Computational Cognitive Science

Lecture 12: Lexicon Acquisition

Chris Lucas
(Slides adapted from S. Goldwater & F. Keller's)

School of Informatics
University of Edinburgh
clucas2@inf.ed.ac.uk

31 October, 2017

Reading: Frank, Goodman, and Tenenbaum (2009).

Word Learning

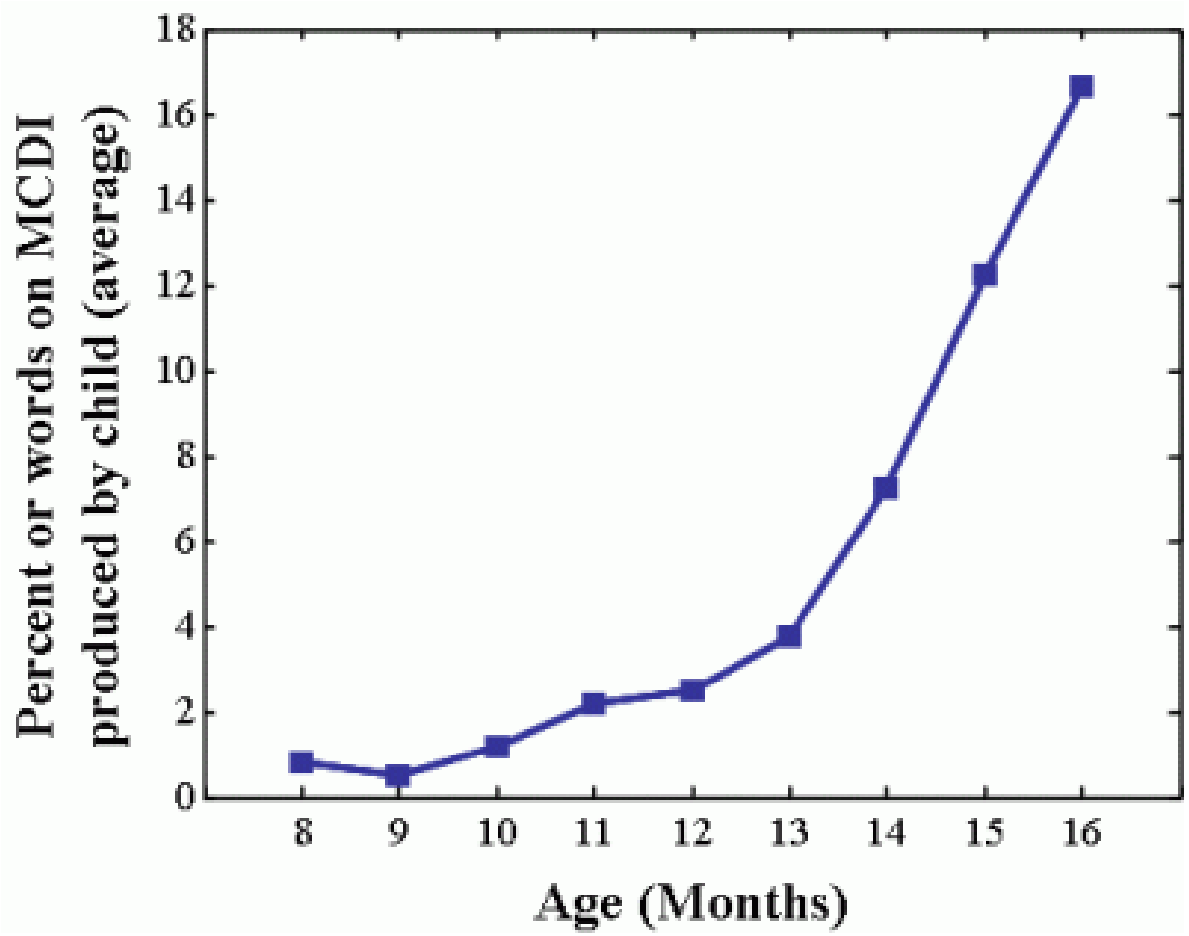
In the last lecture, we discussed how words are processed (word recognition). But these models can't explain how words are learned, i.e., *lexicon acquisition:*

- between birth and adulthood, children learn about 60,000 words (8–10 words per day on average);
- during the second postnatal year, word learning accelerates dramatically: *vocabulary explosion;*
- often children learn a new word based on a single example of its use: *one-trial learning.*

How is this possible?

