On the Origin of Shallow Syntactic Bootstrapping

Afra Alishahi & Grzegorz Chrupała

January 17, 2014

Learning Words

- How children learn the meaning of words?

doggy is playing with a frisbee

Cross-situational Learning:
Using Co-occurrence Statistics

doggy is playing with a frisbee

Cross-situational Learning: Using Co-occurrence Statistics

doggy is playing with a frisbee

She is throwing the frisbee

Syntactic Bootstrapping: Using Sentential Context

doggy is playing with a frisbee

Syntactic Bootstrapping:
Using Sentential Context

doggy is playing with a frisbee

He is playing with a matchbox

Sara is cutting with a knife

Ian is washing with a soapbar

Syntactic Bootstrapping:
Using Sentential Context

doggy is playing with a frisbee

He is playing with a matchbox

Sara is cutting with a knife

Ian is washing with a soapbar

X is DOing with a Y

physical object

Main Questions

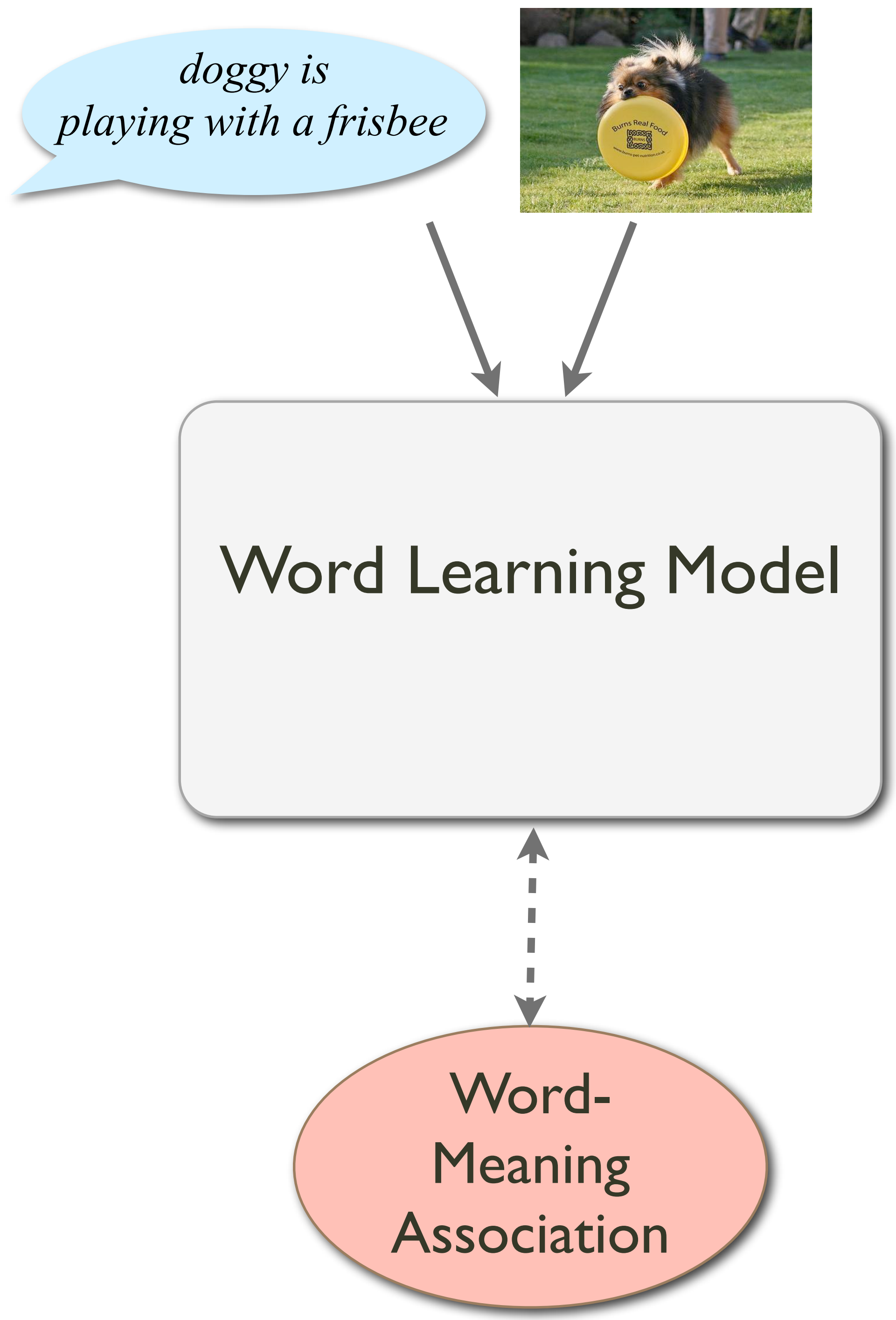
- How can these two mechanisms be integrated into a unified model of word learning?

Main Questions

- How can these two mechanisms be integrated into a unified model of word learning?

- What is the origin and onset of syntactic bootstrapping?

Modeling of Cross-situational Word Learning



Cross-situational Learning [Fazly et al. 2010]

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

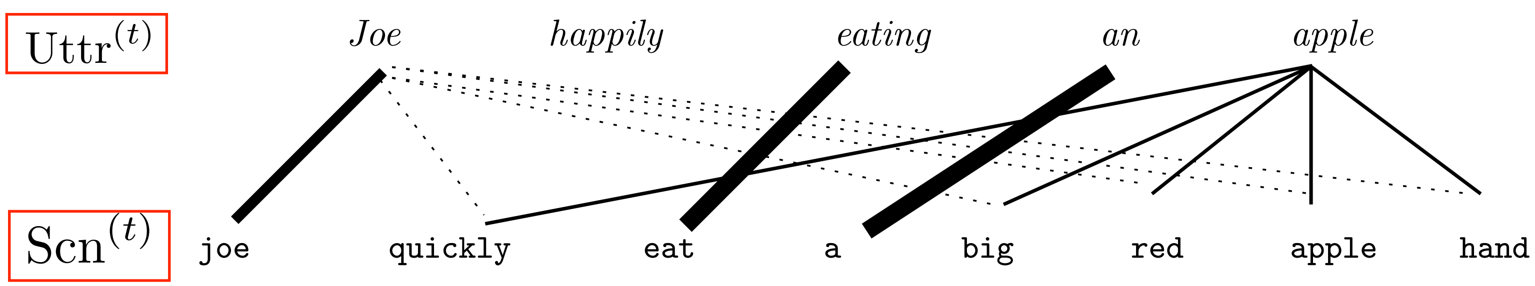
1. Alignment: use previously learned meaning associations to align each word in utterance with each meaning element from the scene

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

Cross-situational Learning [Fazly et al. 2010]

- 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

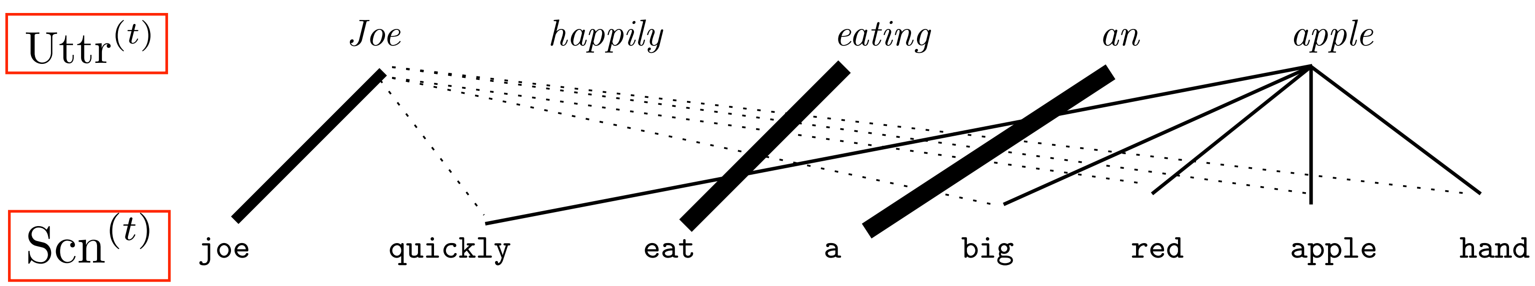


2. Update: use these alignments to update the probabilistic associations between a word and its meaning elements

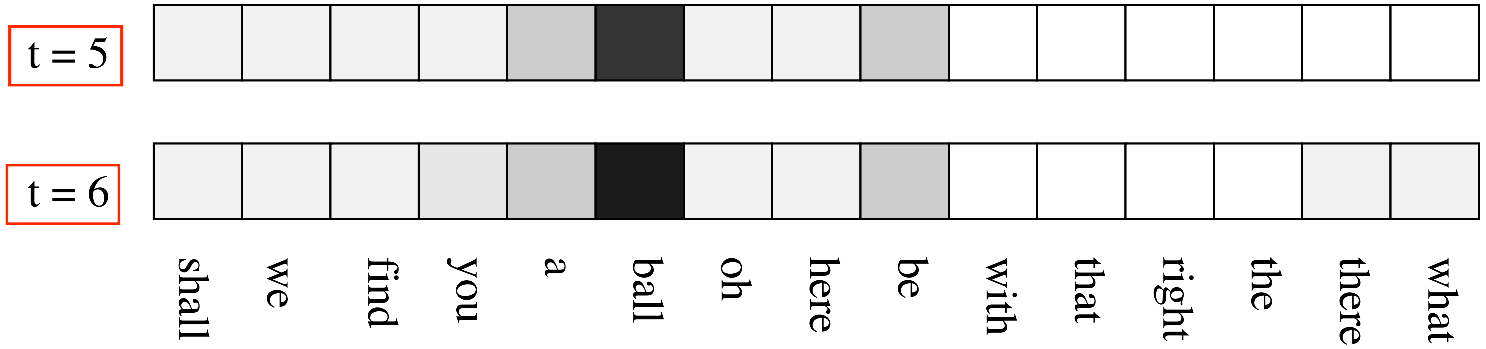
Cross-situational Learning [Fazly et al. 2010]

