A Probabilistic Model of Cross-situational Word Learning from Noisy and Ambiguous Data

Afra Alishahi

Joint work with Afsaneh Fazly and Suzanne Stevenson,
University of Toronto

Word Learning

- Word learning: a mapping between a word and its "meaning".

apple ⟺

- Mappings are learned from exposure to word usages in utterances that describe scenes.

the chimp eats apples

Challenges: Referential Uncertainty

- Which aspect of a scene is described by a corresponding utterance?

a black chimp is sitting on a rock

the chimp eats apples

there are two red apples in his hands

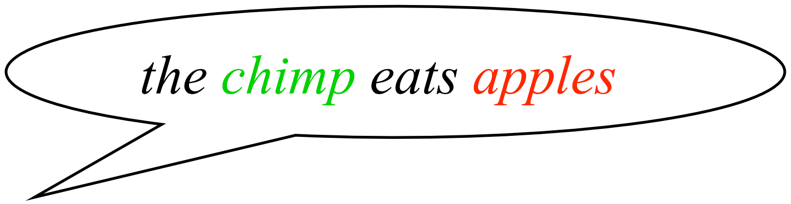
Challenges: Ambiguity

- What word refers to what part of the meaning?

the chimp eats apples

Challenges: Ambiguity

- What word refers to what part of the meaning?



{black, animal, living, chimp, eyes, hands, feet, red, apple, fruit, edible, food, rock, object, green, leaf, action, consume, sit, hold, …}

Cross-situational Learning

- Meaning of a word is learned by detecting meaning elements of a scene in common across several usages of the word. [Pinker89]

the chimp eats apples

daddy is picking apples

A Detailed Account of Word Learning

- Cross-situational learning does not explain various patterns observed in children, such as vocabulary spurt and fast mapping. [e.g., Reznick et. al'92; Carey'78]

- Many specific principles are proposed to explain each pattern, e.g., mutual exclusivity or a change in the learning mechanism. [e.g., Markman et. al'88]

- A unified model of word learning is needed to account for all observed patterns.
  - Computational implementation allows for the evaluation of such a model in a naturalistic setting.

Our Goals

- Implement an incremental probabilistic account of cross-situational learning.
- Explain observed patterns without incorporating mechanisms specific to each phenomenon.
- Handle referential uncertainty and ambiguity.
- Learn word–meaning mappings from naturally occurring child directed utterances.

Input to the Model

- Input is a sequence of utterance–scene pairs:

utterance		scene representation
"the chimp eats an apple"	⟺	{black, animal, living, chimp, eyes, hands, feet, red, apple, fruit, edible, food, rock, object, green, leaf, action, consume, sit, hold, …}

- Meaning of each word is represented as a set of semantic features.

Overview of the Learning Algorithm

- An adaptation of a model for finding corresponding words between sentences in two languages.

  [Brown et al.'93]

- Each input pair is processed in two steps:
  - use previously learned meaning associations to align each word in utterance with meaning elements from the scene.
  - use these alignments to update the (probabilistic) association between a word and its meaning elements.

Formal Definitions

- Alignment probabilities:

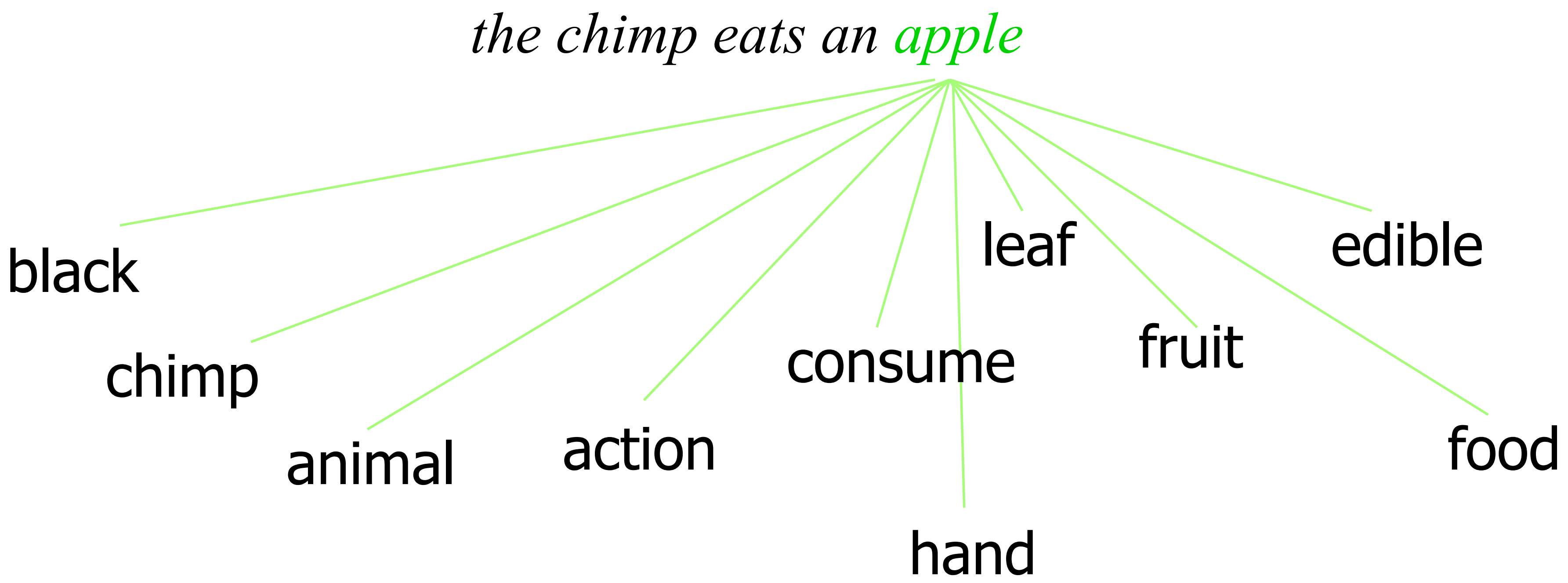
$$a(w \mid m, U^{(t)}) = \frac{p^{(t-1)}(m \mid w)}{\sum_{w_k \in U^{(t)}} p^{(t-1)}(m \mid w_k)}$$