Word Learning

Growth curve from the MacArthur-Bates Communicative Development Inventory (MCDI):



http://www2.psychology.uiowa.edu/faculty/mcmurray/recent/science/

Referential Ambiguity

Philosophers have argued that this should be impossible (Quine, 1960):

"gavagai!"

rabbit

undetached rabbit parts

food

a momentary rabbit-stage

there will be a storm tonight

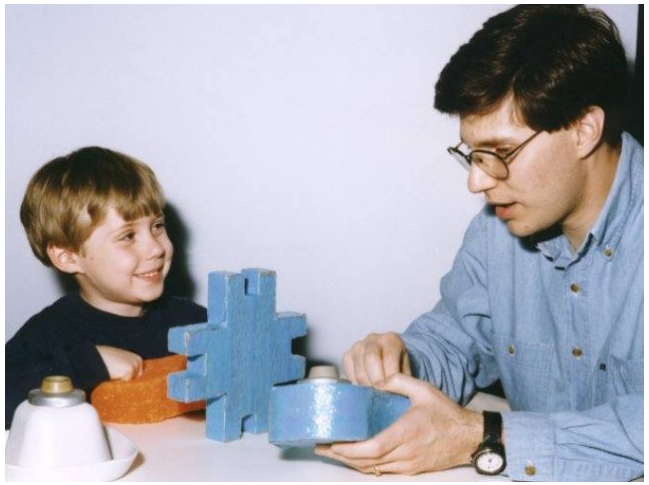
Developmental psychologists think:

- *referential ambiguity* is a serious problem, but:
- concrete nouns are learned first, then verbs, adjectives, abstract nouns follow;
- coincides with the development of syntax (two-word stage);
- one-trial learning happens when a new word is encountered in a context in which the other objects are familiar.

Social Learning

Crucially, children rely on *social context* to learn words. They infer other peoples' intentions based on:

- eye gaze;
- body position;
- pragmatics/saliency.



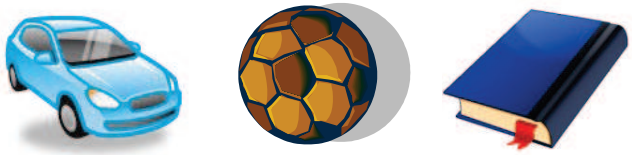
Evidence for this comes from:

- apparent lack of learning from video;
- tracking of others' gaze at six months;
- learning new words using gaze at 18 months.

Image: http://www-cogsci.ucsd.edu/~deak/cdlab/research.html

Cross-situational Learning

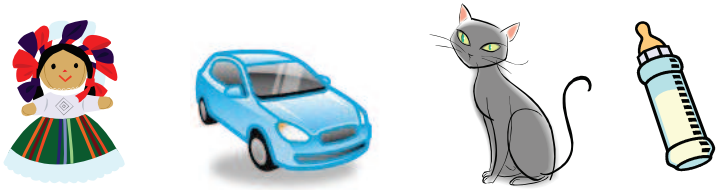
Cross-situational learning resolves referential ambiguity over time:



...car...book...

...ball...cat...car...