- 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



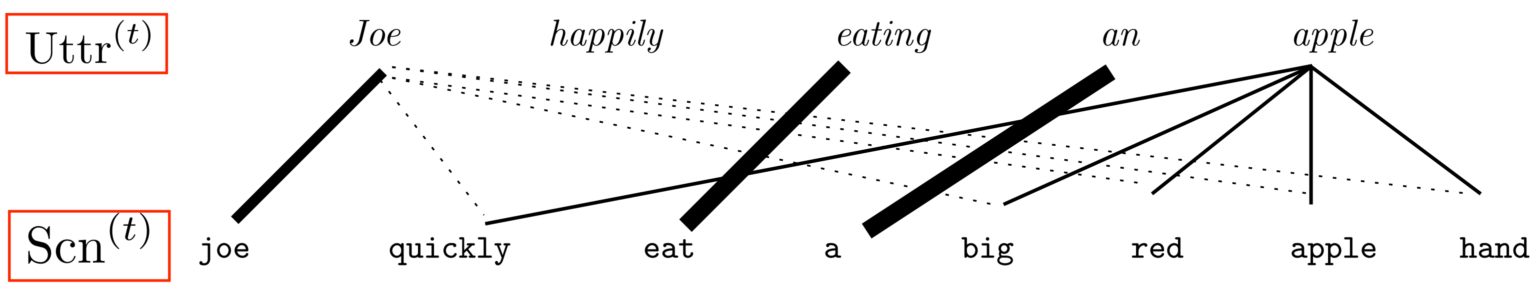
2. Update: use these alignments to update the probabilistic associations between a word and its meaning elements



Cross-situational Learning [Fazly et al. 2010]

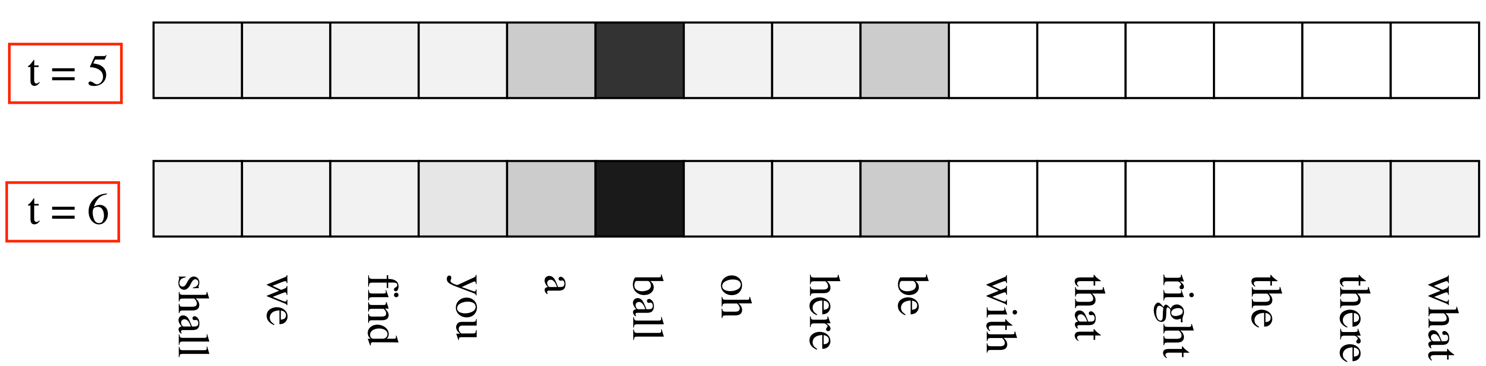
- For every new pair of scene and utterance, $(\mathrm{Uttr}^{(t)}, \mathrm{Scn}^{(t)})$

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



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

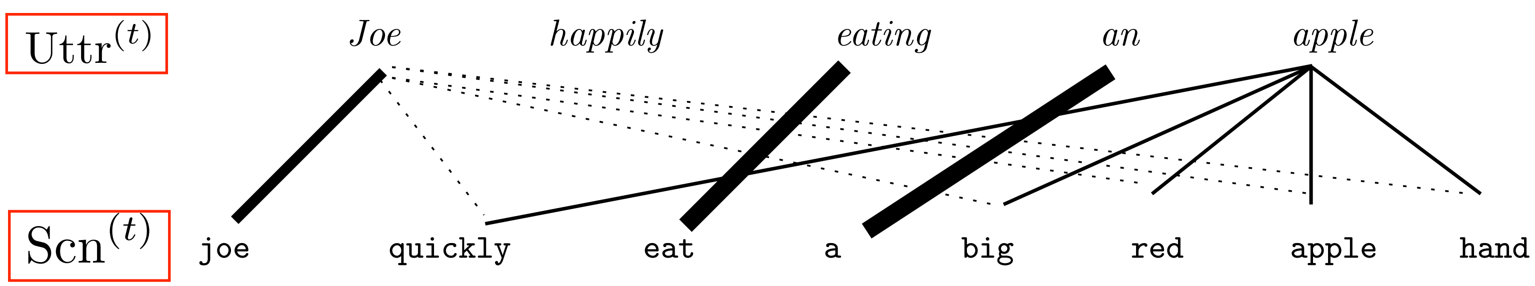
2. Update: use these alignments to update the probabilistic associations between a word and its meaning elements



Cross-situational Learning [Fazly et al. 2010]

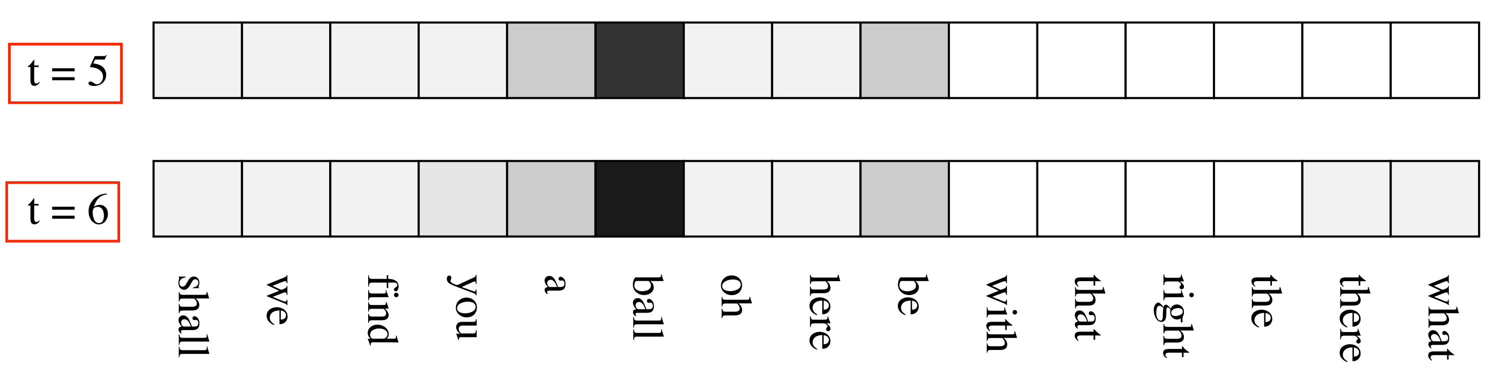
- For every new pair of scene and utterance, $(\mathrm{Uttr}^{(t)}, \mathrm{Scn}^{(t)})$