- Meaning probabilities:

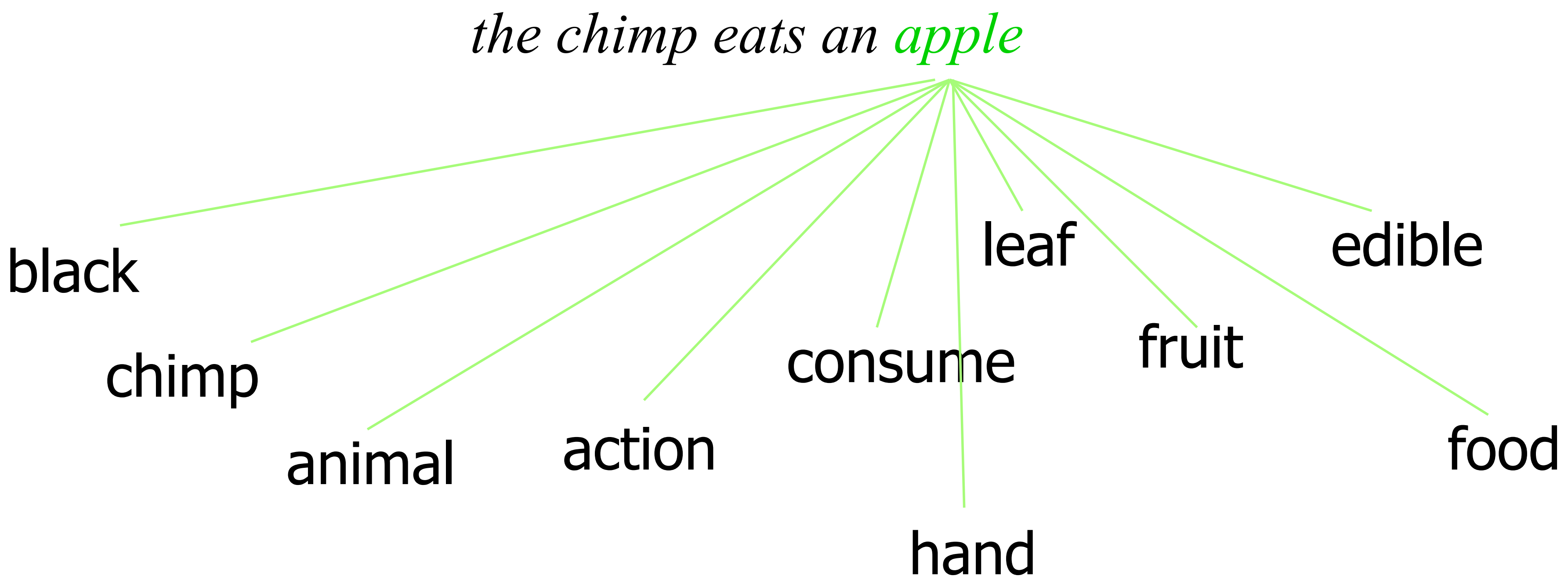
$$p^{(t)}(m \mid w) = \frac{\sum_{s=1}^{t} a(w \mid m, U^{(s)}) + \lambda}{\sum_{m_j \in M} \sum_{s=1}^{t} a(w \mid m_j, U^{(s)}) + \beta \times \lambda}$$

An Example

apple ?



