Informatics 1 Cognitive Science

Lecture 10: Word Learning

Frank Keller

6 February 2025

School of Informatics
University of Edinburgh
keller@inf.ed.ac.uk

Slide credits: Frank Mollica, Chris Lucas, Mirella Lapata

Overview

Recap

In order to acquire a lexicon young children segment speech into words using multiple sources of support; we focused on distributional regularities:

- transitional probability provides cues to word boundaries
- Minimum Description Length help assembling words into a lexicon

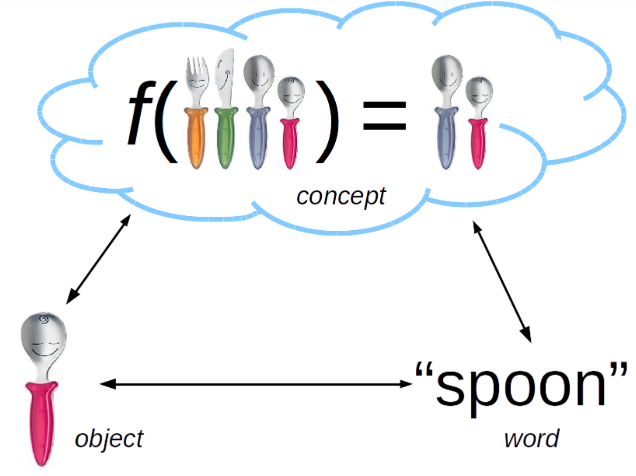
We also saw Bayes Rule as a way of combining prior beliefs with evidence, and updating beliefs in the light of new evidence.

In today's lecture we focus on word learning: How do children associate words with concepts?

We'll see a detailed case study on number words. Bayes Rule will again be important.

Word Learning

Word Learning: The Generalization Problem



Word Learning: The Mapping Problem

W. V. O. Quine (1960). *Word and Object.*

A rabbit!
Our dinner!
Shh, be quiet!
What a cute furry thing!
Rabbit parts!
Get it out!
Don't move!
What long ears!

The child does not know which attribute is being labeled!

Inductive Biases

Inductive Biases

"For any set of data there will be an infinite number of logically possible hypotheses consistent with it. The data are never sufficient logically to eliminate all competing hypotheses." –Ellen Markman

In order to explain how children solve the mapping problem, researchers have hypothesized that they use inductive biases to help with word learning:

- Whole object bias
- Mutual exclusivity

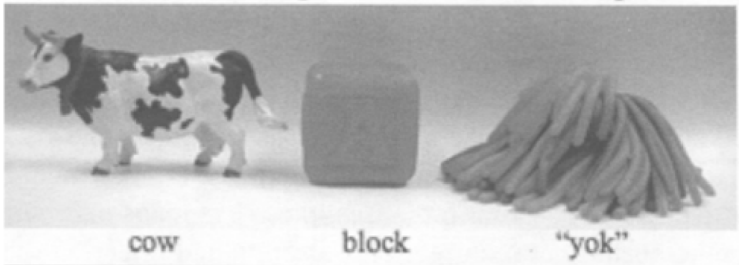
Whole Object Bias

Words refer to the whole object, not its parts.

Mutual Exclusivity

Every object only has one name.

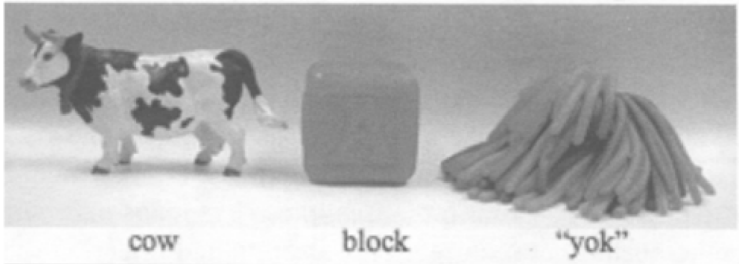
So novel words must apply to objects we don't have a word for already:



Mutual Exclusivity

Every object only has one name.

So novel words must apply to objects we don't have a word for already:



Fast Mapping: Using the whole object bias and mutual exclusivity, children map between a new word and a new object based on a single observation.