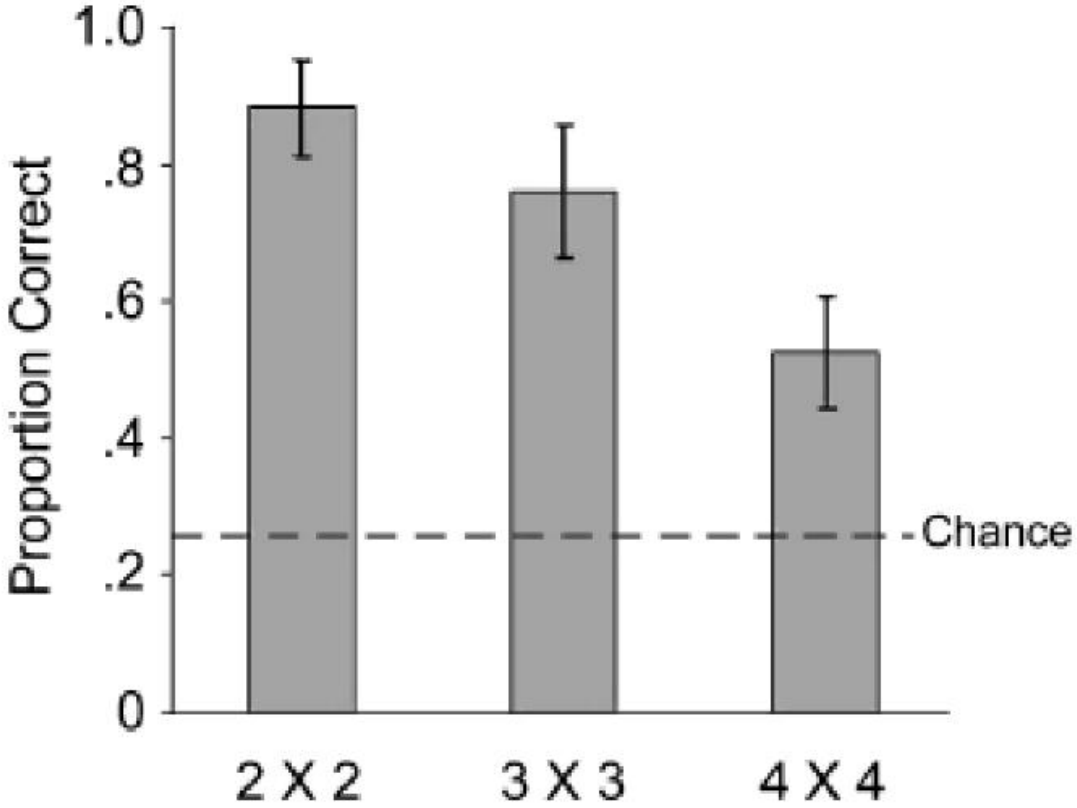
...book...car...bottle...



...car...bottle...cat...

Cross-situational Learning

Evidence: both adults and infants can learn word meanings from cross-situational experience in the lab (Yu & Smith, 2007):



$n \times n$ condition: trial presented n words and n possible referents.

Bayesian Formulation

Frank et al., 2009 propose a Bayesian model of word learning:

- combines cross-situational learning and inferring speaker's intention;
- learns a lexicon L from a video corpus C annotated with objects present and words spoken.

Goal: infer a lexicon given a corpus:

$$P(L|C) \propto P(C|L)P(L)$$

Here, L is a set of word-object pairs. $P(C|L)$ is defined using generative model, and $P(L)$ favors smaller L:

$$P(L) \propto e^{-\alpha|L|}$$

Generative Model

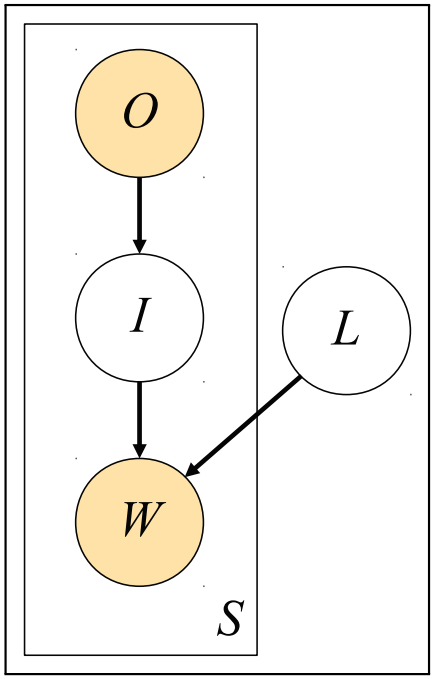
The corpus consists of independent situations. For each situation $s \in C$:

- O is the set of objects present in the situation;
- chose $I \subseteq O$, the set of intended referents (objects the speakers wants to talk about);
- chose $W \subseteq L$, the words the speaker utters (the lexicon L includes referential and non-referential words).