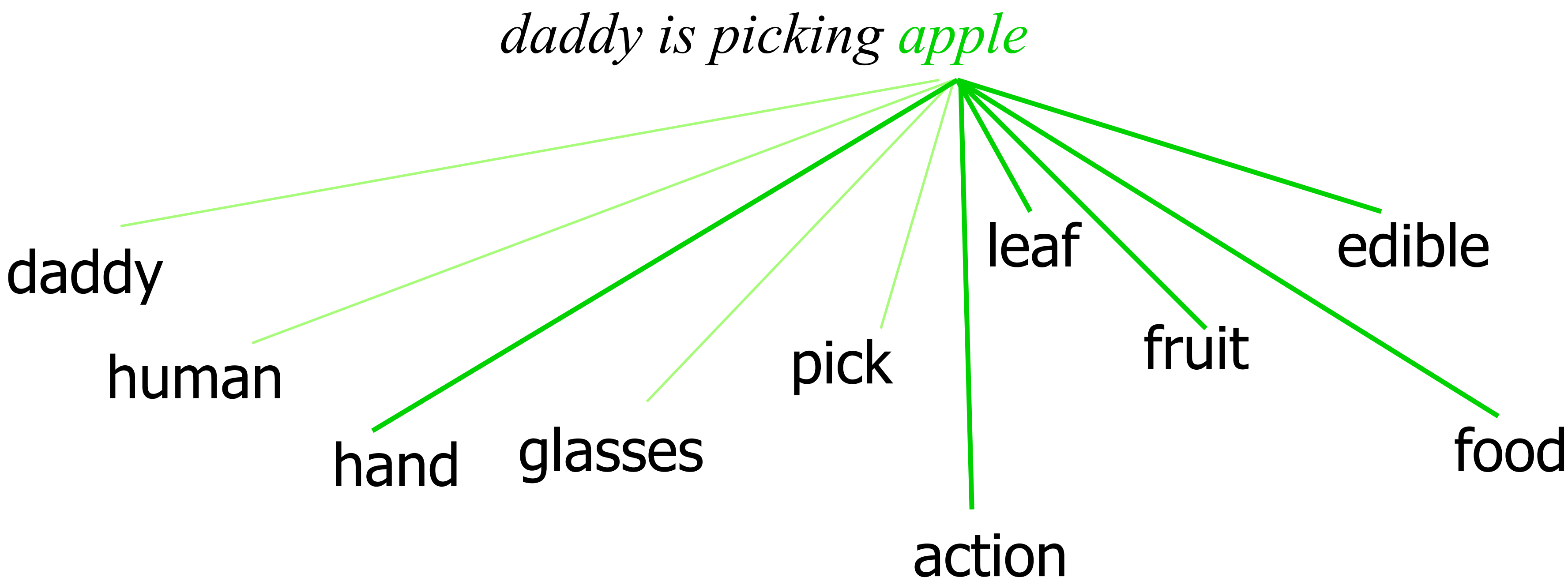
An Example



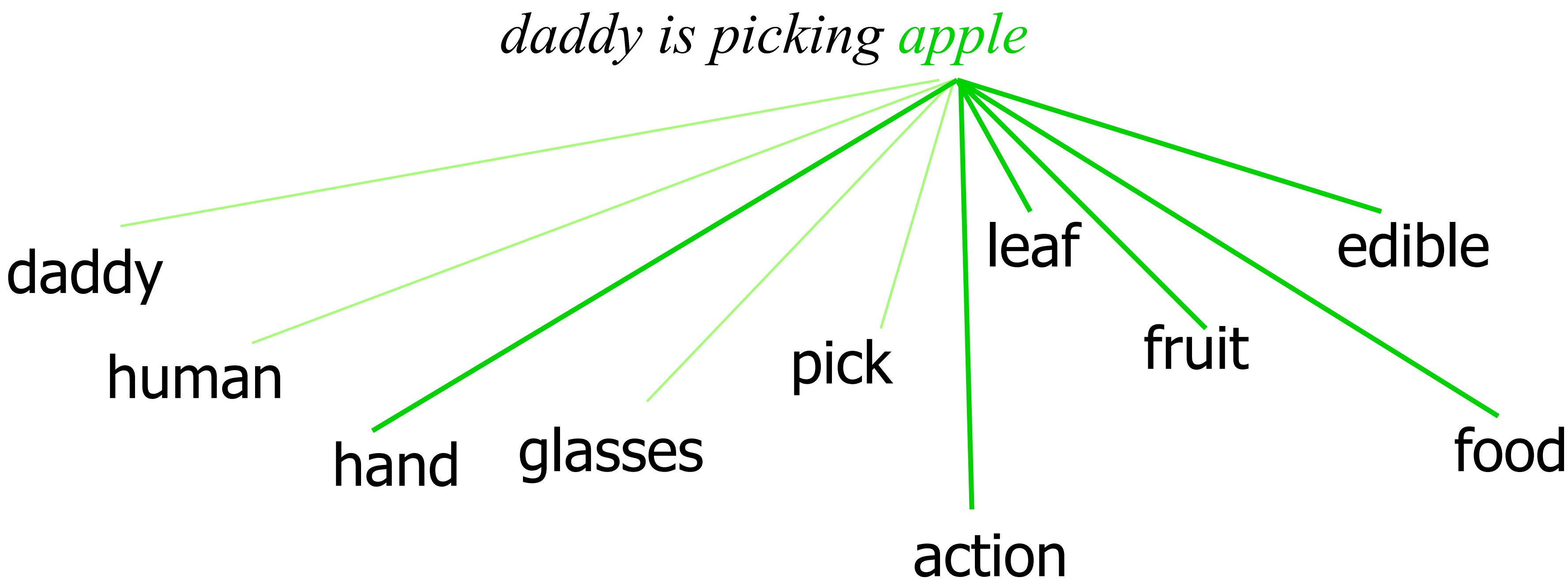
apple black chimp animal action consume hand leaf fruit food edible ...

An Example

apple black chimp animal action consume hand leaf fruit food edible ...