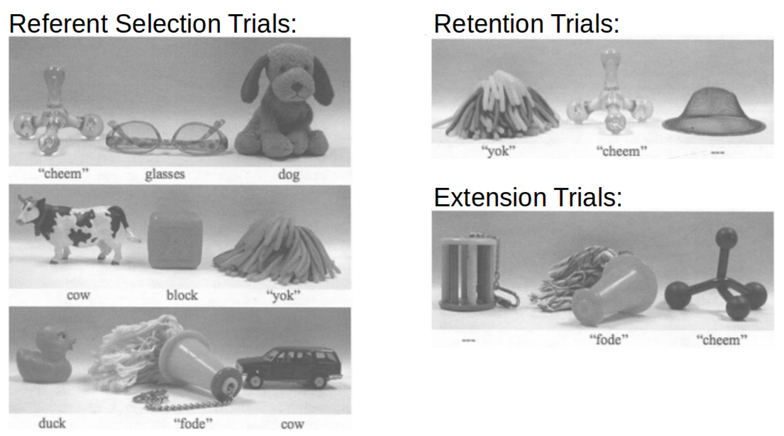
Fast mapping could explain the vocabulary growth spurt that children experience in their second year. But what is the experimental evidence?

Fast Mapping (Horst & Samuelson, 2008)

Children map between a word and an object based on a single observation.

Q1: Do fast mappings last?

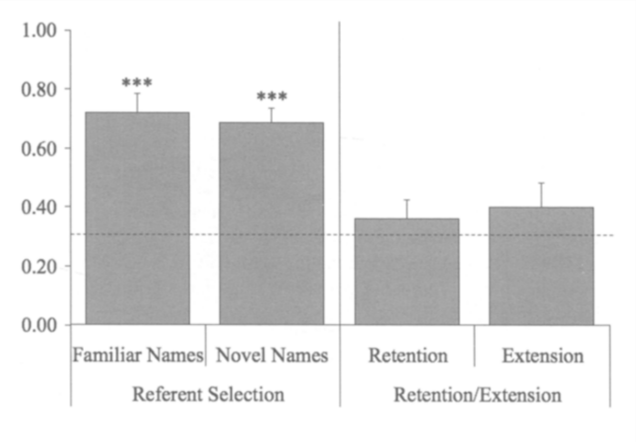
Q2: Do fast mappings also solve the generalization problem?



Fast Mapping (Horst & Samuelson, 2008)

Q1: Do fast mappings last? A: No.

Q2: Do fast mappings also solve the generalization problem? A: No.



Taxonomic Bias and Basic Level Bias

So clearly, the whole object bias and mutual exclusivity are not sufficient to explain word learning.

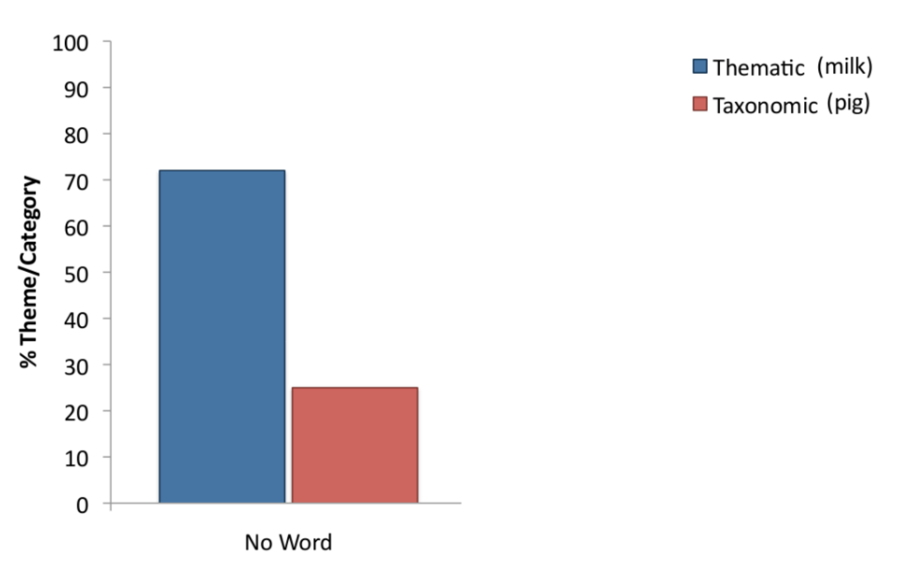
There's evidence for other biases:

- Taxonomic bias: when children hear a new word, they assume its in a taxonomic relationship to words they already know.
- Basic level bias: preference for new words to refer to basic-level categories.