In practice, situation = utterance.

Graphical model notation:

- empty/white-background circle: hidden random variable
- shaded/colored circle: observed random variable
- arrow: conditional dependence
- plate: replicated S times.



Generative Model

$$
\begin{aligned}
P(C|L) &= \prod_{s \in C} P(O_s, W_s|L) \\
&= \prod_{s \in C} \sum_{I_s \subseteq O_s} P(O_s, I_s, W_s|L) \\
&= \prod_{s \in C} \sum_{I_s \subseteq O_s} P(O_s) P(I_s|O_s) P(W_s|I_s, L) \\
&\propto \prod_{s \in C} \sum_{I_s \subseteq O_s} P(I_s|O_s) P(W_s|I_s, L)
\end{aligned}
$$

Generative Model

Generate intentions from objects: uniform distribution:

$$P(I_s|O_s) \propto 1$$

Generate words from intentions and lexicon: words are independent. For each word w in W_s:

- choose referential ($p = \gamma$) or non-referential ($p = 1 - \gamma$):

$$P(W_s|I_s, L) = \prod_{w \in W_s} \left[\gamma \sum_{o \in I_s} \frac{1}{|I_s|} P_R(w|o, L) + (1 - \gamma) P_{NR}(w|L) \right]$$

- $P_R(w|o, L)$: choose uniformly from lexical items that refer to correct object;
- $P_{NR}(w|L)$: choose uniformly from all words in corpus.

Data

Small hand-annotated corpus:

- two 10-minute videos of mothers playing with preverbal infants using a small number of toys;
- speech is transcribed;
- each line of transcription is annotated with all visible mid-size objects, and true intention (for evaluation only).

words	"do bunnies go jumping through the forest?"
objects	BOOK, BIRD, RATTLE, MIRROR, BUNNY
intention	BUNNY

Evaluation

Evaluate model predictions against:

- gold-standard lexicon (word-object pairings);
- gold-standard intentions for each utterance (coded manually).

Compute *precision* (proportion of pairings that were correct) and *recall* (proportion of total correct pairings that were found).

Compare against related models:

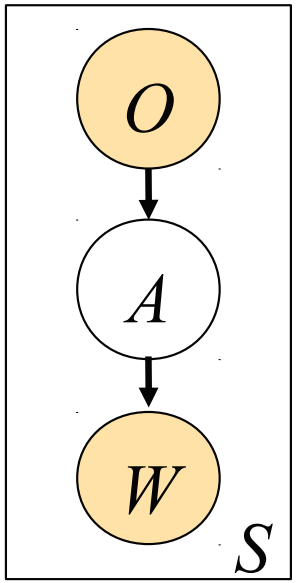
- simple statistics (co-occurrence frequency, conditional probability, mutual information);
- cross-situational model without intentions: IBM Machine Translation Model 1 (associative model).

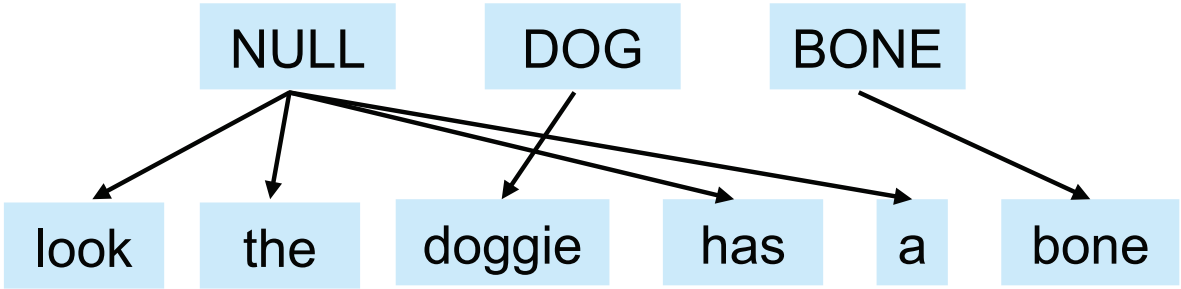
Machine Translation Model

Given objects $O = o_1 \dots o_J$ in current situation:

- choose number of words K;
- choose an alignment $a_1 \dots a_K$ between objects and words (including special NULL object);
- for each k, choose w_k given object that is aligned to it.