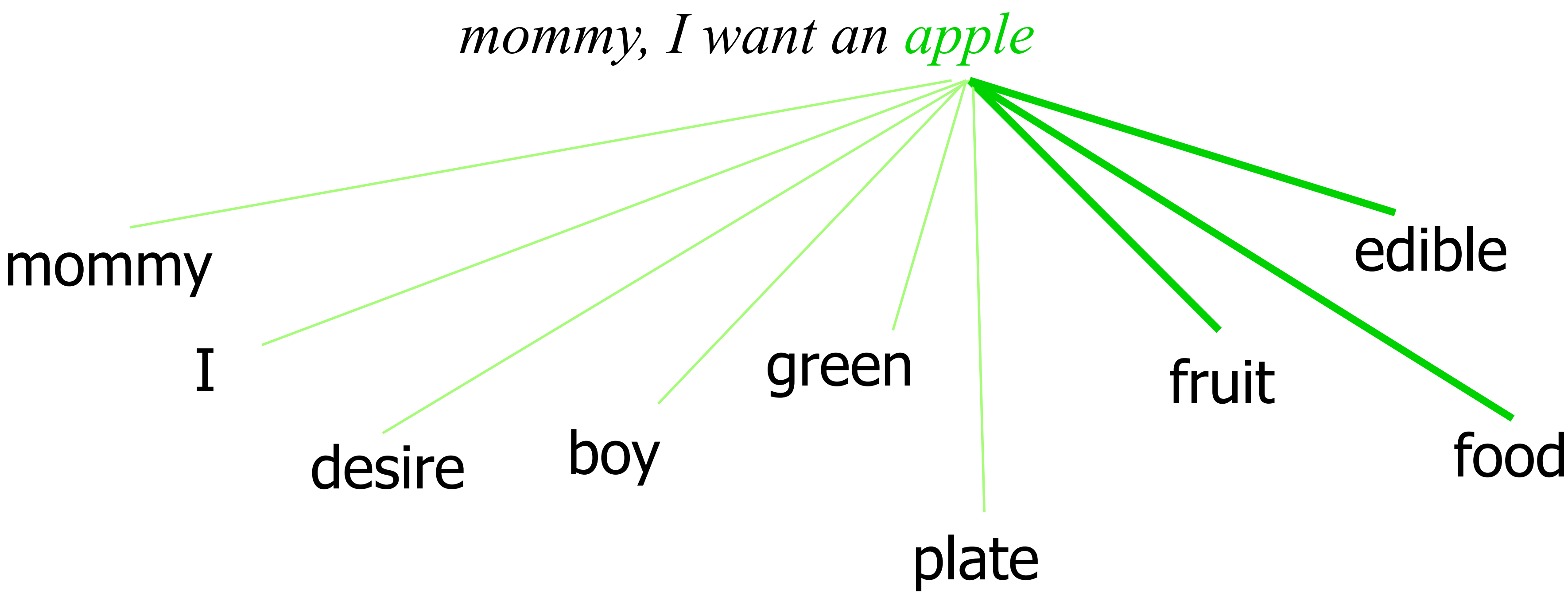
An Example



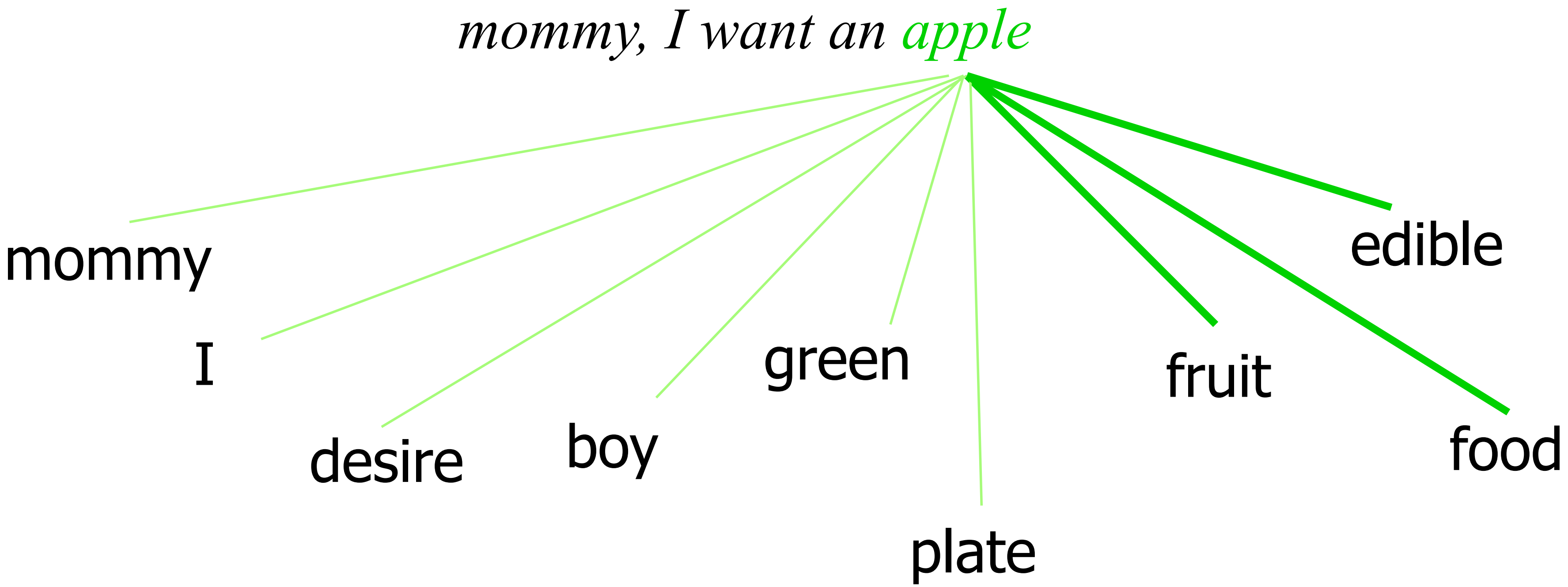
apple black chimp animal action consume hand leaf fruit food edible daddy human glasses pick …

An Example

apple black chimp animal action consume hand leaf fruit food edible daddy human glasses pick …



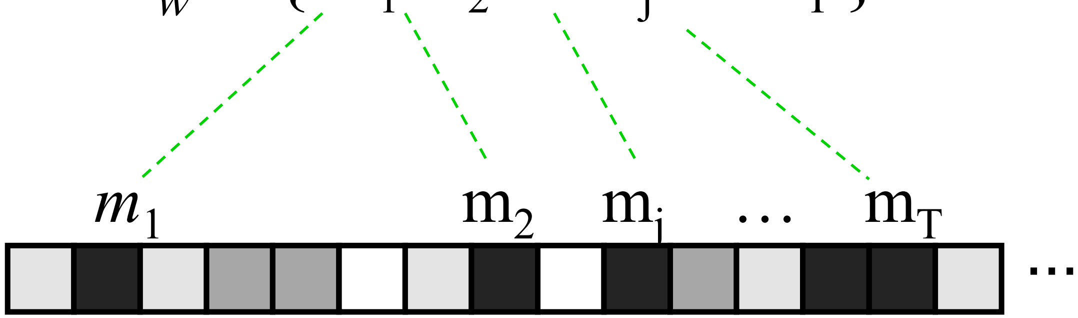
An Example



apple black chimp animal action consume hand rock leaf fruit food edible daddy human glasses pick mommy I desire plate green …

When is a Word "Learned"?