We will come back to these when we talk about categories in the next lecture.

Taxonomic Bias

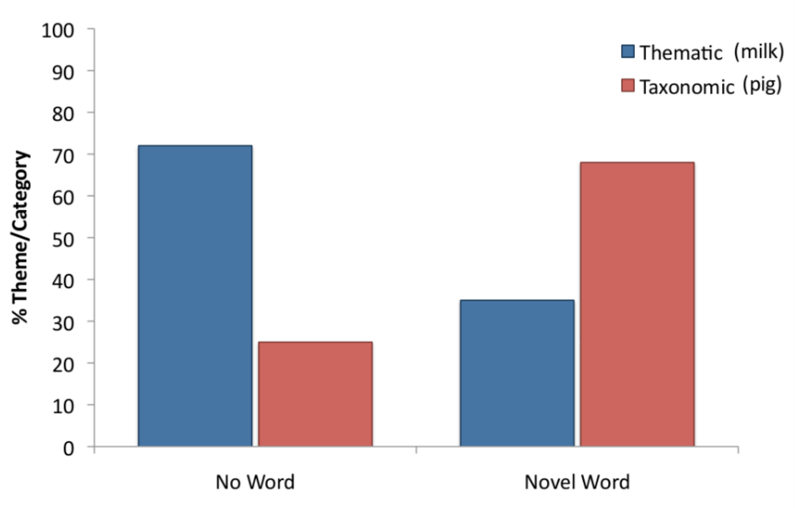
Taxonomic Bias



Taxonomic Bias

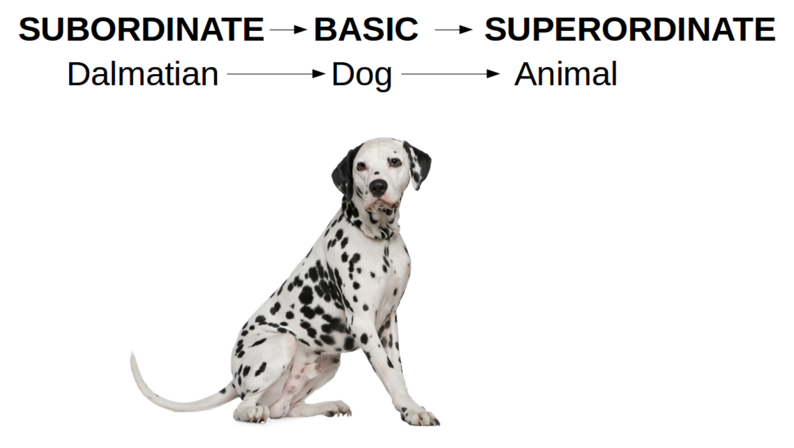
Exposure	Testing
	 
no word: "I'm going to show you something!"	"Can you show me another one?"
novel word: "I'm going to show you a dax!"	"Can you show me another dax?"

Taxonomic Bias

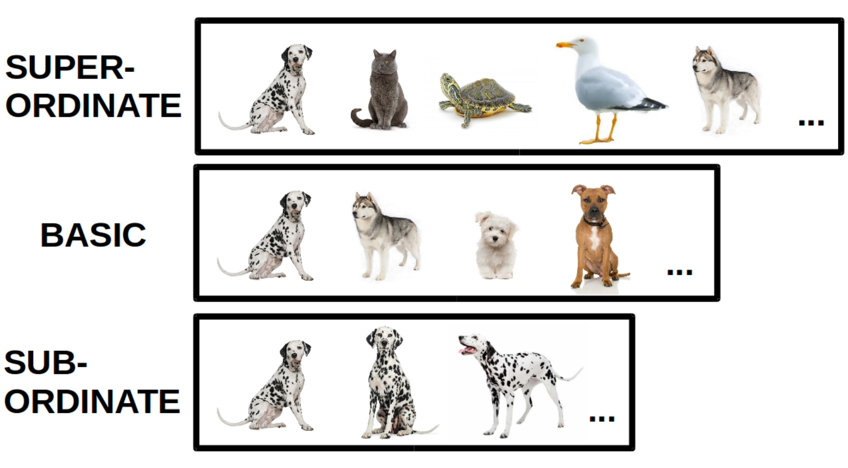


New words refer to taxonomic relationships ("farm animal"), not to associations ("things to do with cows").