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



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

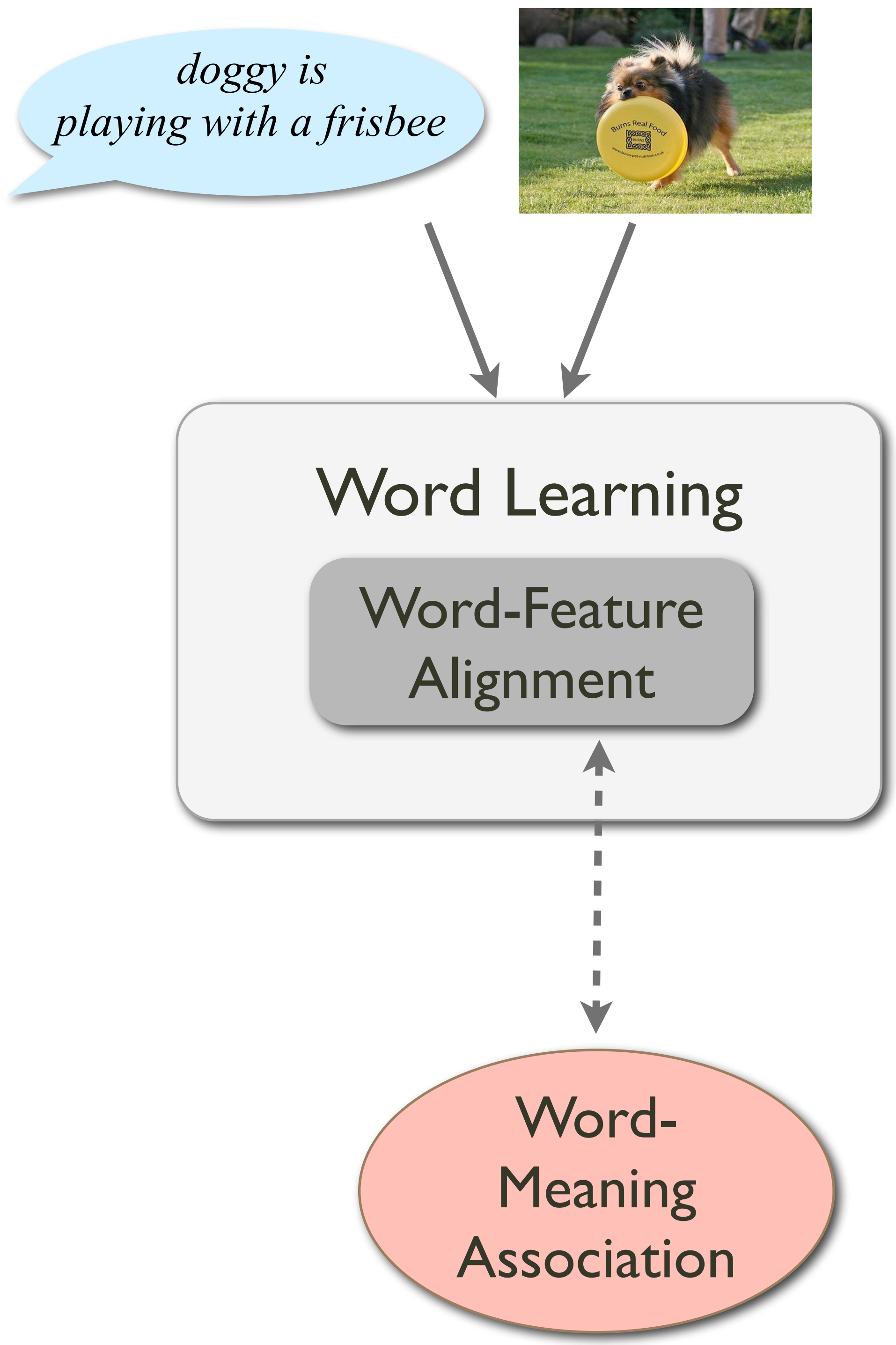
2. Update: use these alignments to update the probabilistic associations between a word and its meaning elements



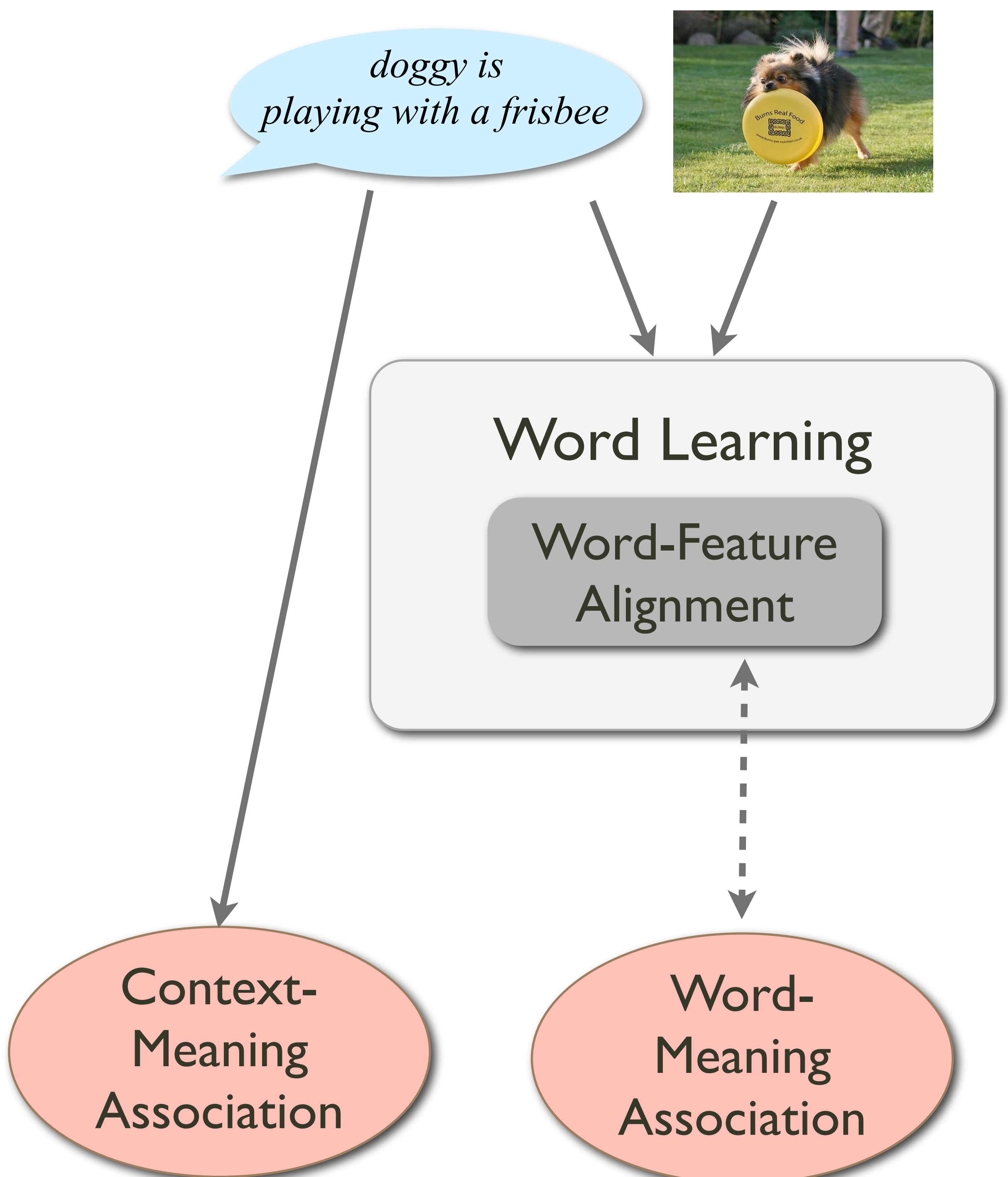
$$\mathrm{assoc}^{(t)}(w,f) = \mathrm{assoc}^{(t-1)}(w,f) + a(w|f, \mathrm{Uttr}^{(t)})$$