- A word is learned when most of its probability mass is concentrated on its correct meaning elements.
  - correct: $T_w = \{ m_1 \; m_2 \; \ldots \; m_j \; \ldots \; m_T \}$
  - learned:



- Comprehension score:

$$c^{(t)}(w) = \sum_{m_j \in T_w} p^{(t)}(m_j \mid w)$$

Data: Input Corpora

- Utterances from Manchester corpus in CHILDES database: [Theakston et. al'01; MacWhinney'95]

that is an apple

do you like apple?

do you want to give dolly an apple?

can teddy bear give penguin a kiss?

.

.

.

Data: Input Corpora

- … paired with meaning primitives extracted from WordNet and a resource by Harm (2002):

that is an apple	definite, be, edible, fruit, …
do you like apple?	do, person, you, desire, edible, fruit, …
do you want to give dolly an apple?	do, person, you, want, location, physical property, artifact, object, …
can teddy bear give penguin a kiss?	artifact, object, teddy, animal, bear, touch, deed, …
.	.
.	.
.	.

Data: Input Corpora

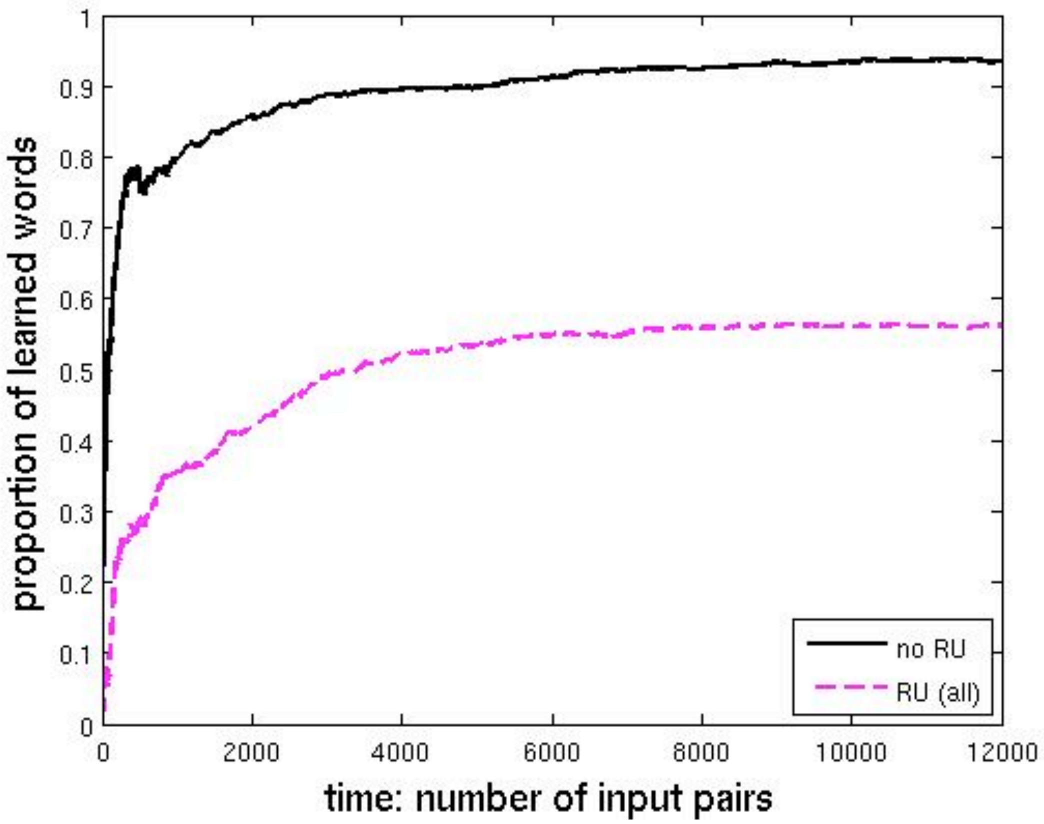
- … and subsequent primitive sets are combined to simulate referential uncertainty:

that is an apple	definite, be, edible, fruit, …
do you like apple?	do, person, you, desire, edible, fruit, …
do you want to give dolly an apple?	do, person, you, want, location, physical property, artifact, object, …
can teddy bear give penguin a kiss?	artifact, object, teddy, animal, bear, touch, deed, …