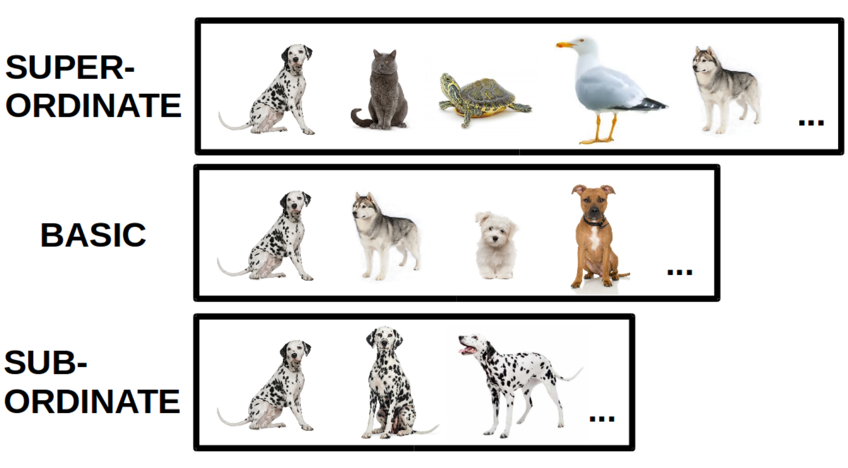
Basic Level Bias



Basic Level Bias



Basic Level Bias



Size Principle: $P(d|h) = \frac{1}{|h|}$. Penalizes hypotheses that pick out sets that are larger than what is required to capture the data.

Inductive Biases in Word Learning

Time for a short quiz on Wooclap!

https://app.wooclap.com/PPUKKP

Modeling Word Learning

Modeling Word Learning

We will look at a model that uses Bayesian hypothesis testing to capture the learning of number words.

Modeling Word Learning

We will look at a model that uses Bayesian hypothesis testing to capture the learning of number words.

When building a model, we need to consider:

- *Input:* What information is your model considering?
- *Output:* What responses are allowed?
- *Hypothesis Space:* What mappings between input and output are possible?
- *Inductive Bias:* How does the model perform when there's no data?
- *Environment:* What training data is available to the model?

Number Words

Children learn number words in stages.

We assess their knowledge using the Give-N task.

(Wyn, 1990; 1992)

Number Words

Children learn number words in stages.

We assess their knowledge using the Give-N task.

(Wyn, 1990; 1992)

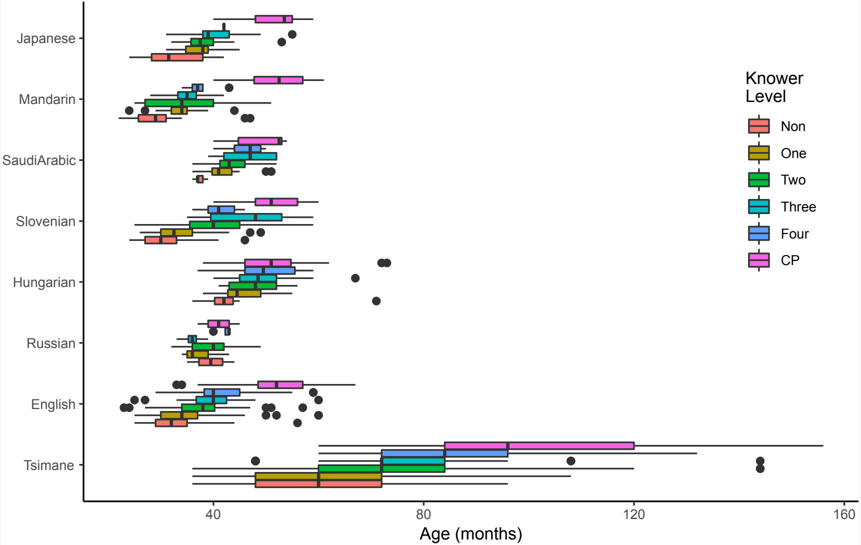
Number Words

Children learn number words in stages.