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

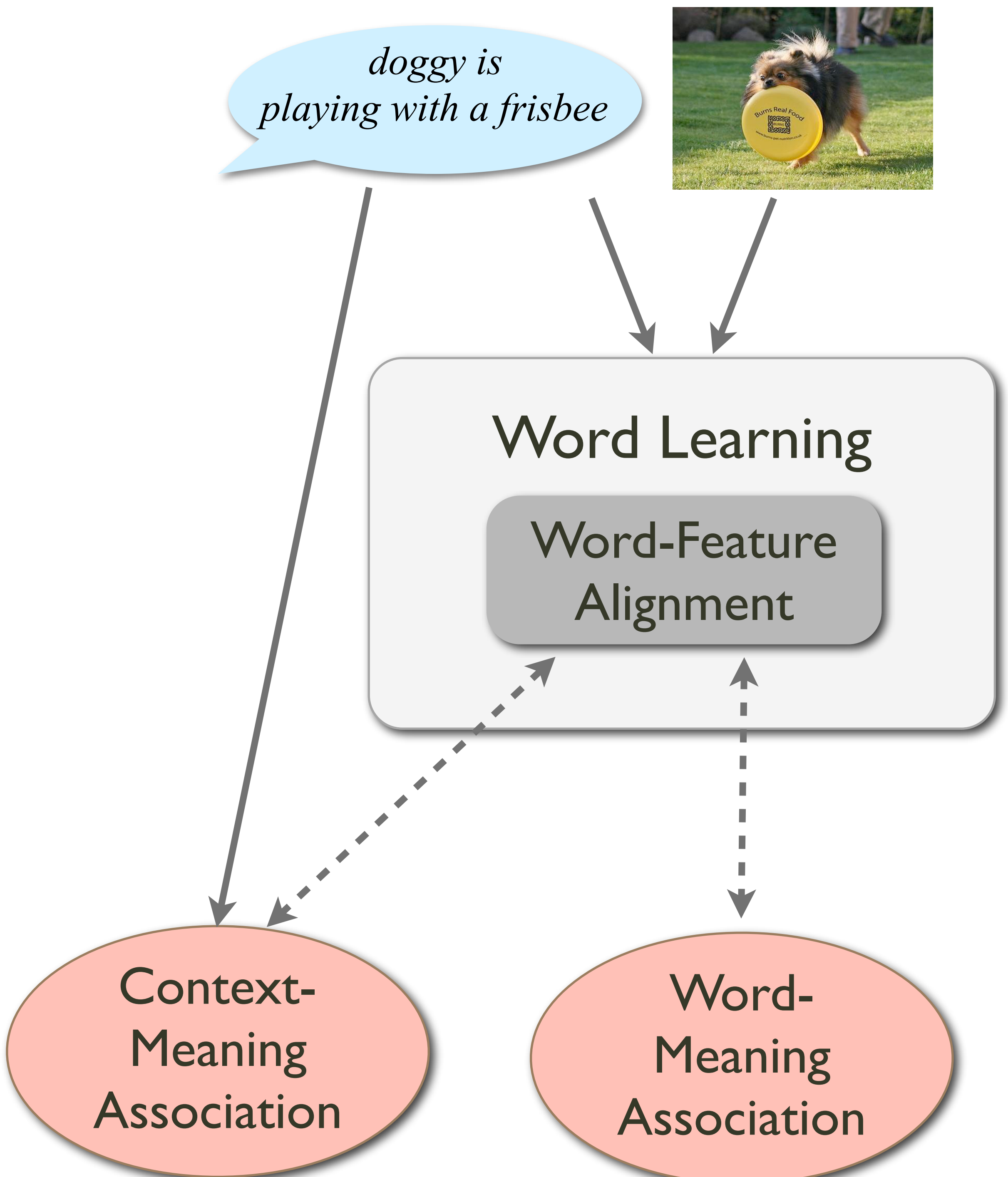
Adding Evidence from Sentential Context



Adding Evidence from Sentential Context



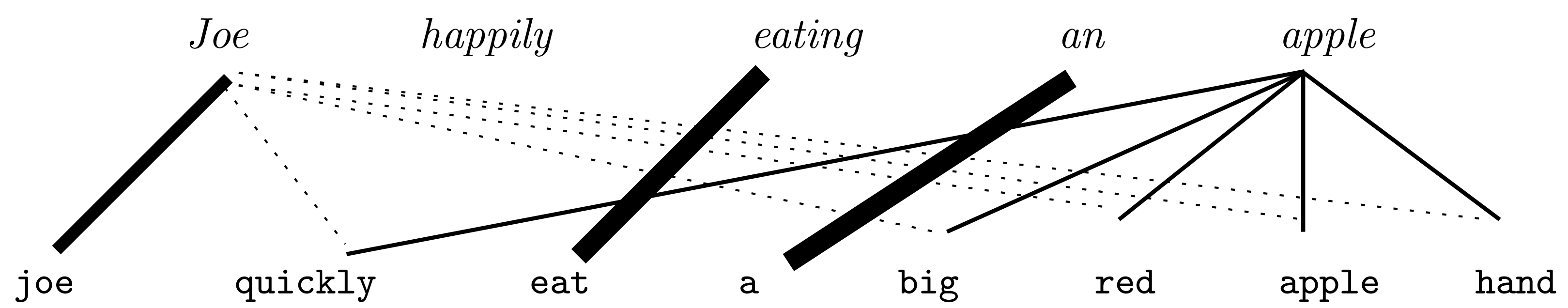
Adding Evidence from Sentential Context



Lexical Categories as a Source for “Shallow” Syntactic Bootstrapping

Lexical Categories as a Source for "Shallow" Syntactic Bootstrapping

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



Lexical Categories as a Source for "Shallow" Syntactic Bootstrapping

- 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)

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

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

Lexical Categories as a Source for "Shallow" Syntactic Bootstrapping

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

Cross-situational Word Meaning

P(f|w)

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

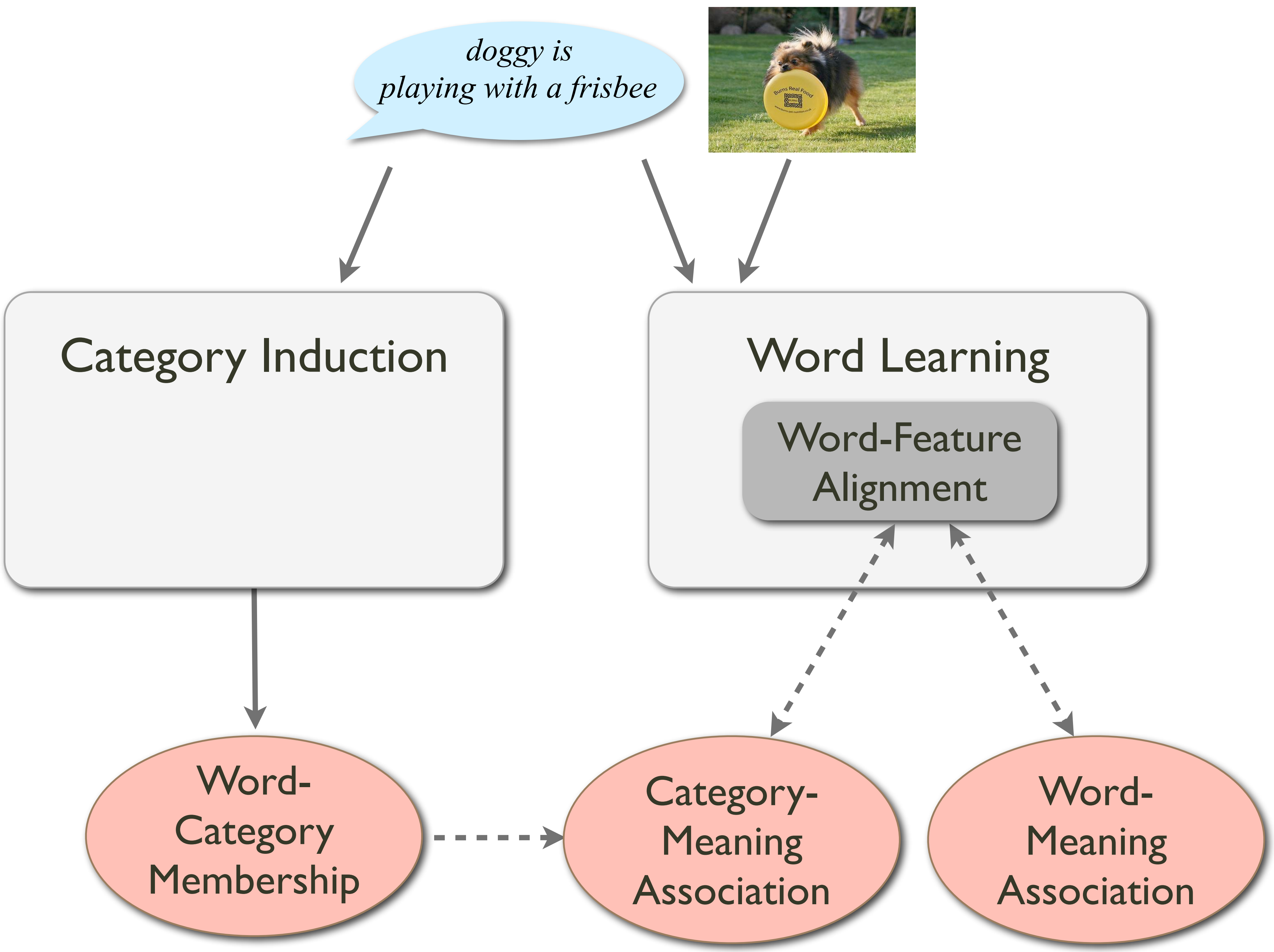
Meaning Associated w. Lexical Category

P(f|C)

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

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

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


doggy is
playing with a frisbee
Category Induction
Word Learning
Word-Feature
Alignment
Word-
Category
Membership
Category-
Meaning
Association
Word-
Meaning
Association

Automatically Induced Categories

- Latent Dirichlet Allocation-based model
  - A hierarchical Bayesian model for inducing a topic structure from a collection of documents

$$
\begin{aligned}
\phi_k &\sim \mathrm{Dirichlet}(\beta), & k &\in [1, K] \\
\theta_d &\sim \mathrm{Dirichlet}(\alpha), & d &\in [1, D] \\
z_{n_d} &\sim \mathrm{Categorical}(\theta_d), & n_d &\in [1, N_d] \\
w_{n_d} &\sim \mathrm{Categorical}(\phi_{z_{n_d}}), & n_d &\in [1, N_d]
\end{aligned}
$$

Automatically Induced Categories

- Latent Dirichlet Allocation-based model
  - A hierarchical Bayesian model for inducing a topic structure from a collection of documents

$$
\begin{aligned}
\phi_k &\sim \text{Dirichlet}(\beta), & k &\in [1, K] \xrightarrow{} \textit{topics} \\
\theta_d &\sim \text{Dirichlet}(\alpha), & d &\in [1, D] \xrightarrow{} \textit{documents} \\
z_{n_d} &\sim \text{Categorical}(\theta_d), & n_d &\in [1, N_d] \xrightarrow{} \textit{words} \\
w_{n_d} &\sim \text{Categorical}(\phi_{z_{n_d}}), & n_d &\in [1, N_d]
\end{aligned}
$$

Automatically Induced Categories

- Chrupala (2011) reinterpretation of LDA:
  - Word types correspond to documents
  - Context words correspond to words in documents

$$
\begin{aligned}
\phi_k &\sim \text{Dirichlet}(\beta), & k &\in [1, K] \longrightarrow \textit{word classes} \\
\theta_d &\sim \text{Dirichlet}(\alpha), & d &\in [1, D] \longrightarrow \textit{word types} \\
z_{n_d} &\sim \text{Categorical}(\theta_d), & n_d &\in [1, N_d] \longrightarrow \textit{context features} \\
w_{n_d} &\sim \text{Categorical}(\phi_{z_{n_d}}), & n_d &\in [1, N_d]
\end{aligned}
$$

Automatically Induced Categories

- Chrupala (2011) reinterpretation of LDA:
  - Word types correspond to documents
  - Context words correspond to words in documents