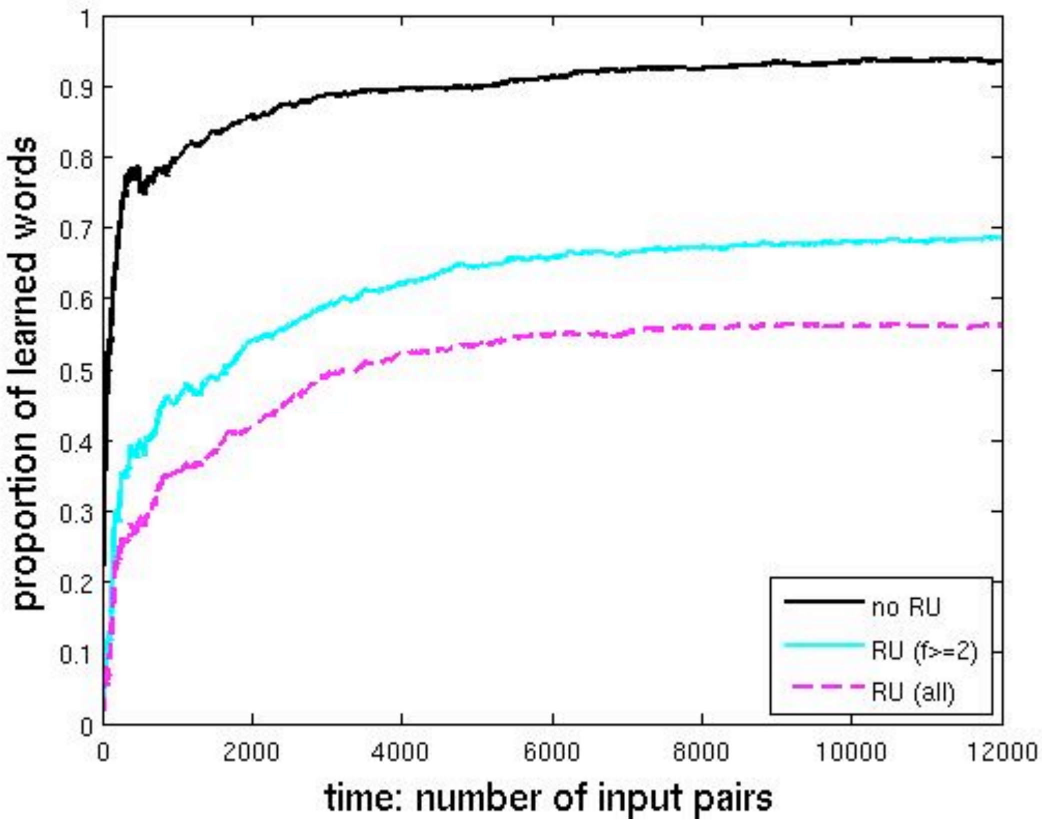
⋮ ⋮

Learning Rates: Referential Uncertainty

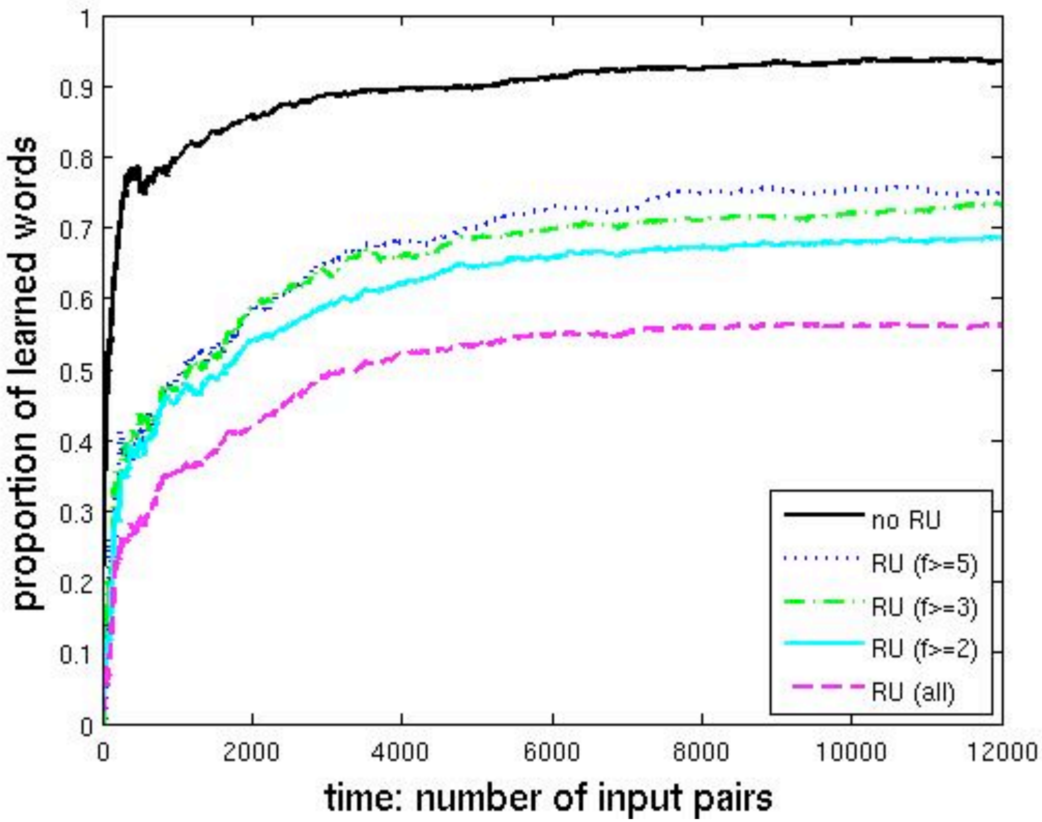
- Change in proportion of learned words over time:



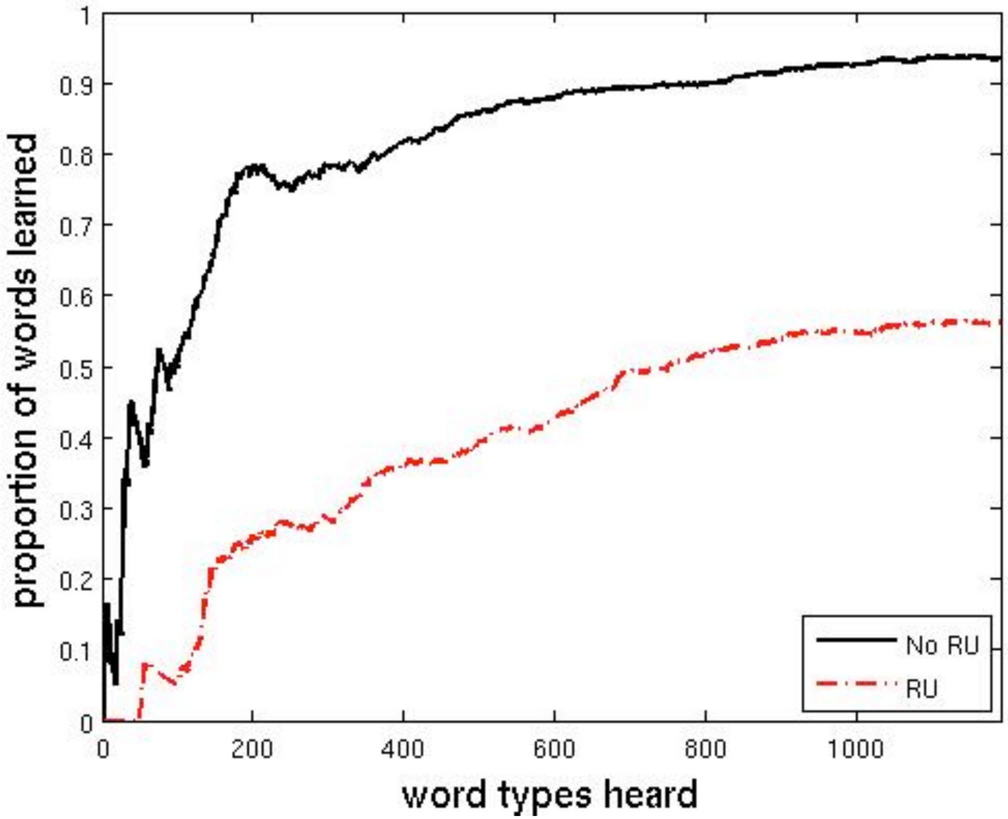
Learning Rates: Effect of Frequency



Learning Rates: Effect of Frequency



Vocabulary Spurt



- We observe a sudden increase in learning rate; no change in the learning mechanisms is needed.

Fast Mapping [Carey'78]

Can you show me the dax?

Fast Mapping [Carey'78]

Can you show me the dax?

- Young children can easily determine the meaning of a novel word if used in a familiar context.
  - referent selection

Fast Mapping and Word Leaning

What is this?

Fast Mapping and Word Leaning

What is this?

- Not clear whether children "learn" the meaning of a fast-mapped word.
  - retention (through comprehension or production)

Possible Explanations

- Fast mapping is due to a specialized mechanism for word leaning:
  - e.g., mutual exclusivity, novel name—nameless category, switching to referential learning.

[Markman & Wachtel'88; Golinkoff et al.'92; Gopnik & Meltzoff'87]

- Fast mapping arises from general processes of learning and communication:
  - e.g., induction using knowledge of acquired words, inference on the intent of the speaker.

[Clark'90; Diesendruck & Markson'01, Halberda'06]

An Example

- Input: a sequence of utterance–scene pairs:

"the chimp eats an apple"	⟺	{ THE, CHIMP, EAT, AN, APPLE, SIT, ON, ROCK, HAND, LEAF }
"daddy is picking apple"	⟺	{ DADDY, PICK, APPLE, TREE, SUNGLASSES, LEAF }
"see the apple on the rock"	⟺	{ SEE, THE, RED, APPLE, ON, ROCK, GREEN, PLATE}

- Output: a probability distribution over meaning elements:

apple

Referent Selection

- Familiar target:

give me the apple

- Novel target:

give me the dax

➢ Different mechanisms might be at work in the two conditions.

[Halberda'06]

Referent Selection

- Familiar target:

give me the apple

Referent Selection

- Familiar target:

give me the apple

  - correct referent is selected upon hearing target word

Referent Selection

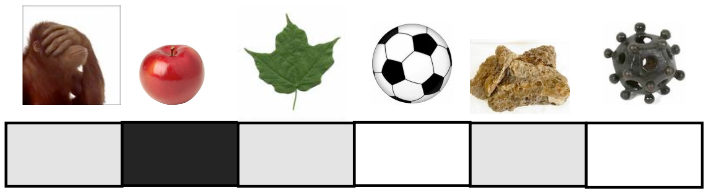
- Familiar target:

give me the *apple*

  - correct referent is selected upon hearing target word

- Use meaning probability $p(.|apple)$

apple



Referent Selection

- Familiar target:

give me the *apple*

  - correct referent is selected upon hearing target word

- Use meaning probability $p(.|apple)$

$p($ $\|apple)$	$p($ $\|apple)$
0.8430±0.056	« 0.0001

Referent Selection

- Novel target:

give me the dax

Referent Selection

- Novel target:

give me the *dax*

  - correct referent is selected by performing induction

Referent Selection

- Novel target:

give me the dax

  - correct referent is selected by performing induction

- Meaning probabilities are not informative:

dax						

Referent Selection

- Novel target:

give me the dax

  - correct referent is selected by performing induction

- Meaning probabilities are not informative:

dax						

- Use referent probability $rf(dax\,|.)$:

$$rf(w\,|\,m) = \frac{p(m\,|\,w)\,p(w)}{\sum_{w' \in V} p(m\,|\,w')\,p(w')}$$

Referent Selection

- Novel target:

give me the dax

  - correct referent is selected by performing induction

- Use referent probability $rf(dax\,|.)$:

$rf(dax|$  $)$	$rf(dax$  $)$
0.127±0.127	0.993±0.002

Retention (2-OBJECT)

- Referent Selection Trial (1):

  give me the dax


- Referent Selection Trial (2):

  give me the cheem


- Retention Trial:

  give me the dax


Retention (2-OBJECT)

- Perform induction over recently-acquired knowledge about the meaning of the two novel words:

rf (*dax*|)	*rf* (*dax*|)
0.996±0.001	0.501±0.068

- The model correctly maps dax to its referent.