We assess their knowledge using the Give-N task.

(Wyn, 1990; 1992)

Number Words



CP: cardinal principle (child can produce any number)

Number Words: Hypotheses Space

One-knower

```
λ S . (if (singleton? S)
          “one”
          undef)
```

Two-knower

```
λ S . (if (singleton? S)
          “one”
          (if (doubleton? S)
              “two”
              undef))
```

Three-knower

```
λ S . (if (singleton? S)
          “one”
          (if (doubleton? S)
              “two”
              (if (tripleton? S)
                  “three”
                  undef))
```

CP-knower

```
λ S . (if (singleton? S)
          “one”
          (next (L (set-difference S
                                   (select S)))))
```

(Piantadosi, Tenenbaum & Goodman, 2012)

Number Words: Hypotheses Space

Singular-Plural

```
λ S . (if (singleton? S)
          "one"
          "two")
```

Mod-5

```
λ S . (if (or (singleton? S)
              (equal-word? (L (set-difference S)
                                    (select S))
                           "five"))
          "one"
          (next (L (set-difference S
                                   (select S)))))
```

2-not-1-knower

```
λ S . (if (doubleton? S)
          "two"
          undef)
```

2N-knower

```
λ S . (if (singleton? S)
          "one"
          (next (next (L (set-difference S (select S))))))
```

(Piantadosi, Tenenbaum & Goodman, 2012)

Number Word Learning (simplified from Piantadosi, Tenenbaum & Goodman, 2012)

Program Induction: Which hypothesis (program) h led to the speaker uttering word w when counting set s?

$$P(h|D) = P(h|w,s) \propto P(w|s,h)P(h)$$

Input

(word, set) pairs.

For example:

(*three*, • • •)

Prior

Simplicity bias: simpler programs h are more likely.

N: number of words in the count sequence.

Output

A knower level: 1, 2, 3, 4, CP

Likelihood: Noisy size principle

$$P(w|s,h) = \begin{cases} \frac{1}{N} & \text{if } h(s) = \text{undef} \\ \alpha + (1-\alpha)\frac{1}{N} & \text{if } h(s) = w \\ (1-\alpha)\frac{1}{N} & \text{else} \end{cases}$$

where α is the probability of uttering word w computed by program h for set s (typically, close to 1)

Number Words: Simplicity Prior

Time for a short quiz on Wooclap!

https://app.wooclap.com/PPUKKP

Number Words: Simplicity Prior

How do we define the simplicity prior? We combine:

- rational rules prior: programs with fewer primitives are more probable
- penalty for recursion: programs that use recursion are less probably

Number Words: Simplicity Prior

How do we define the simplicity prior? We combine:

- rational rules prior: programs with fewer primitives are more probable
- penalty for recursion: programs that use recursion are less probably

$$P(h) \propto \begin{cases} \gamma \cdot P_{RR}(h) & \text{if } h \text{ uses recursion} \\ (1-\gamma) \cdot P_{RR}(h) & \text{otherwise} \end{cases}$$

