Language Acquisition
Fall 2010/Winter 2011

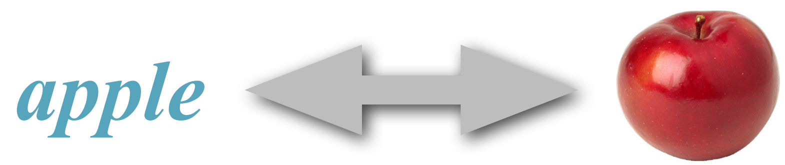
Learning Word Meanings

Afra Alishahi, Heiner Drenhaus

Computational Linguistics and Phonetics
Saarland University

Learning Words

- Learning the meaning of words: associating a mental representation, or concept, with a word form



Challenges of Word Learning

- Sentential context
    - Most words are not used in isolation, but in a a multi-word utterance
- Referential uncertainty
    - Learners may perceive aspects of a scene are unrelated to the utterance they hear
- Noise
    - Error in perception or interpretation of the heard utterance or the observed scene

Sentential Context

The chimp eats apples

Referential Uncertainty

a black chimp is sitting on a rock

The chimp eats apples

the chimp eats apples

there are two red apples in his hands

Perception Error (Noise)

The chimp eats ???

Suggested Learning Mechanisms

- Associative learning
  - Simple associative mechanisms are used to map a word form with a concept
- Referential learning
  - A variety of attention mechanisms are used to narrow down the intended meaning
- Cross-situational learning
  - Inferring correct word-meaning mappings by observing regularities across usages of a word

Associative Learning

- Ideas and experiences reinforce one another
  - a new word form may be learned through repeated association to an already learned concept
- Classic conditioning, e.g. Pavlov's dog:

$$\begin{tikzcd} \textbf{meat powder} \arrow[r, leftrightarrow] \arrow[d] & \textbf{bell ring} \arrow[d] \\ \textbf{salivation} & \textbf{salivation} \end{tikzcd}$$

Referential Learning

- Using specific biases for restricting the referents
  - Whole object bias: a novel word is likely to refer to the entirety of an object
  - Taxonomic bias: labels refer to objects of the same kind (often basic-level categories)
- Using social and visual cues
  - Joint attention through pointing or gaze helps narrow down possible referents of a novel label

Cross-situational Learning

- Detecting common meaning elements across several usages of a word:

kitty is playing with yarn

Sam is knitting the green yarn

Developmental Patterns

- Vocabulary spurt
  - Vocabulary learning is slow at the early stages, then proceeds to a rapid pace
- Fast mapping
  - Young children can map a novel word to a novel object in a familiar context
- Second labels
  - Early on, children show difficulty in learning homonymous and synonymous words (i.e., one-to-many and many-to-one mappings)

Vocabulary Spurt

- Following a slow start, rate of word learning rapidly increases
  - Usually around the time the child's vocabulary has about 50 words
- Vocabulary spurt is suggested to arise from a qualitative change in the nature of lexical acquisition, such as
  - shift from associative to referential learning
  - sudden realization that objects have names
  - development of categorization aibilities
  - onset of word learning constraints

Fast Mapping

Can you show me the dax?

- Young children can easily determine the referent of a novel word in a familiar context
- Fast mapping is attributed to a specialized mechanism:
  - principle of Mutual Exclusivity
  - bias to map novel names to nameless objects
  - change in children's underlying word learning mechanism

Second Labels

- Young children exhibit difficulty in learning synonyms
  - one-to-many and many-to-one mappings are hard at first

- Suggestions:
  - Children have an (innate or learned) bias towards one-to-one mappings
  - They must overcome this bias in order to learn synonymous and homonymous words

How to Explain These Patterns?

- A change in the underlying learning mechanism?
  - A shift from associative to referential learning
- Task-specific biases and constraints?
  - Principle of Mutual Exclusivity: words pick out mutually exclusive concept categories
  - Principle of Contrast: every two word forms contrast in meaning
  - Name-Nameless Category Principle (NjC): children tend to find names for nameless objects/categories
- Statistical properties of the input and the learning process explain the changing behaviour of children?

Computational Models of Word Learning

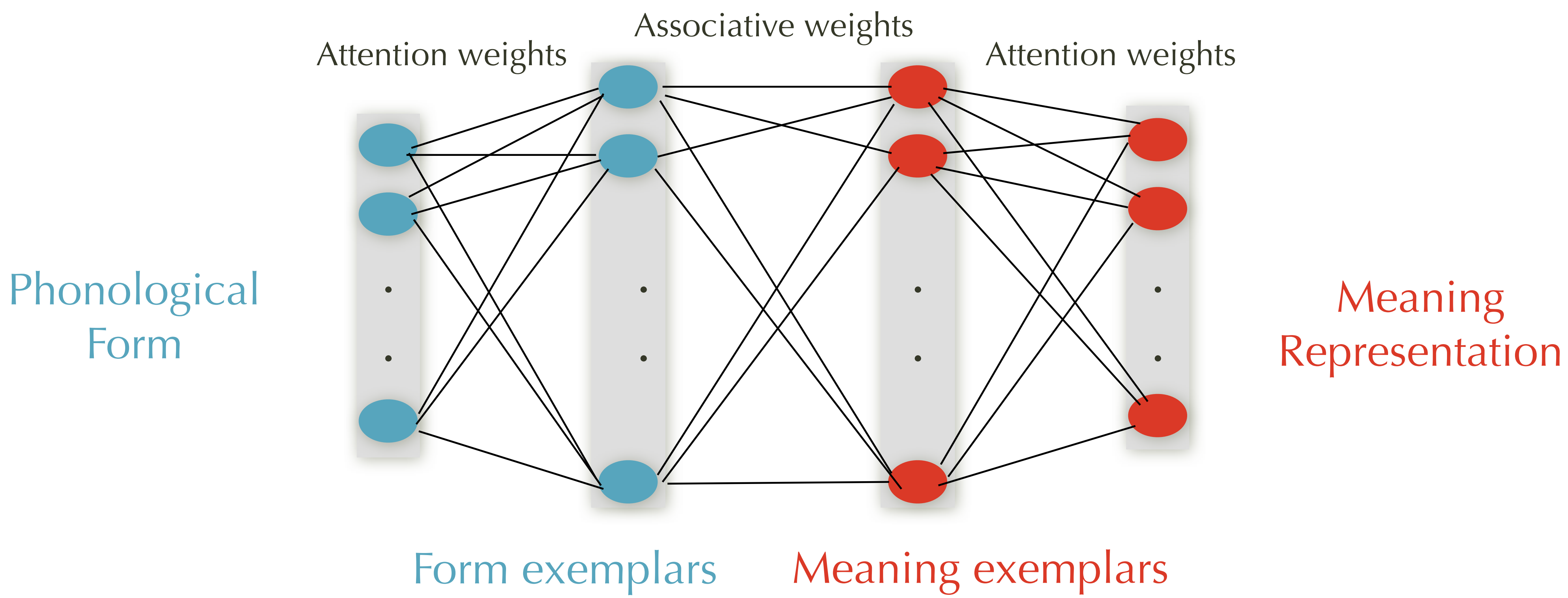
- Computational modeling is a powerful tool for investigating the hypothesized mechanisms of word learning
  - Reproduction: does the model imitate the experimental patterns observed in children?
  - Consistency: does the model need a change in the underlying mechanism to account for the observed patterns?
  - Realistic input: can the model perform on realistic data, containing noise and referential uncertainty?