Retention (3-OBJECT)

- Referent Selection Trial (1):

give me the *dax*

- Referent Selection Trial (2):

give me the *cheem*

- Retention Trial (w/ a third unfamiliar object):

give me the *dax*

Retention (3-OBJECT)

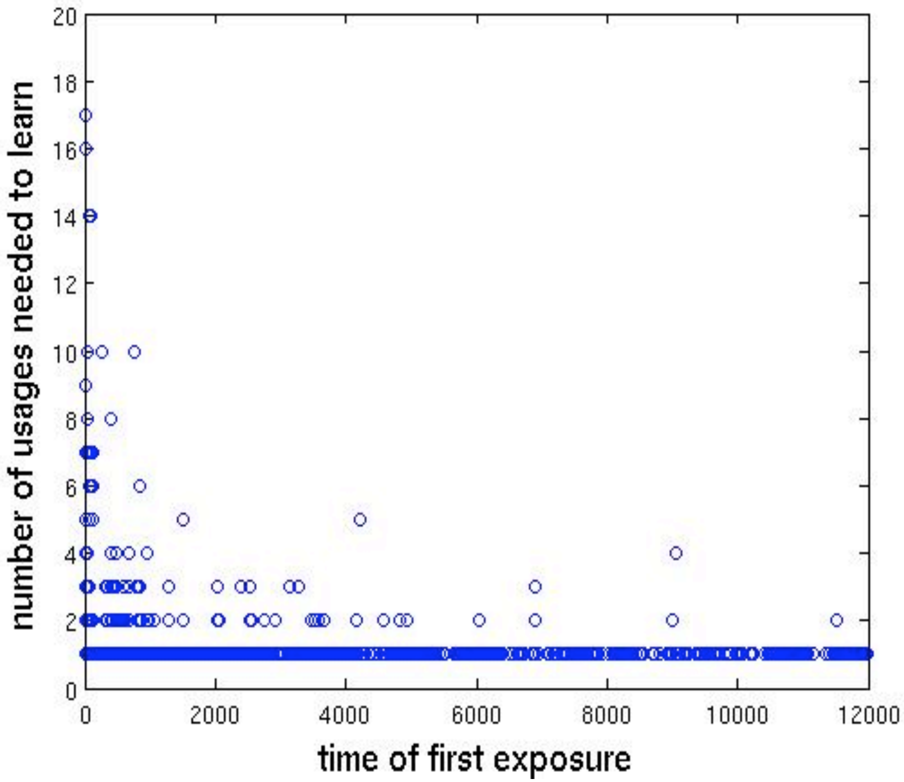
- Induction over two recently fast-mapped and one novel object:

$rf(dax \mid$  $)$	$rf(dax \mid$  $)$	$rf(dax \mid$  $)$
0.995±0.001	0.407±0.062	0.990±0.001

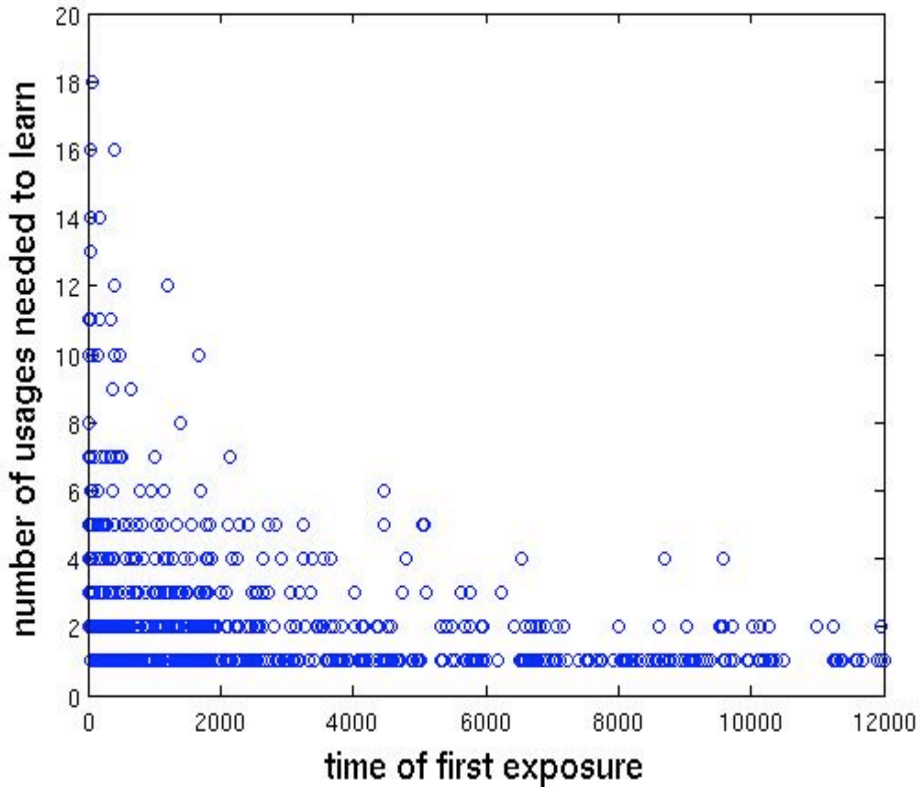
- The presence of a third novel object can be confusing, as also seen in experiments on children.

Fast Mapping Effects

no RU



RU



- Mapping novel words to their meanings becomes easier with more exposure to input.

Summary and Future Directions

- Developed an incremental probabilistic model that
  - learns word–meaning mappings from naturalistic data, in the face of ambiguity and referential uncertainty.
  - incorporates a single learning mechanism that accounts for many learning patterns observed in children.

- Future directions:
  - study the role of syntax in word learning.
  - learn semantic and/or syntactic categories of words from the acquired word–meaning associations.