Model is asymmetric; try both $O|W$ and $W|O$.





Results: Lexicon Accuracy

Model	Precision	Recall	F score
Association frequency	.06	.26	.10
Conditional probability (object|word)	.07	.21	.10
Conditional probability (word|object)	.07	.32	.11
Mutual information	.06	.47	.11
Translation model (object|word)	.07	.32	.12
Translation model (word|object)	.15	.38	.22
Intentional model	**.67**	**.47**	**.55**
Intentional model (one parameter)	.57	.38	.46

(Frank et al., 2009)

Results: Lexicon Accuracy

Word	Object
bear	**bear**
bigbird	**bird**
bird	**duck**
birdie	**duck**
book	**book**
bottle	bear
bunnies	**bunny**
bunnyrabbit	**bunny**
hand	**hand**
hat	**hat**
hiphop	mirror
kittycat	**kitty**
lamb	**lamb**
laugh	cow
meow	baby
mhmm	hand
mirror	**mirror**
moocow	**cow**
oink	pig
on	ring
pig	**pig**
put	ring
ring	**ring**
sheep	**sheep**

(Frank et al., 2009)

Results: Referent Accuracy

Model	Precision	Recall	F score
Association frequency	.27	**.81**	.40
Conditional probability (object|word)	.59	.36	.45
Conditional probability (word|object)	.32	.79	.46
Mutual information	.36	.37	.37
Translation model (object|word)	.57	.41	.48
Translation model (word|object)	.40	.57	.47
Intentional model	**.83**	.45	**.58**
Intentional model (one parameter)	.77	.36	.50

(Frank et al., 2009)