Overview of the Existing Computational Models of Word Learning

- Implementing biases and constraints in a symbolic framework, using artificially generated input (e.g. Siskind 1996)
- Learning associations btw a word form and its meaning from isolated, simplified word usages, often in a connectionist framework (e.g. Regier 2005, Li et al. 2004, 2007)
- Probabilistic interpretation of cross-situational learning (e.g. Yu 2005, Fazly et al. 2008)
- Incorporating attention mechanisms such as intentional and social cues (e.g. Yu 2006, Frank et al. 2007, Yu and Ballard 2008)
- Generalizing category meaning from examples of word usages (e.g. Xu and Tenenbaum 2007)

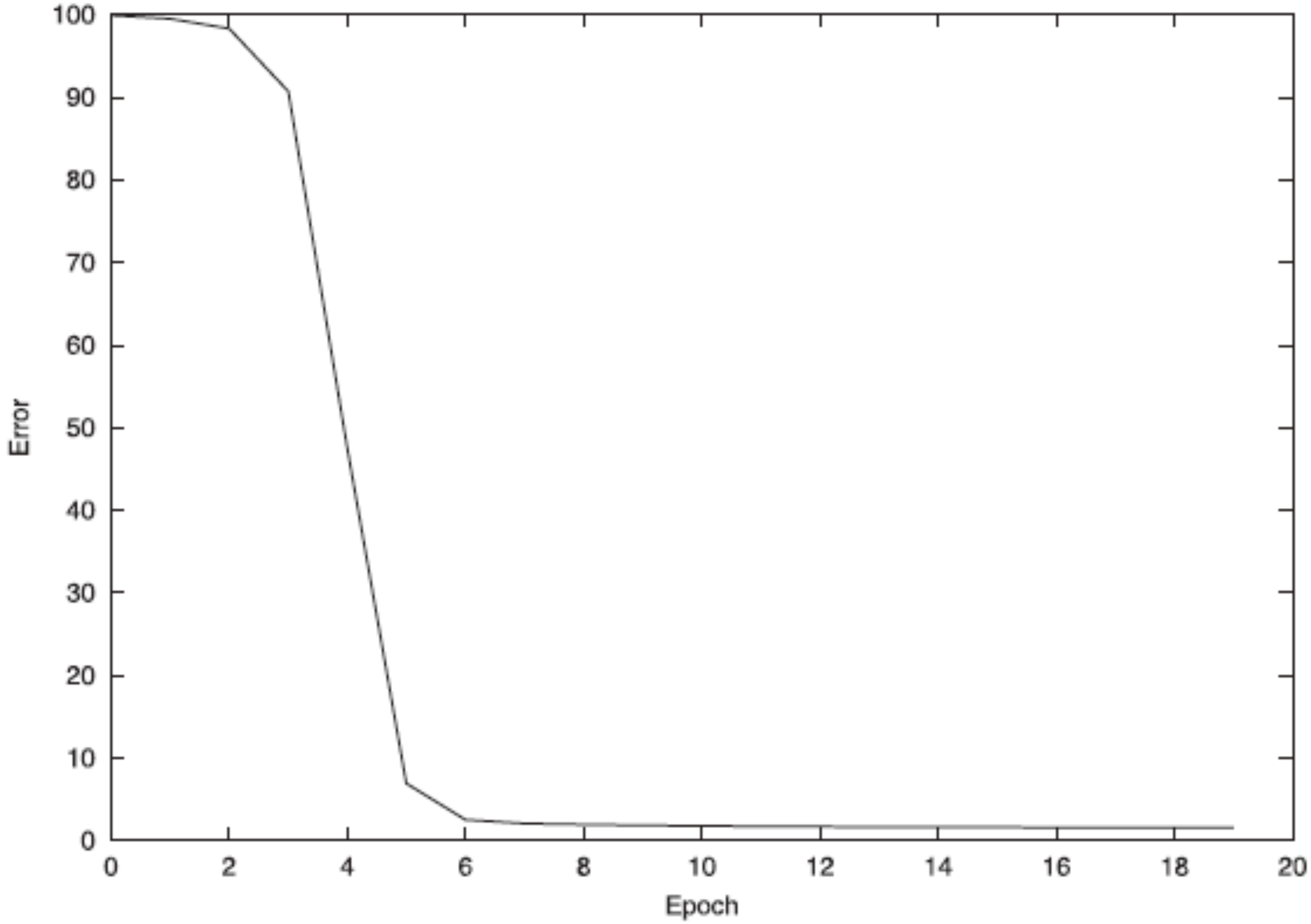
Case Study: Regier (2005)