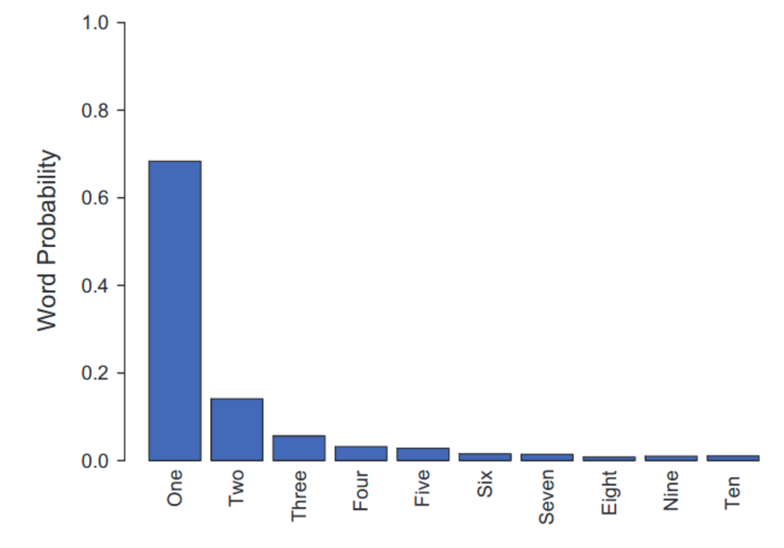
where $P_{RR}(h)$ is the prior of h according to the rational rules model (next page); γ is a parameter that penalizes recursion.

Number Words: Simplicity Prior

Functions mapping sets to truth values	
(*singleton? X*)	Returns true iff the set *X* has exactly one element
(*doubleton? X*)	Returns true iff the set *X* has exactly two elements
(*tripleton? X*)	Returns true iff the set *X* has exactly three elements
Functions on sets	
(*set-difference X Y*)	Returns the set that results from removing *Y* from *X*
(*union X Y*)	Returns the union of sets *X* and *Y*
(*intersection X Y*)	Returns the intersect of sets *X* and *Y*
(*select X*)	Returns a set containing a single element from *X*
Logical functions	
(*and P Q*)	Returns TRUE if *P and Q* are both true
(*or P Q*)	Returns TRUE if either *P or Q* is true
(*not P*)	Returns TRUE iff *P* is false
(*if P X Y*)	Returns *X* iff *P* is true, *Y* otherwise
Functions on the counting routine	
(*next W*)	Returns the word after *W* in the counting routine
(*prev W*)	Returns the word before *W* in the counting routine
(*equal-word? W V*)	Returns TRUE if *W* and *V* are the same word
Recursion	
(*L S*)	Returns the result of evaluating the entire current lambda expression on set *S*

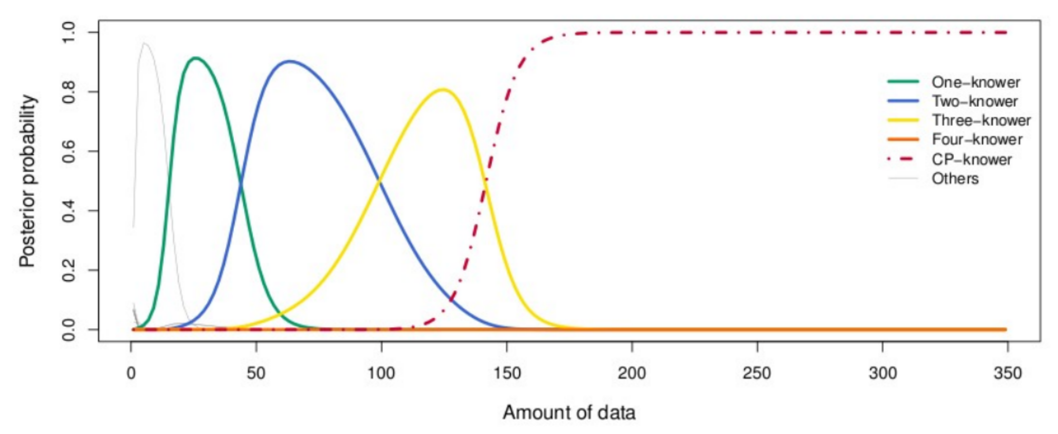
(Piantadosi, Tenenbaum & Goodman, 2012)

Number Words: Environment



(Piantadosi, Tenenbaum & Goodman, 2012)

Number Words: Results



(Piantadosi, Tenenbaum & Goodman, 2012)

Summary

- In word learning, children face a generalization problem: they need to map words to concepts.
- The have inductive biases which make the problems easier: whole world bias, taxonomic bias, basic level bias.
- We can combine knowledge about the environment, inductive biases and learning to model how children acquire word meanings.
- We illustrated this for number word learning by combining Bayes Rule with a simplicity prior.