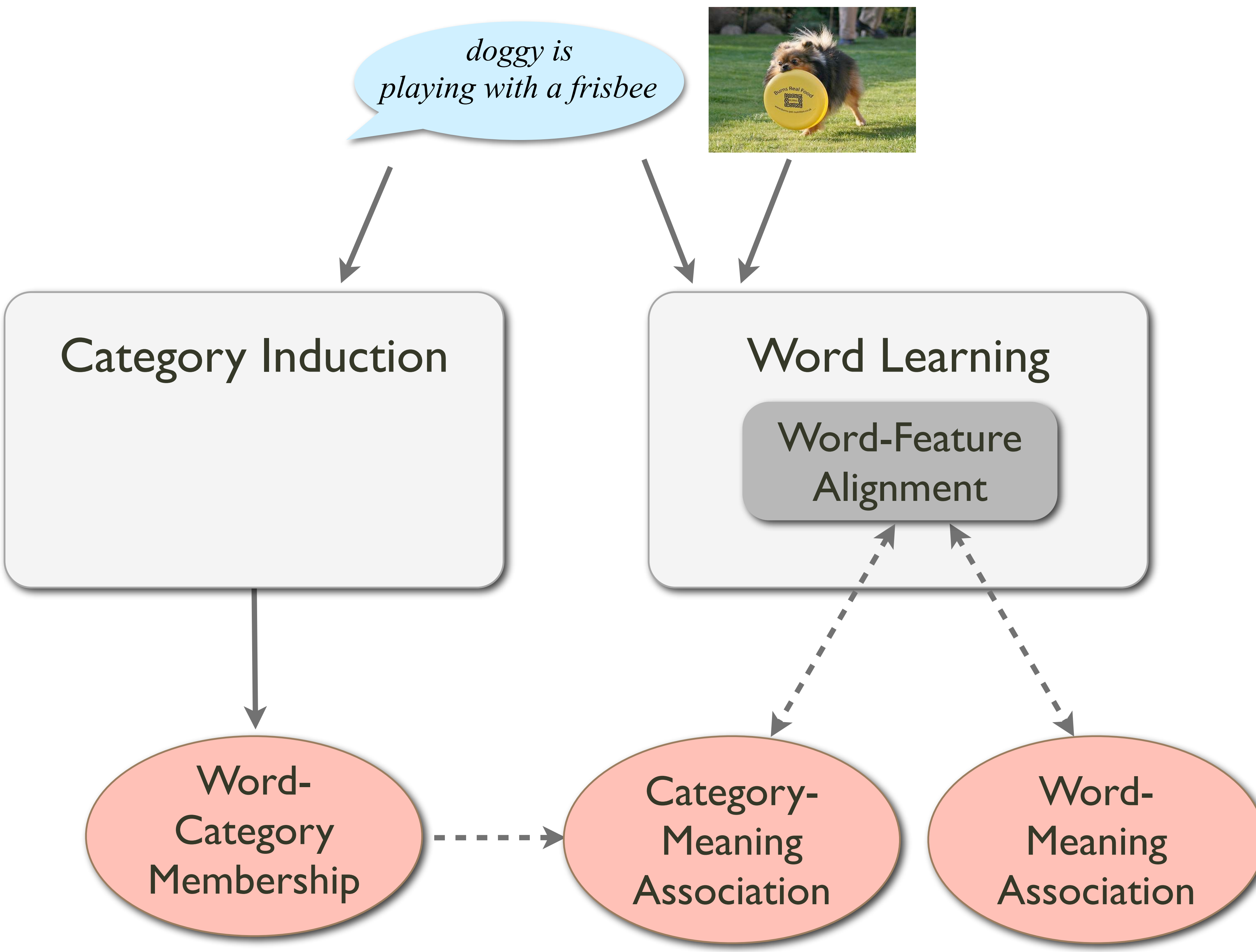
$$
\begin{aligned}
\phi_k &\sim \text{Dirichlet}(\beta), & k &\in [1, K] \\
\theta_d &\sim \text{Dirichlet}(\alpha), & d &\in [1, D] \\
z_{n_d} &\sim \text{Categorical}(\theta_d), & n_d &\in [1, N_d] \\
w_{n_d} &\sim \text{Categorical}(\phi_{z_{n_d}}), & n_d &\in [1, N_d]
\end{aligned}
$$

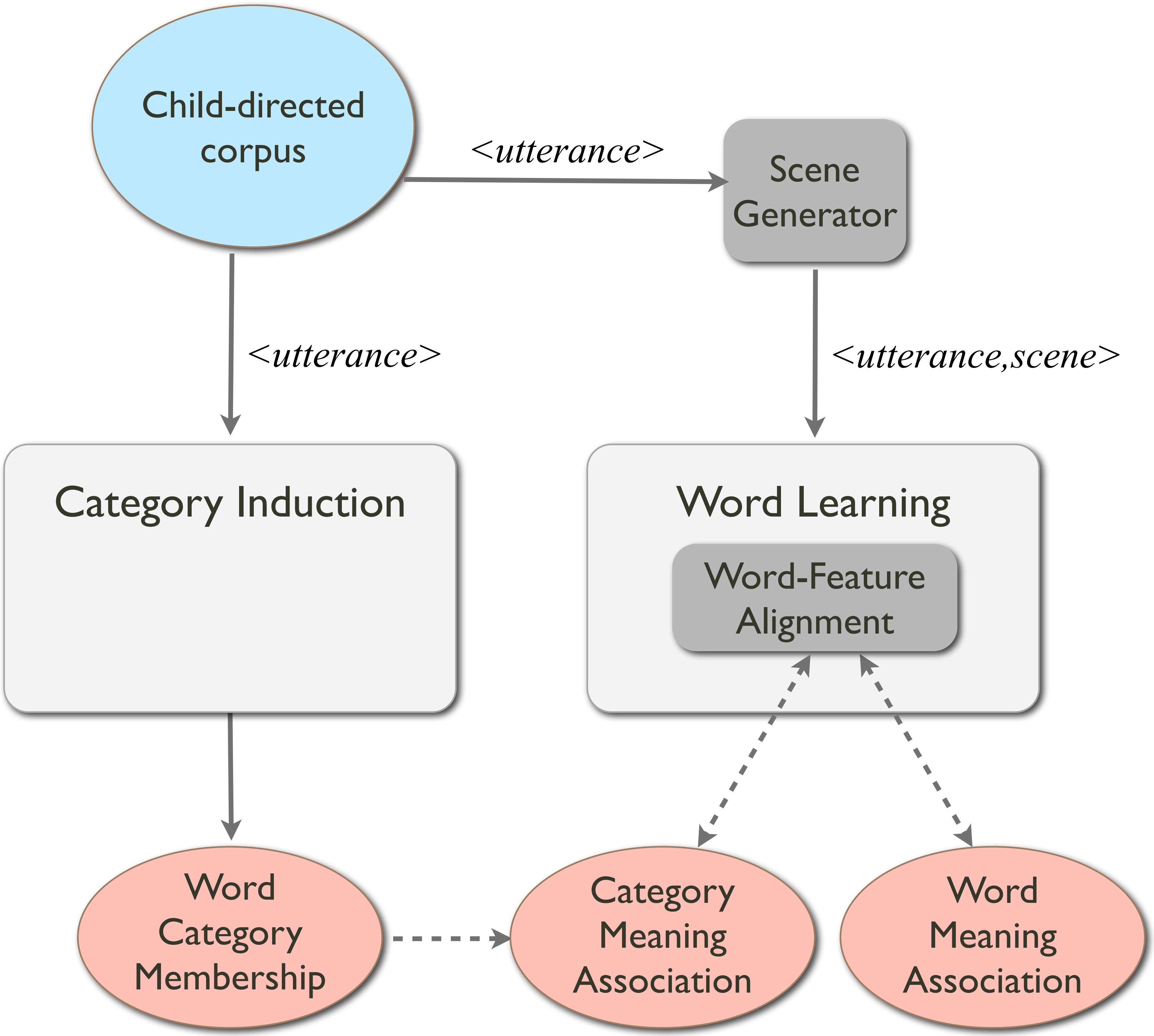
word classes

word types

context features

class membership of word d given context Nd


doggy is
playing with a frisbee
Category Induction
Word Learning
Word-Feature
Alignment
Word-
Category
Membership
Category-
Meaning
Association
Word-
Meaning
Association



Child-directed corpus
<utterance>
Scene Generator
<utterance>
<utterance,scene>
Category Induction
Word Learning
Word-Feature Alignment
Word Category Membership
Category Meaning Association
Word Meaning Association

Input Data

- A sample input item:

Utterance:	{	*mommy, ate, broccoli* }
Scene:	{	ANIMATE, HUMAN, …,
		CONSUMPTION, ACTION, …
		BROCCOLI, VEGETABLE, …
		PLATE, OBJECT, … }

Input Data

- A sample input item:

Utterance:	{	*mommy, ate, broccoli* }
Scene:	{	ANIMATE, HUMAN, …,
		CONSUMPTION, ACTION, …
		BROCCOLI, VEGETABLE, …
		PLATE, OBJECT, … }

- Child-adult interaction data from CHILDES [MacWhinney'95]
  - Manchester corpus [Theakston et al.'01]
  - Pearl-Sprouse corpus [Pearl & Sprouse'13]

Input Data

- Child-directed utterances from each corpus

that is an apple

do you like apple?

do you want to give dolly an apple?

can teddy bear give penguin a kiss?

...