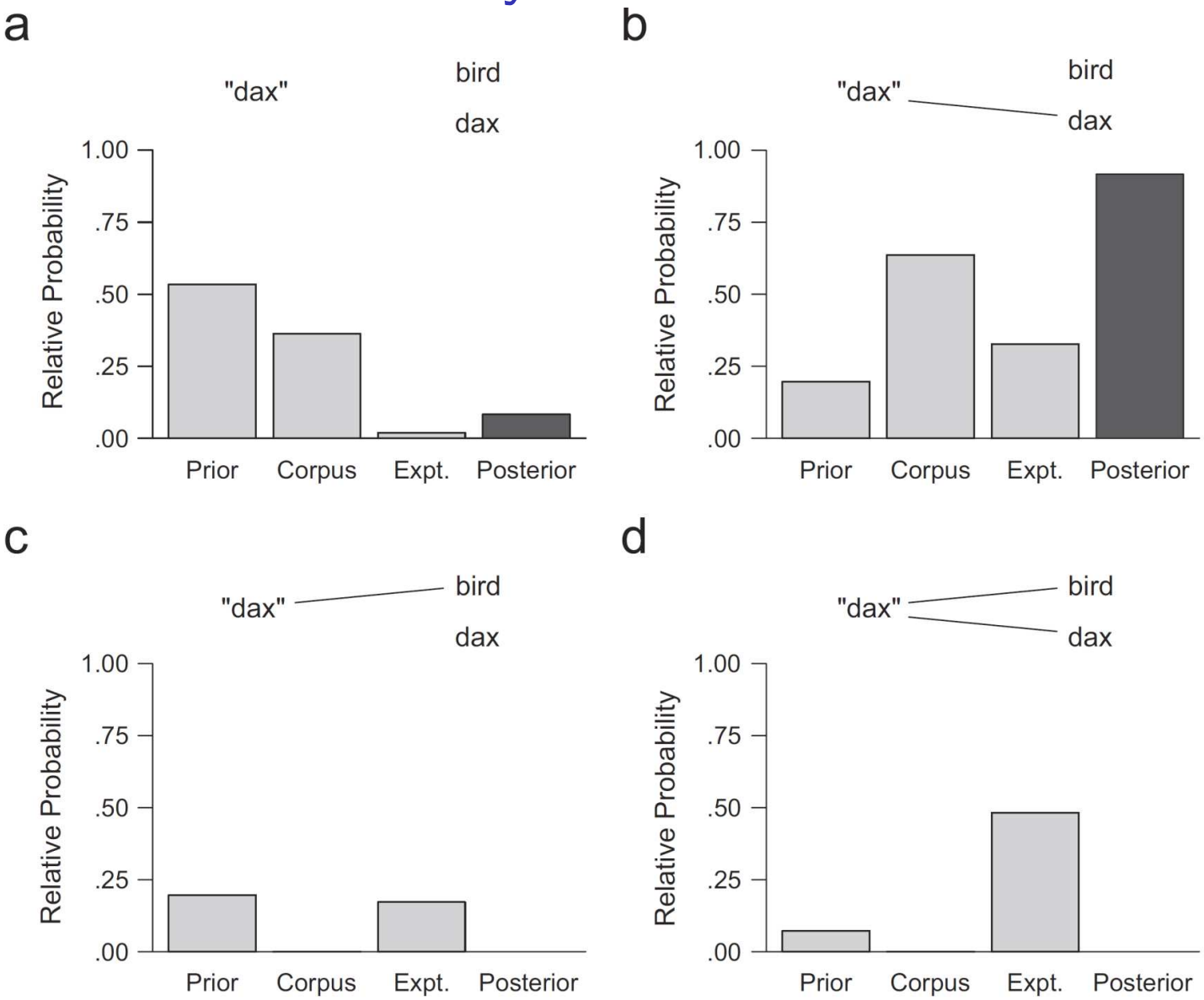
Results: Mutual Exclusivity

Children as young as 16 months map novel words to novel objects:

"Where is the dax?"

- some researchers have postulated a principle of mutual exclusivity to account for this;
- but it could also be general pragmatic principles at work;
- is mutual exclusivity learned or innate?

Results: Mutual Exclusivity



(Frank et al., 2009)

Results: Mutual Exclusivity

The model is able to capture mutual exclusivity:

- mapping "dax" to BIRD is unlikely:
  - highly coincidental that no other BIRDs are "dax";
  - likelihood is low;
- prior favors not mapping "dax" to anything, but this lowers the probability of the corpus;
- many of the other models also predict mutual exclusivity, suggesting no special principle is needed.

This example also shows that the model captures *one-trial learning.*

Results: Object Individuation

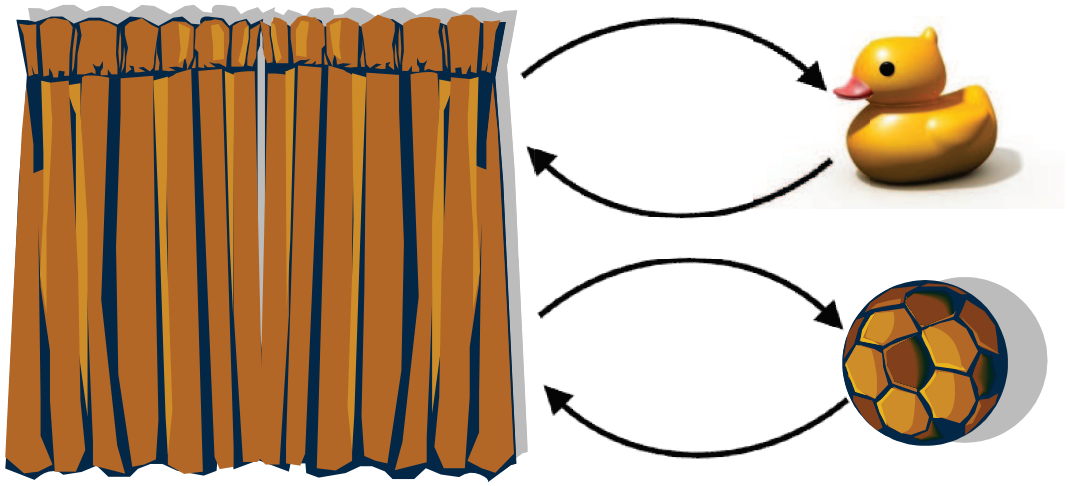
Infants as young as 12 months can use words to individuate objects. Experiment Xu, 2002:

1. an object emerges behind a screen, they hear a word, the object disappears again;
2. a different object emerges and disappears, they either hear the same word or a different word;
3. the screen disappears revealing either one or two objects.

Children show longer looking times when the number of objects in step 3 mismatches the number of words in steps 1 and 2.

Results: Object Individuation

1.



"Look, a toy!"

2.

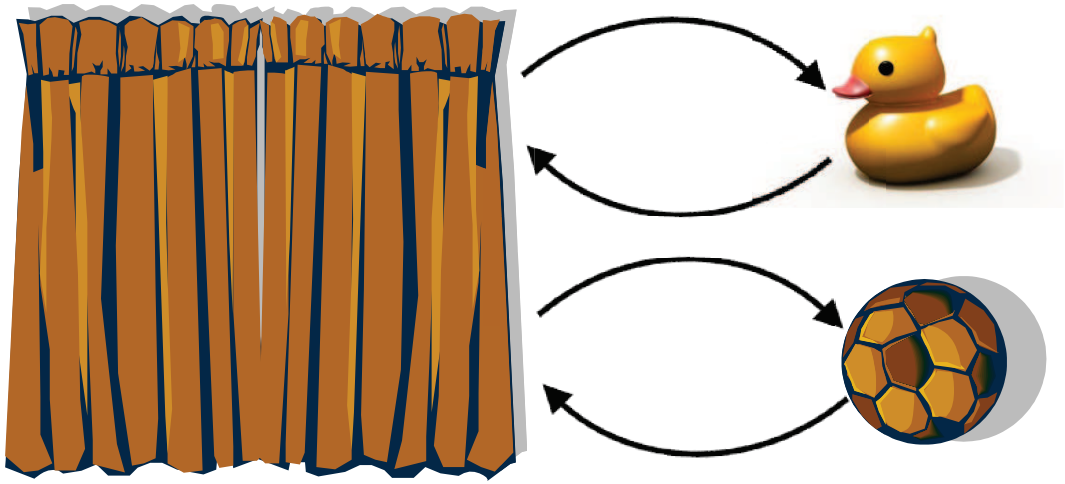
"Look, a toy!"

3.

vs.

Results: Object Individuation

1.



"Look, a duck!"

2.

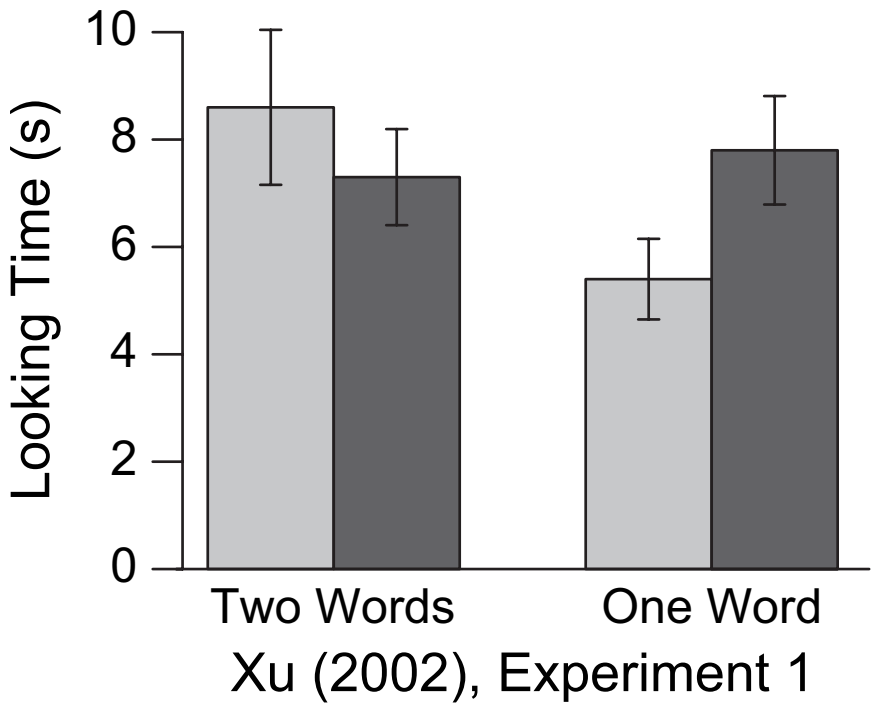
"Look, a ball!"