- An associative, exemplar-based model

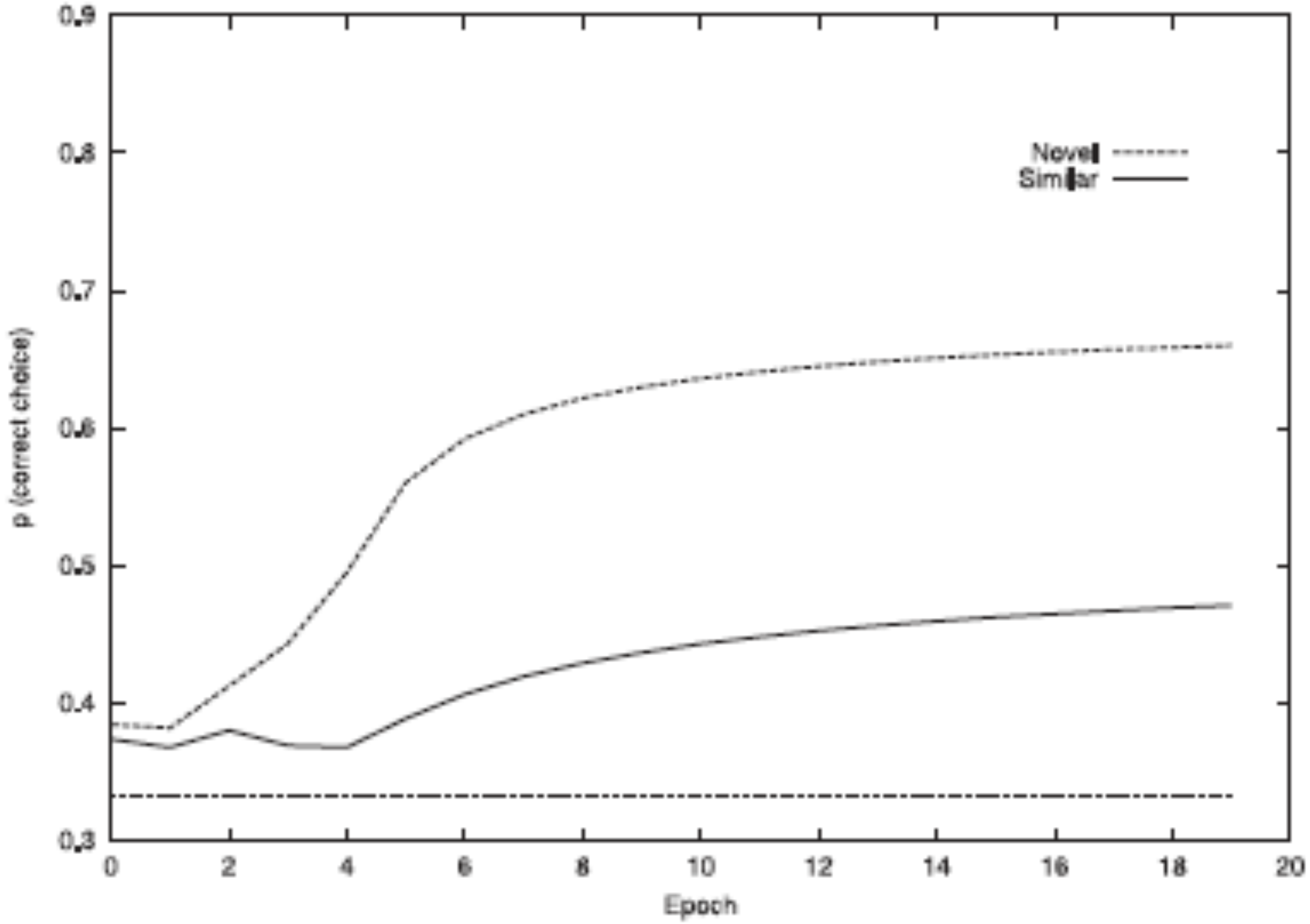


- Phonological form: a vector of phonological features (e.g. voicing)
- Meaning representation: a vector of semantic features (e.g. shape)

Regier (2005): Ease of Learning



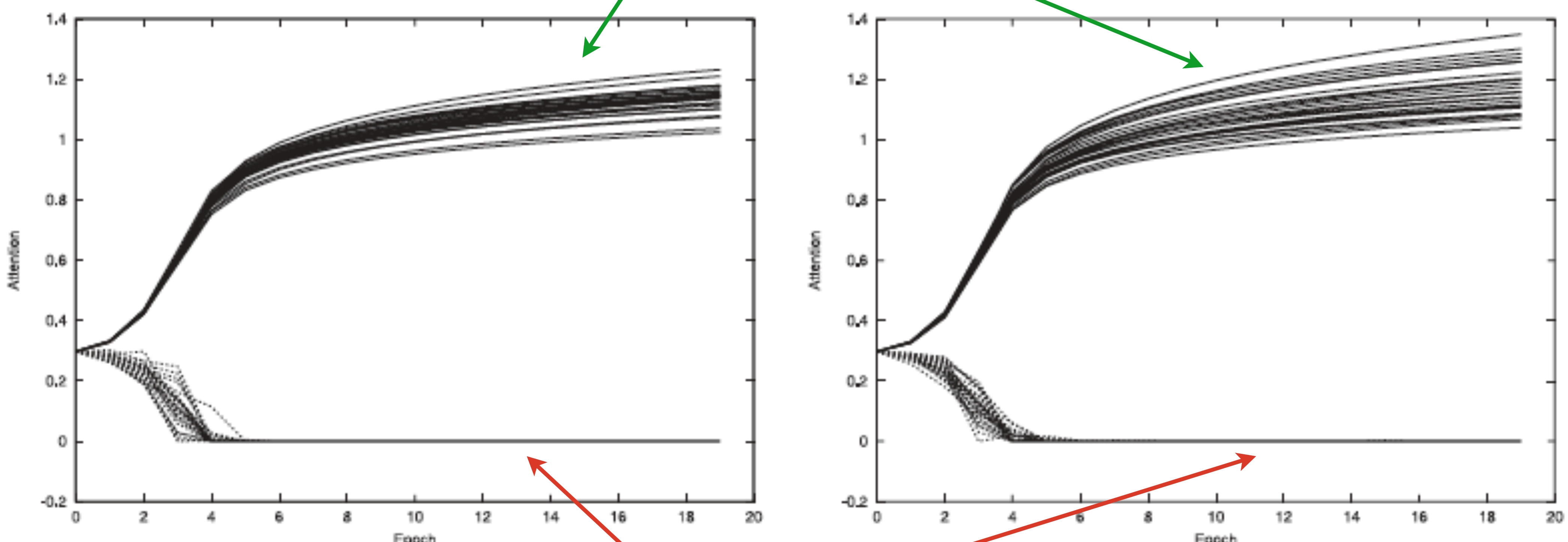
Regier (2005): Learning Second Labels



- “Similar” is a new meaning for an existing word (synonymy)

Regier (2005): Honing of Form and Meaning

Attention weights for significant features



Attention weights for non-significant features

Case Study: Siskind (1996)

- A symbolic model of cross-situational learning
  - Input: artificially generated sentence and scene representations

John went to school.	⟷	GO(**John**,TO(**school**)) MOVE(**John**,**feet**) WEAR(**John**,RED(**shirt**))