Input Data

- … paired with meaning primitives extracted from WordNet

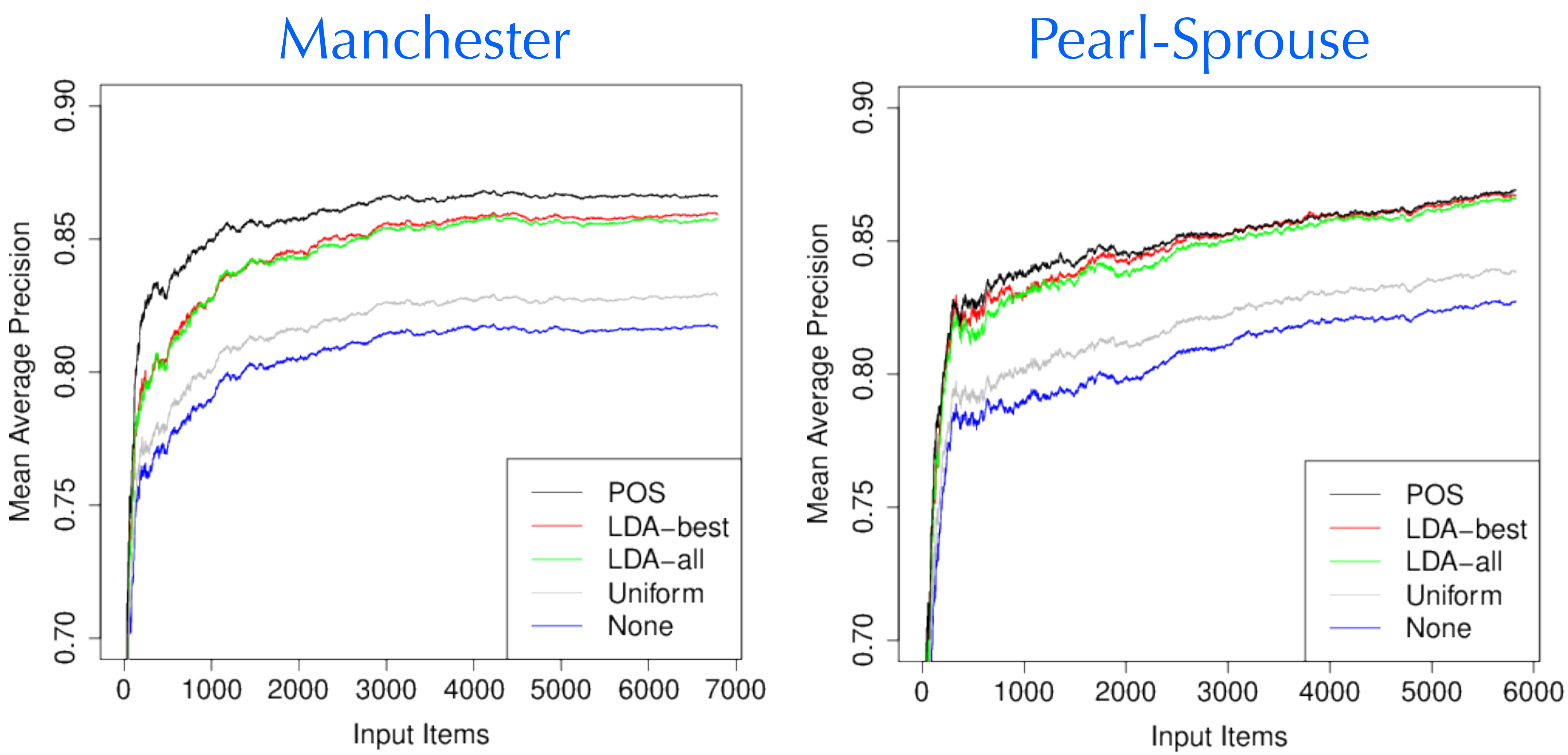
that is an apple	definite, be, edible, fruit, …
do you like apple?	do, person, you, desire, edible, fruit, …
do you want to give dolly an apple?	do, person, you, want, location, artifact, …
can teddy bear give penguin a kiss?	artifact, object, teddy, animal, bear, …
…	…

Input Data

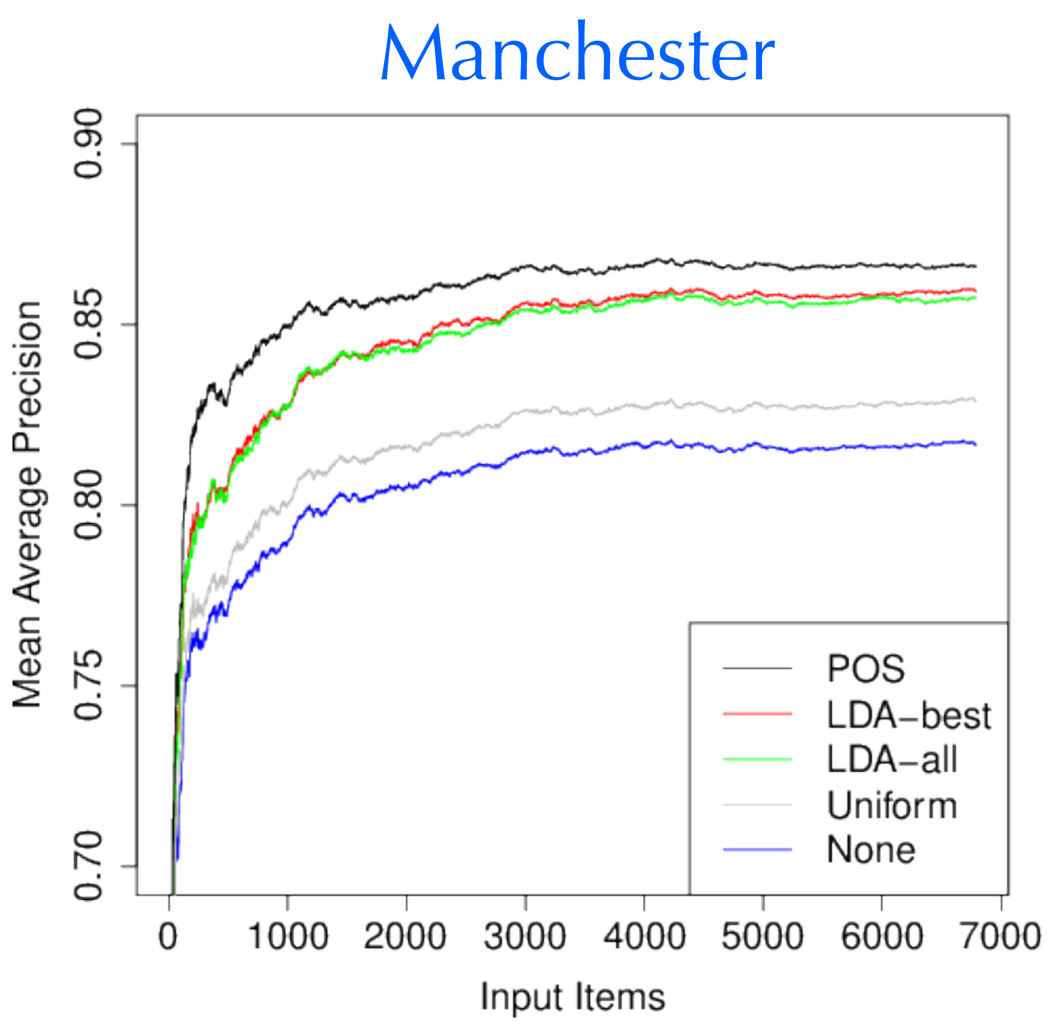
- ... and subsequent primitive sets combined to simulate referential uncertainty:

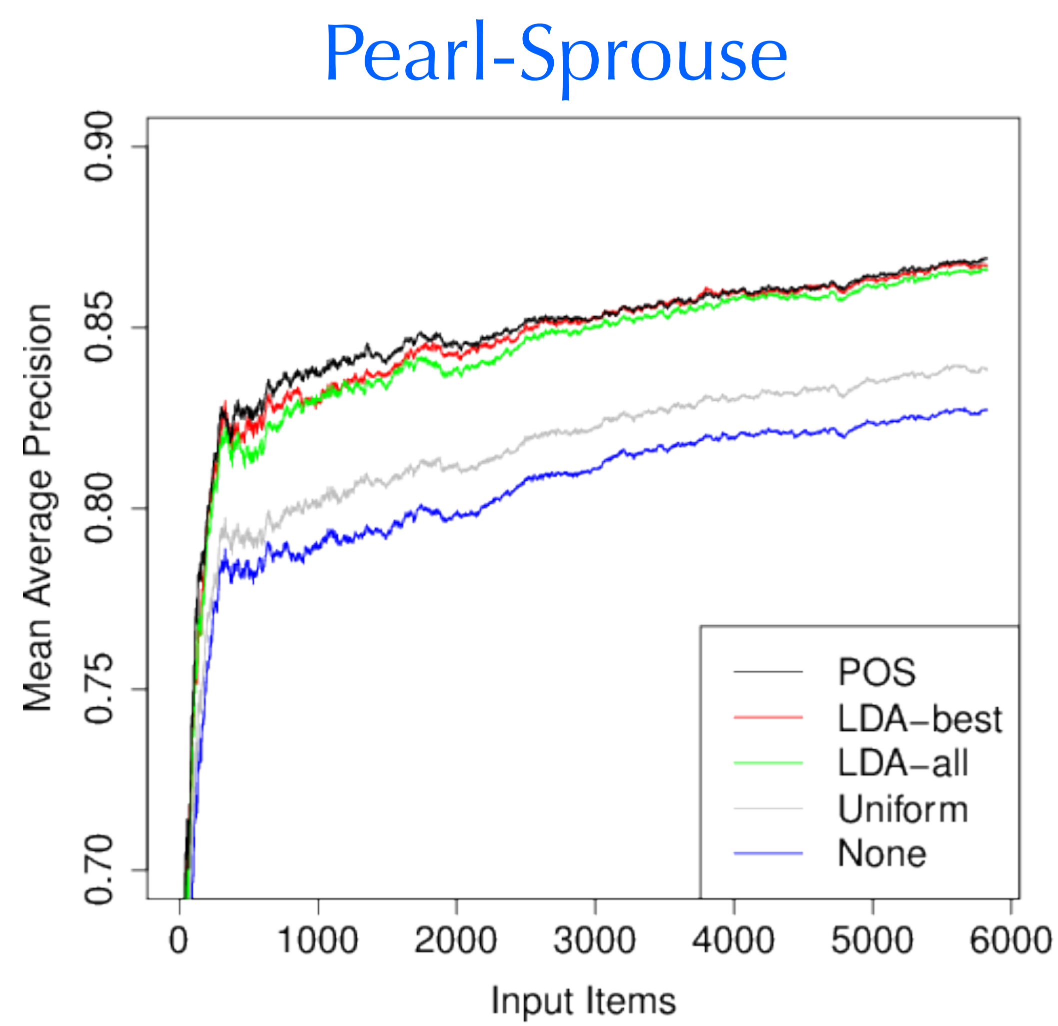
that is an apple	definite, be, edible, fruit, ...
do you like apple?	do, person, you, desire, edible, fruit, ...
do you want to give dolly an apple?	do, person, you, want, location, artifact, ...
can teddy bear give penguin a kiss?	artifact, object, teddy, animal, bear, ...
...	...

