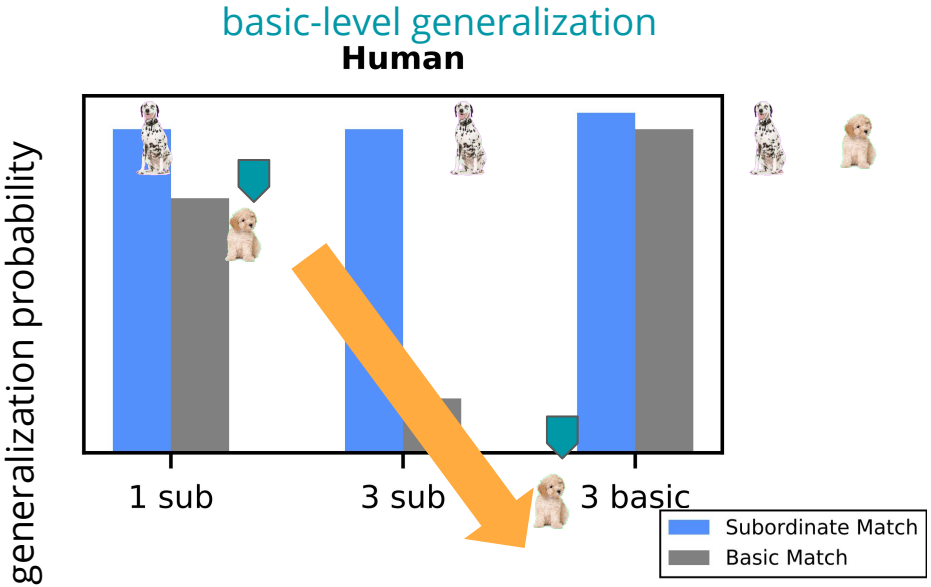
trial 1 *This is a dax.*

Test

Pick everything that is a dax

Generalization in People

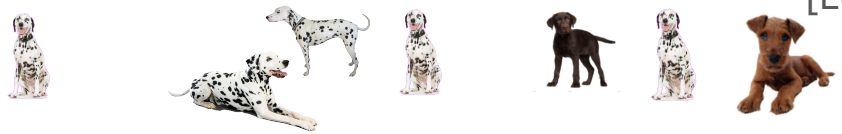
[Xu & Tenenbaum, Psych Rev 2007]



Generalize to the basic-level with only subordinate examples: a basic-level bias.

Basic-level generalization is attenuated.

[Abbott, Austerweil, & Griffiths, CogSci 2012]
[Lewis & Frank, Psych Sci 2018]



Why Are the Results Interesting?

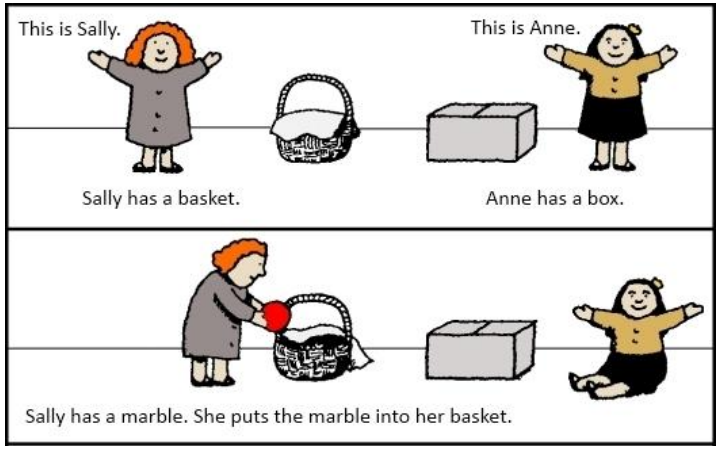
People learn a novel word (“dax”) only from positive examples.

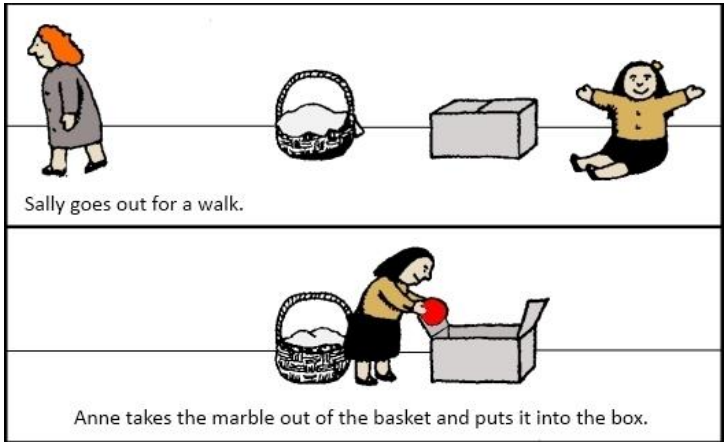
They exhibit a bias towards the basic-level category: is this bias learned or innate? [Nematzadeh *et al.*, EMNLP 2015]

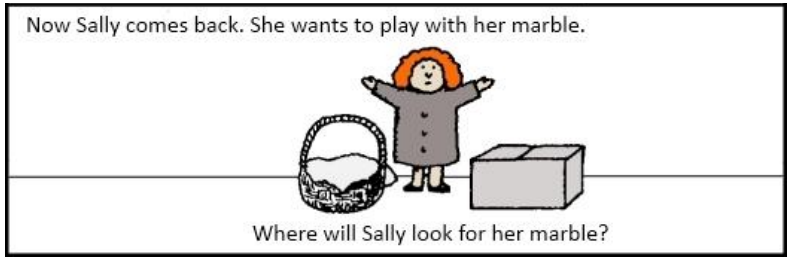
Their generalization is sensitive to the number of examples in a category.