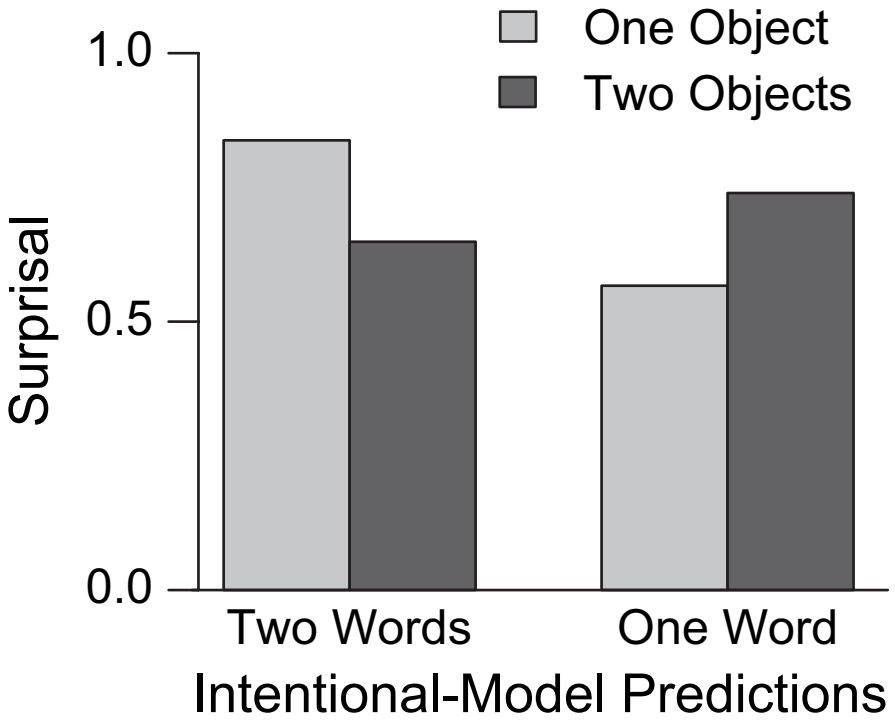
3.

vs.

Results: Object Individuation

This can be captured by the model in terms of surprisal (more on surprisal later):





Discussion

Strengths:

- model combines cross-situational and social/intentional learning;
- more accurate learning of lexicon and referents than previous models;
- explains various experimental phenomena without special principles.

Weaknesses:

- only tested on very small corpus;
- only deals with concrete nouns;
- no model of syntax.

Summary

- Infants learn words based on multiple cues (eye gaze, body position, pragmatics/salience);
- encountering a word in multiple situations is crucial for word learning;
- Frank et al.'s model combines visual information (objects) and gaze (intended referents) to infer a mapping from words to meanings;
- it aggregates information across situations;
- it outperforms a simple translation (alignment) model and captures mutual exclusivity and object individuation.

References

- Frank, M. C., Goodman, N., & Tenenbaum, J. (2009). Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science*, *20*, 579–585.
- Quine, W. (1960). *Word and object*. Cambridge, MA: MIT Press.
- Xu, F. (2002). The role of language in acquiring object concepts in infancy. *Cognition*, *85*, 223–250.
- Yu, C. & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, *18*(5), 414–420.