  - Meaning representation: two sets of symbols for each word

words	Necessary meanings	Possible meanings
John	{**John**}	{**John**,**ball**}
took	{CAUSE}	{CAUSE,WANT,GO,TO,**arm**}

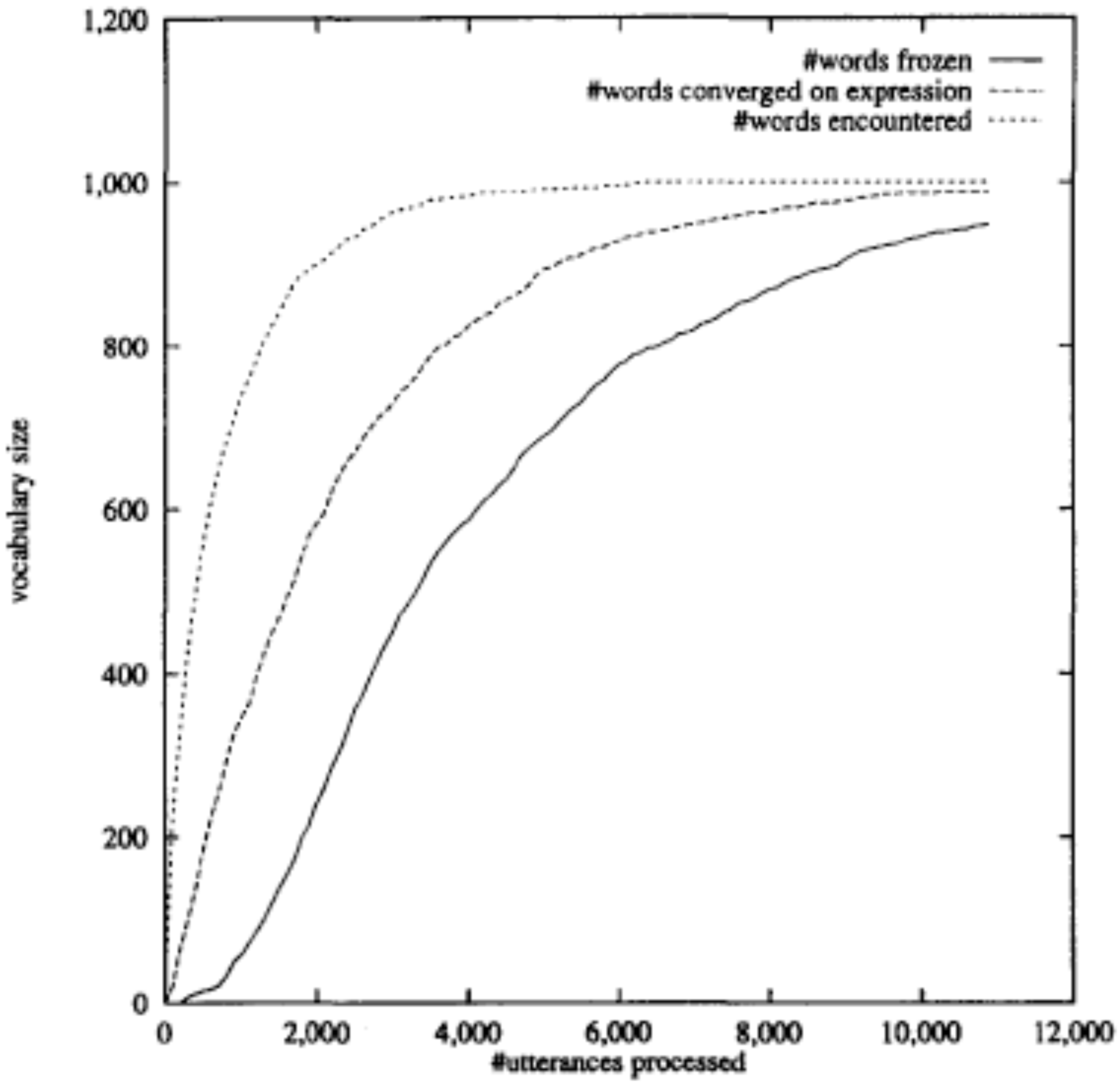
Case Study: Siskind (1996)

- Learning mechanism:
  - Start with empty N and P sets for all words
  - For each word in a new sentence update N & P according to the specific rules
  - Declare a word learned when N=P for that word

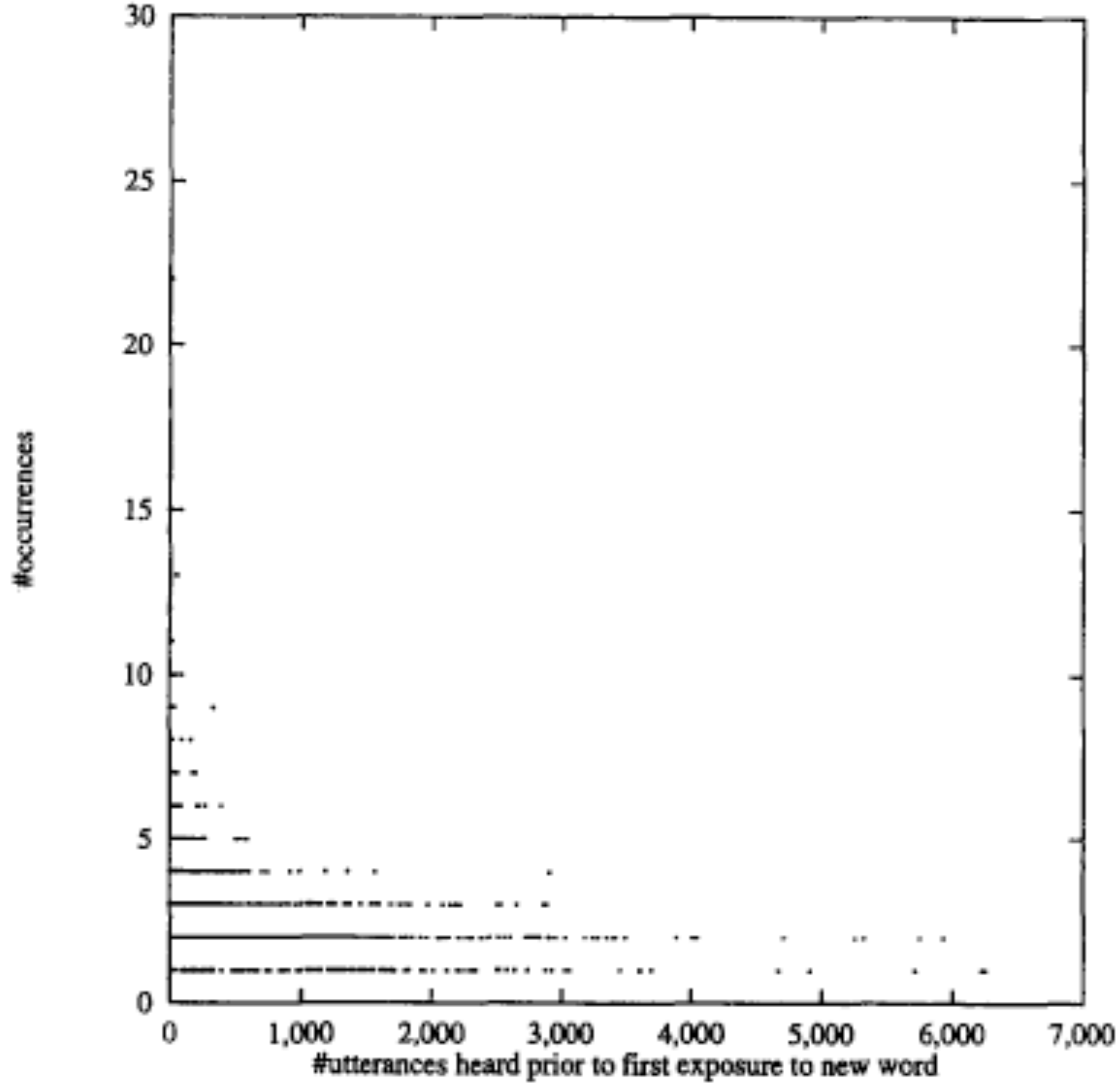
- A sample rule:

For each word symbol in the utterance, rule out any conceptual symbols that do not appear in some remaining utterance meaning

Siskind (1996): Learning Curves



Siskind (1996): Age of Exposure Effects



Case Study: Fazly et al., 2008

- A probabilistic, incremental model
  - Input: a sequence of utterance-scene pairs from CHILDES:

kitty is playing with the yarn

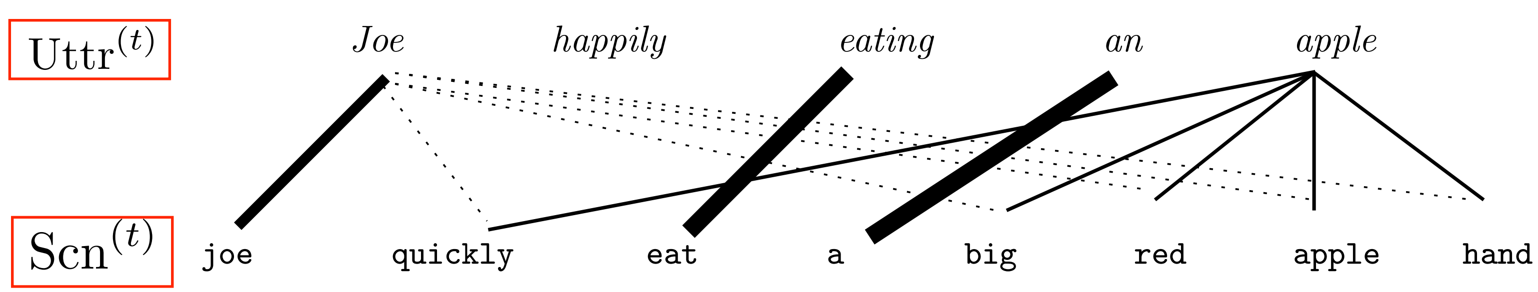
{ animal, living thing, cat, eyes, toes, tail, red, yarn, object, action, move, play, ... }

  - Meaning of a word: a probability distribution over all semantic features, or $p(.|w)$
  - Word acquisition score: a measure of how closely the meaning of a word resembles its true meaning

Fazly et al. (2008): Learning Mechanism

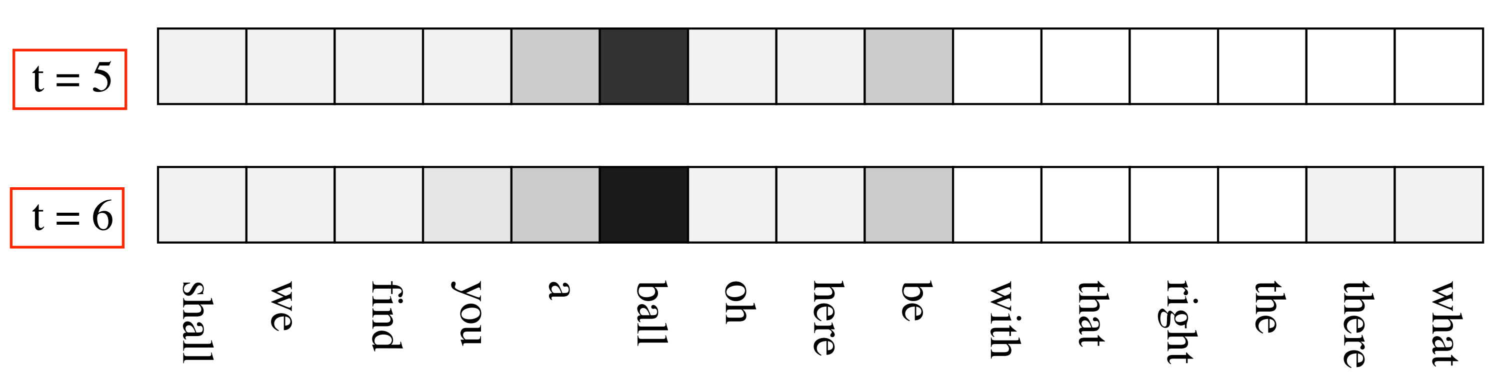
- For every new pair of scene and utterance, $(\mathrm{Uttr}^{(t)}, \mathrm{Scn}^{(t)})$

1. Alignment: use previously learned meaning associations to align each word in utterance with each meaning element from the scene



$$a(w|f,\, \mathrm{Uttr}^{(t)}) = \frac{p^{(t-1)}(f|w)}{\sum_{w_k \in \mathrm{Uttr}^{(t)}} p^{(t-1)}(f|w_k)}$$