Automatically Induced Categories



Automatically Induced Categories

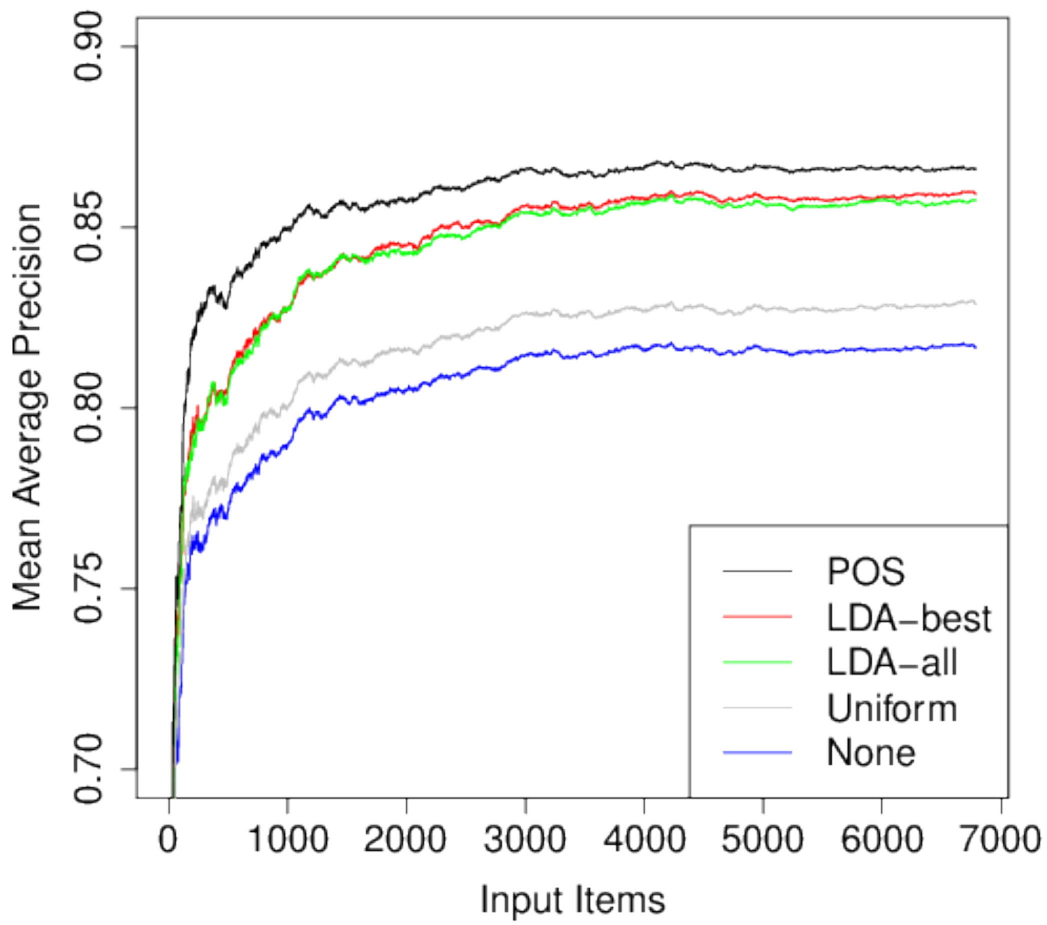




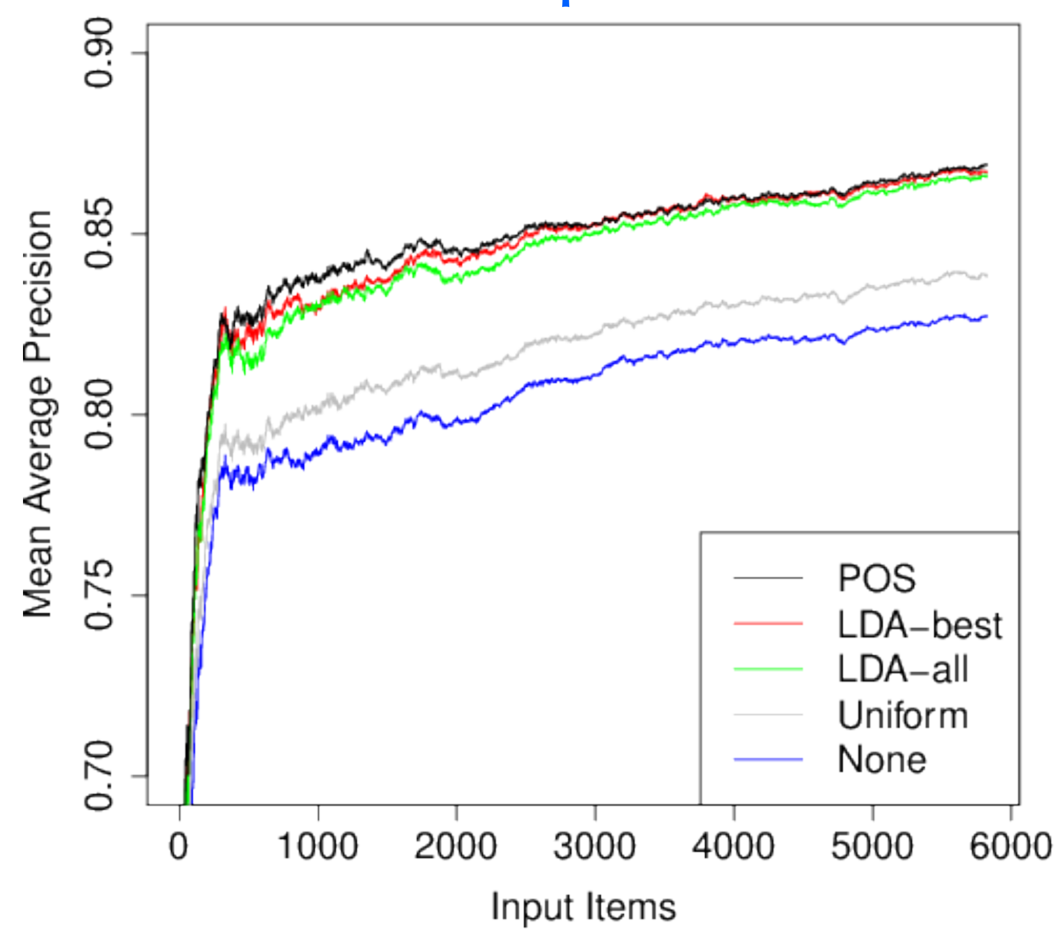
➡ Categories significantly improve word learning performance

Automatically Induced Categories

Manchester



Pearl-Sprouse



➡ Categories significantly improve word learning performance

➡ LDA-based categories are comparable to manually-annotated, "gold" POS categories