Language Understanding

Theory of Mind: Reasoning About Beliefs





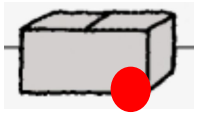


False-belief or
Sally-Anne task
[Baron-Cohen *et al*., 1985]

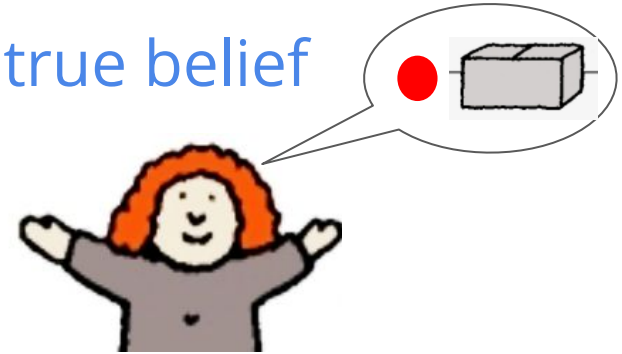
Need to reason about others' beliefs & maintain multiple representations.

True or False Beliefs

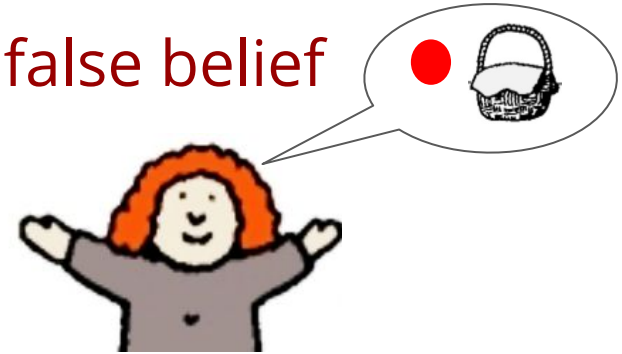
reality



true belief

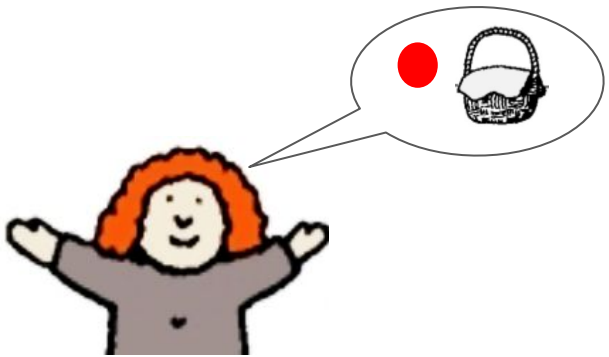


false belief



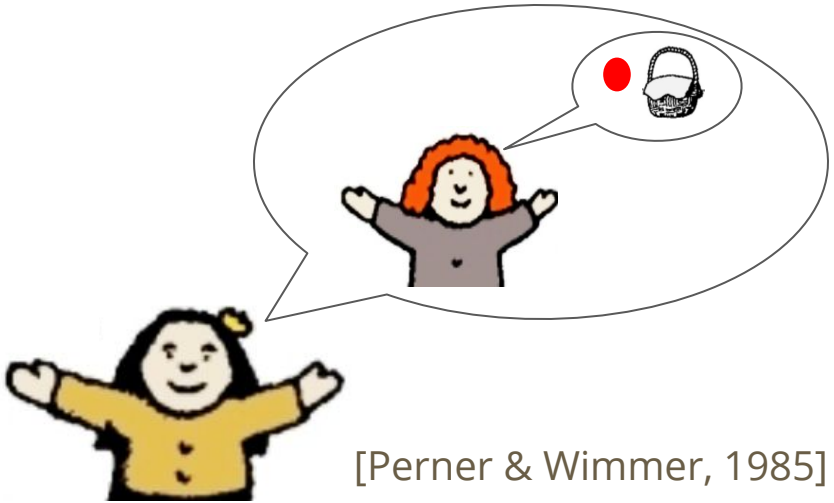
Beliefs About Beliefs

First-order belief: Sally's belief about marble's location.



false

Second-order belief: Anne's belief about Sally's belief.



[Perner & Wimmer, 1985]

True Belief

Anne entered the kitchen.
Sally entered the kitchen.
The milk is in the fridge.
Anne moved the milk to the pantry.

Memory Where was the milk at the beginning?
Reality Where is the milk really?
First-order Where will Sally look for the milk?
Second-order Where does Anne think that Sally searches for the milk?

[Nematzadeh *et al.*, EMNLP 2018]

False Belief

Anne entered the kitchen.
Sally entered the kitchen.
The milk is in the fridge.
Sally exited the kitchen.
Anne moved the milk to the pantry.

Memory Where was the milk at the beginning?
Reality Where is the milk really?
First-order Where will Sally look for the milk?
Second-order Where does Anne think that Sally searches for the milk?

[Nematzadeh *et al.*, EMNLP 2018]

Second-order False Belief

Anne entered the kitchen.
Sally entered the kitchen.
The milk is in the fridge.
Sally exited the kitchen.
Anne moved the milk to the pantry.
Anne exited the kitchen.
Sally entered the kitchen.

Memory Where was the milk at the beginning?
Reality Where is the milk really?
First-order Where will Sally look for the milk?
Second-order Where does Anne think that Sally searches for the milk?

[Nematzadeh *et al.*, EMNLP 2018]