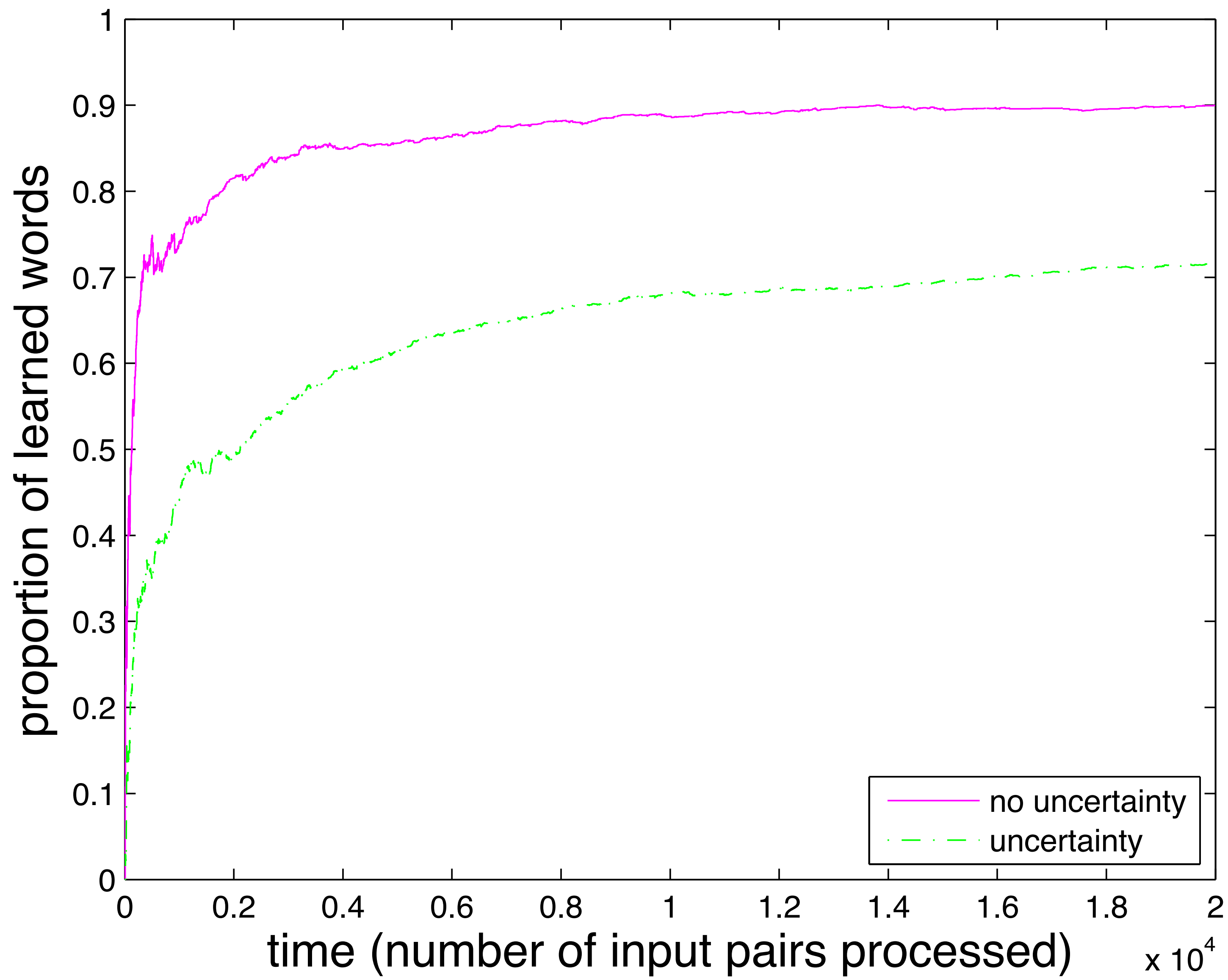
2. Update: use these alignments to update the probabilistic associations between a word and its meaning elements



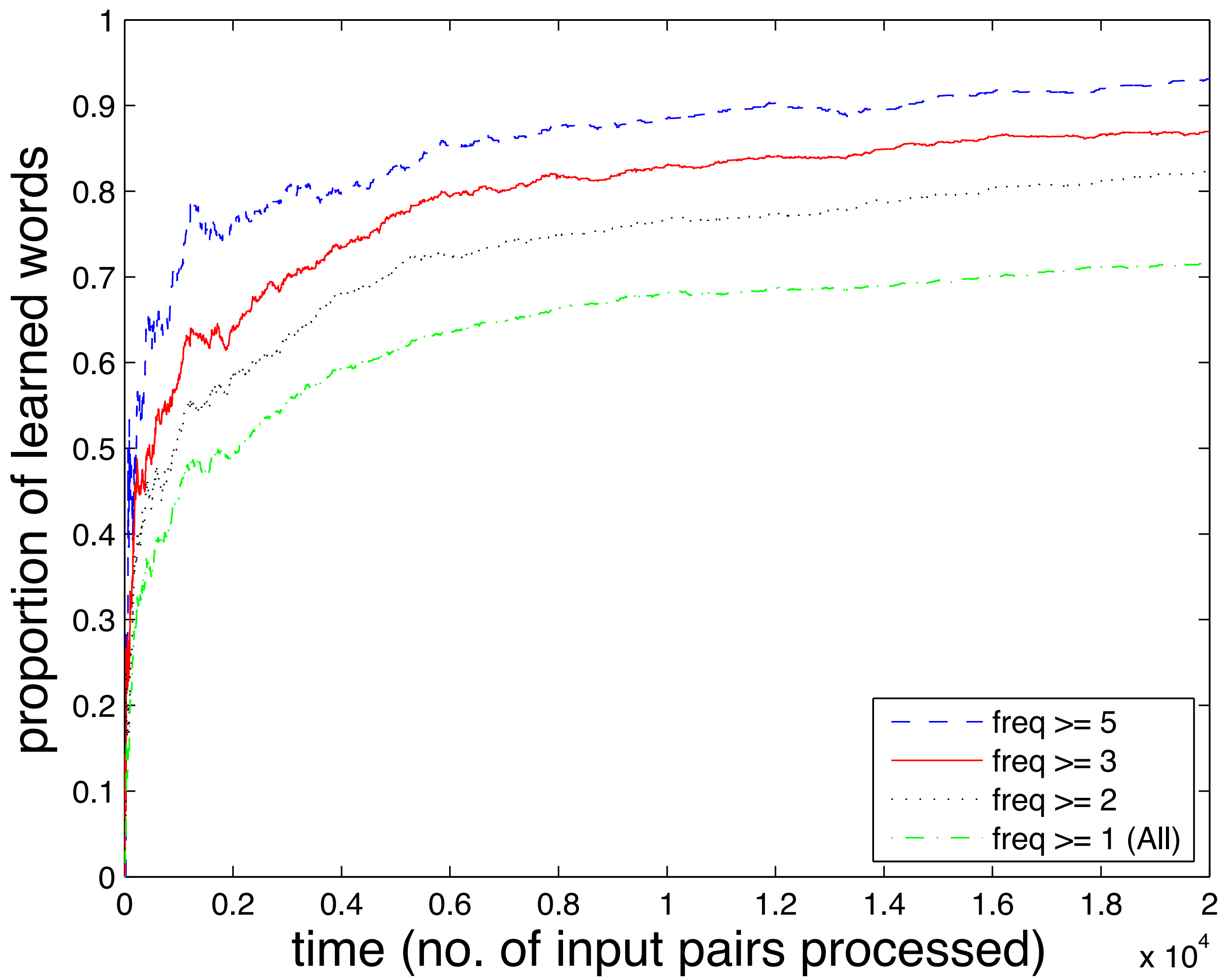
$$\mathrm{assoc}^{(t)}(w, f) = \mathrm{assoc}^{(t-1)}(w, f) + a(w|f, \mathrm{Uttr}^{(t)})$$

$$p^{(t)}(f|w) = \frac{\mathrm{assoc}^{(t)}(f,\, w)}{\sum_{f_j \in \mathcal{F}} \mathrm{assoc}^{(t)}(f_j,\, w)}$$

Fazly et al. (2008): Referential Uncertainty



Fazly et al. (2008): Frequency Effects



Role of Sentential Context

kitty is playing with yarn

He is playing with matches

Sara is cutting with scissors

Ian is washing with soap

X is DOing with Y

physical object

Syntactic Bootstrapping

- Children can learn aspects of word meaning by drawing on syntactic structure of the sentence (Gleitman, 1990)
  - E.g., differences in meaning of chase and flee cannot be fully learned through cross-situational learning
- Using syntactic structure in word learning has been computationally modeled in limited settings
  - Niyogi'02, Yu'06, Maurits et al.'09

Alishahi & Fazly (2010): Integrating Syntactic Categories

- An extension of the probabilistic model of Fazly et al (2008)
  - Integrate cross-situational and syntactic evidence
  - Assumption: the syntactic category of each word can be determined based on its context

- Input: use manually assigned PoS tags as lexical categories

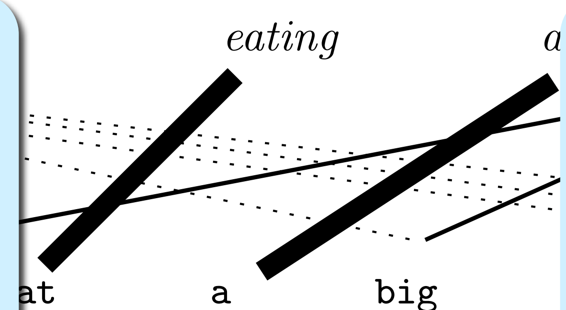
that is an apple	DET AUX DET N
do you like apple?	AUX N V N

Alishahi & Fazly: Integration Mechanism

- Aligning words and meaning elements: combine cross-situational evidence with lexical categories

Cross-situational Word Meaning

$P(f|w)$



Meaning Associated w. Lexical Category

$P(f|C)$

$$a_w(w|f,\, \mathrm{Uttr}^{(t)}) = \frac{p^{(t-1)}(f|w)}{\sum_{w_k \in \mathrm{Uttr}^{(t)}} p^{(t-1)}(f|w_k)}$$

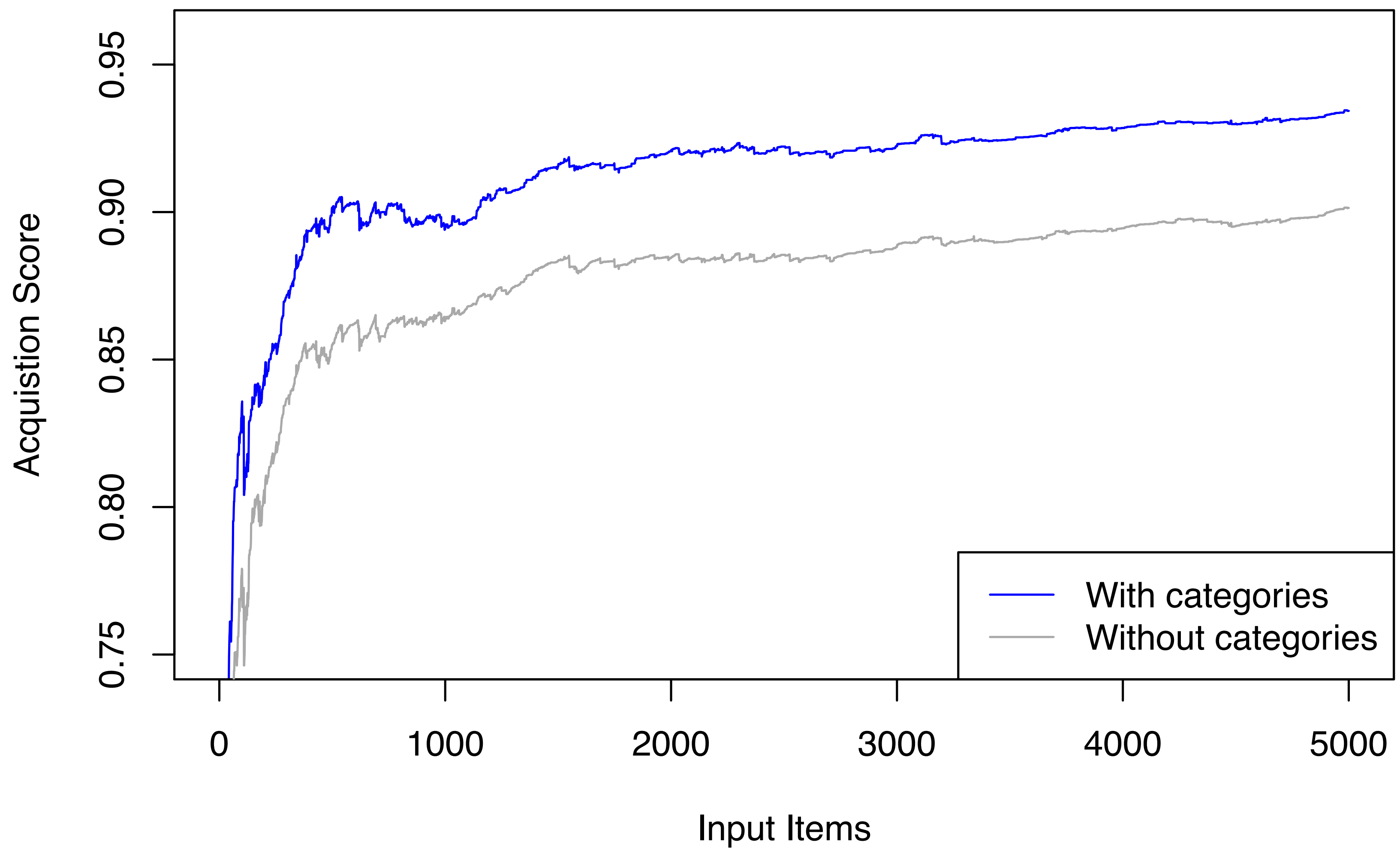
$$a_c(w|f,\, \mathrm{Uttr}^{(t)}) = \frac{p^{(t-1)}(f|\mathrm{cat}(w))}{\sum_{w_k \in \mathrm{Uttr}^{(t)}} p^{(t-1)}(f|\mathrm{cat}(w_k))}$$

$$\begin{aligned} a(w|f,\, \mathrm{Uttr}^{(t)}) &= \mathit{weight}(w) \cdot a_w(w|f,\, \mathrm{Uttr}^{(t)}) \\ &+ (1 - \mathit{weight}(w)) \cdot a_c(w|f,\, \mathrm{Uttr}^{(t)}) \end{aligned}$$

$$\mathit{weight}(w) = \frac{\mathit{freq}(w)}{\mathit{freq}(w) + 1}$$

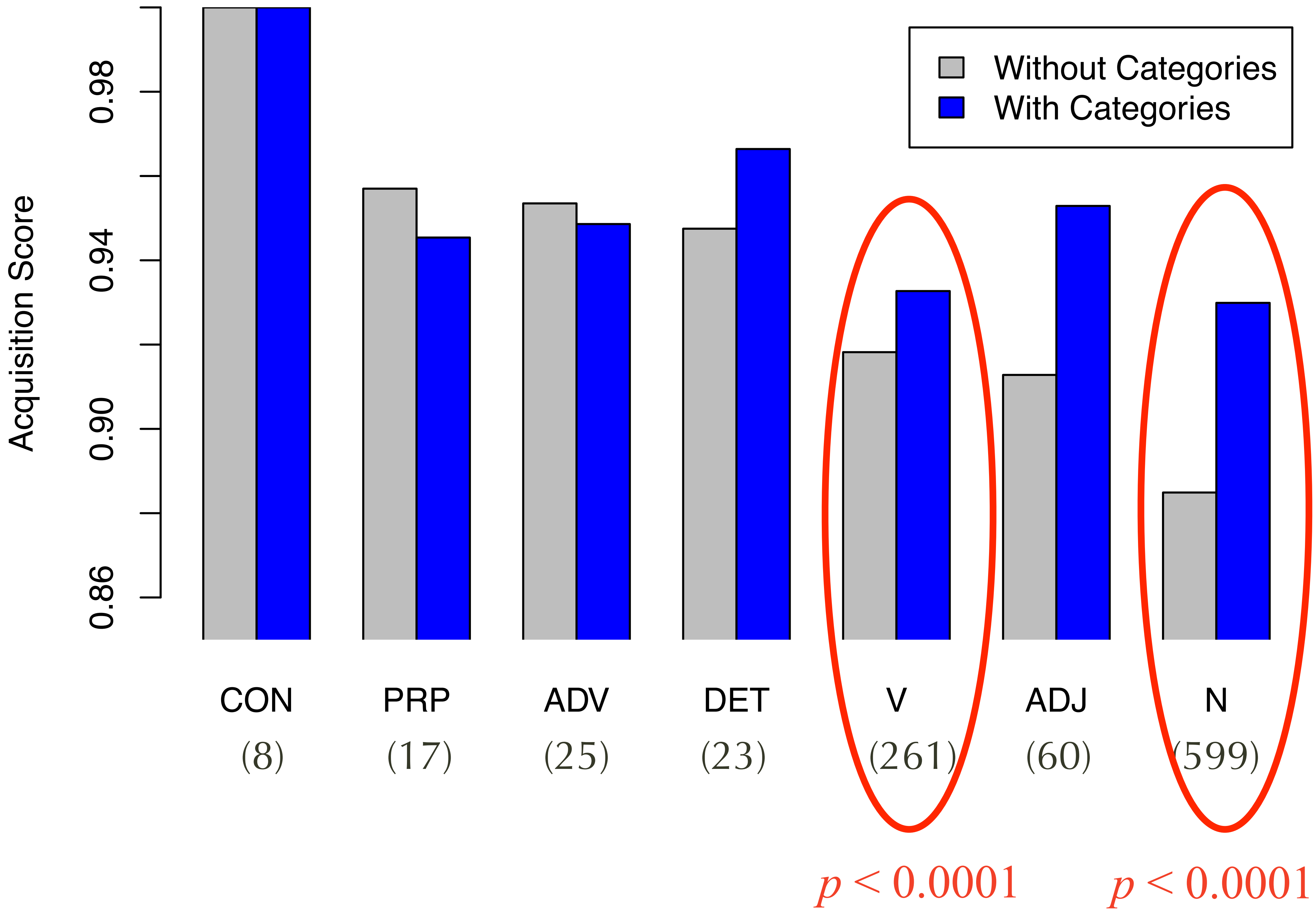
Alishahi & Fazly: Overall Learning Rates

- Learning rate over time:



➡ Integrating lexical categories in word learning improves overall performance

Alishahi & Fazly: Comparing Categories



Computational Word Learning

- Many computational models of word learning suggest that
  - several behavioural patterns can be a by-product of the statistical properties of the input that children receive
  - children's behavioural changes are not necessarily due to a shift in the underlying learning mechanism
  - a unified learning mechanism can explain a variety of effects that have been attributed to task-specific constraints or biases

Open Questions

- Most existing models do not use a realistic representation of semantic information
- Word learning studies are generally limited to mappings between nouns and concrete objects
- In particular, relational or abstract meaning representations are often ignored
- Computational studies of word learning have mostly been carried in isolation and independently of the other aspects of language acquisition