Guessing Meaning from Context

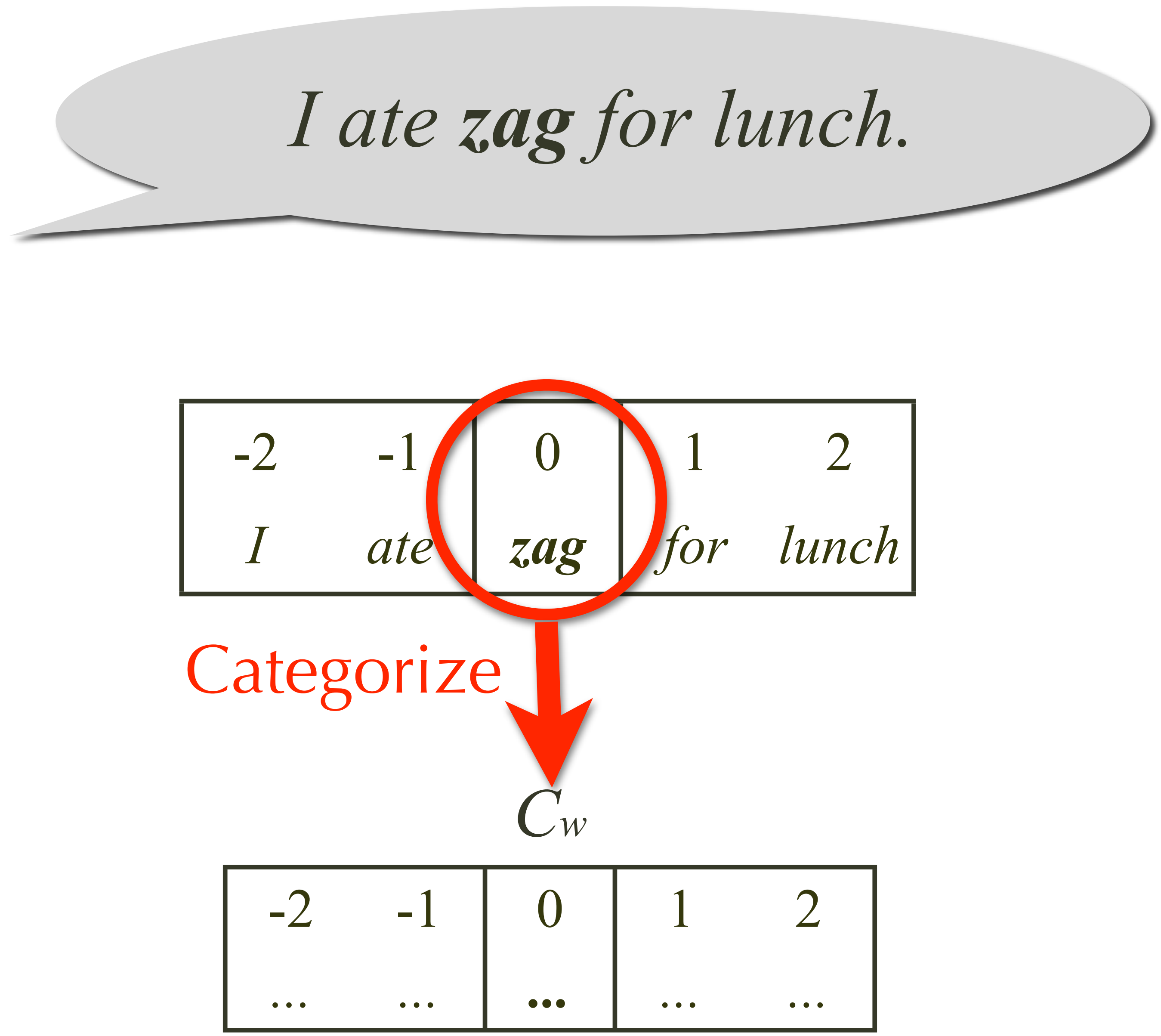
*I ate **zag** for lunch.*

Guessing Meaning from Context

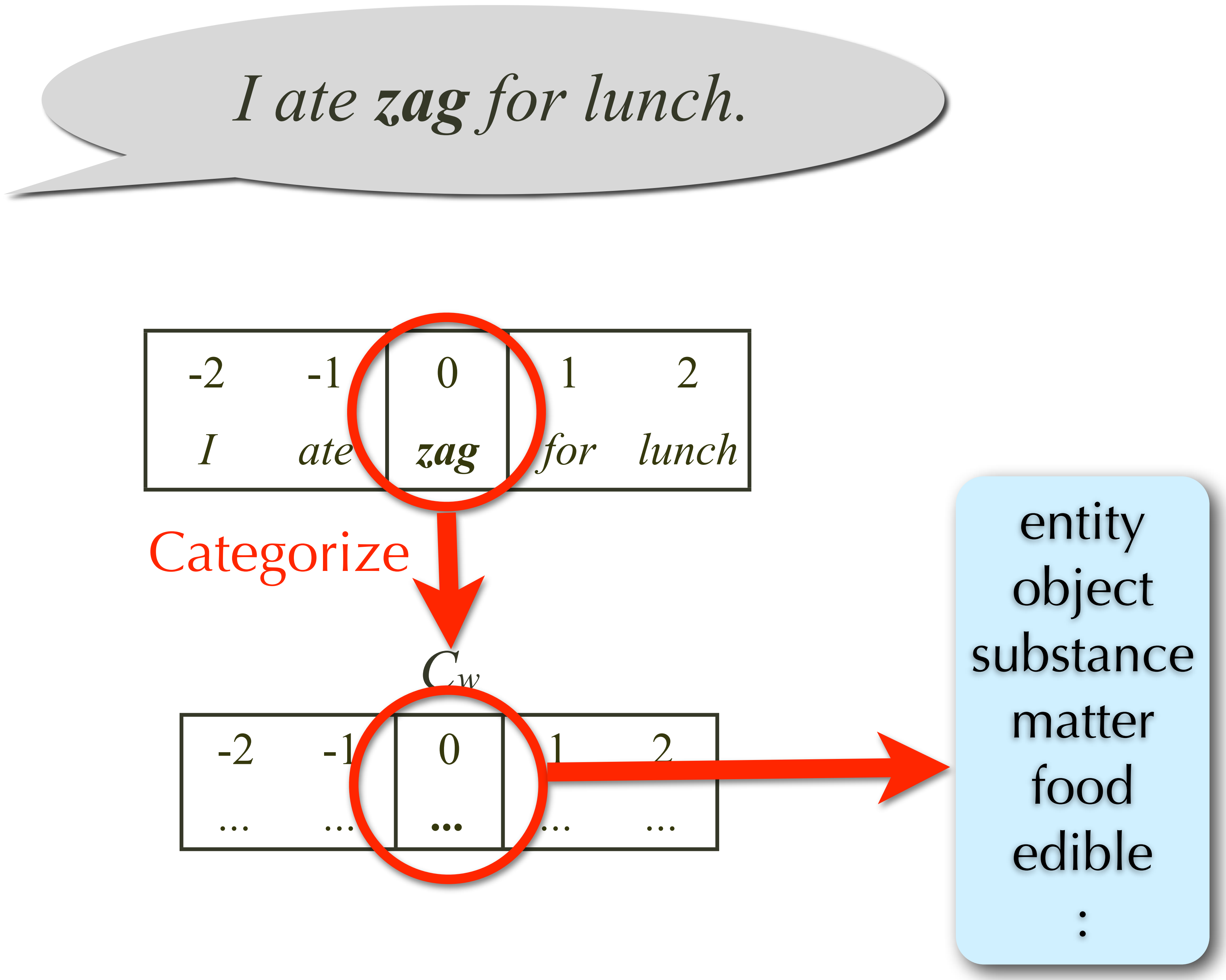
*I ate **zag** for lunch.*

-2	-1	0	1	2
I	*ate*	***zag***	*for*	*lunch*

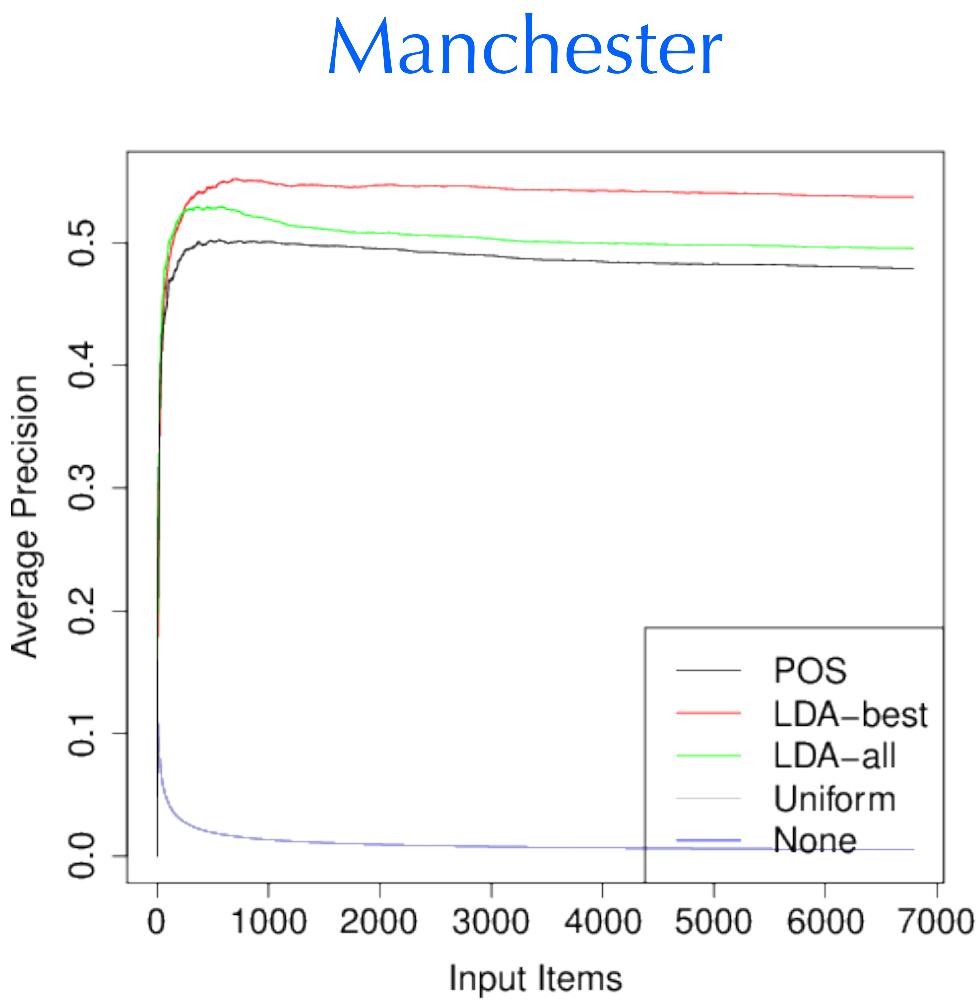
Guessing Meaning from Context

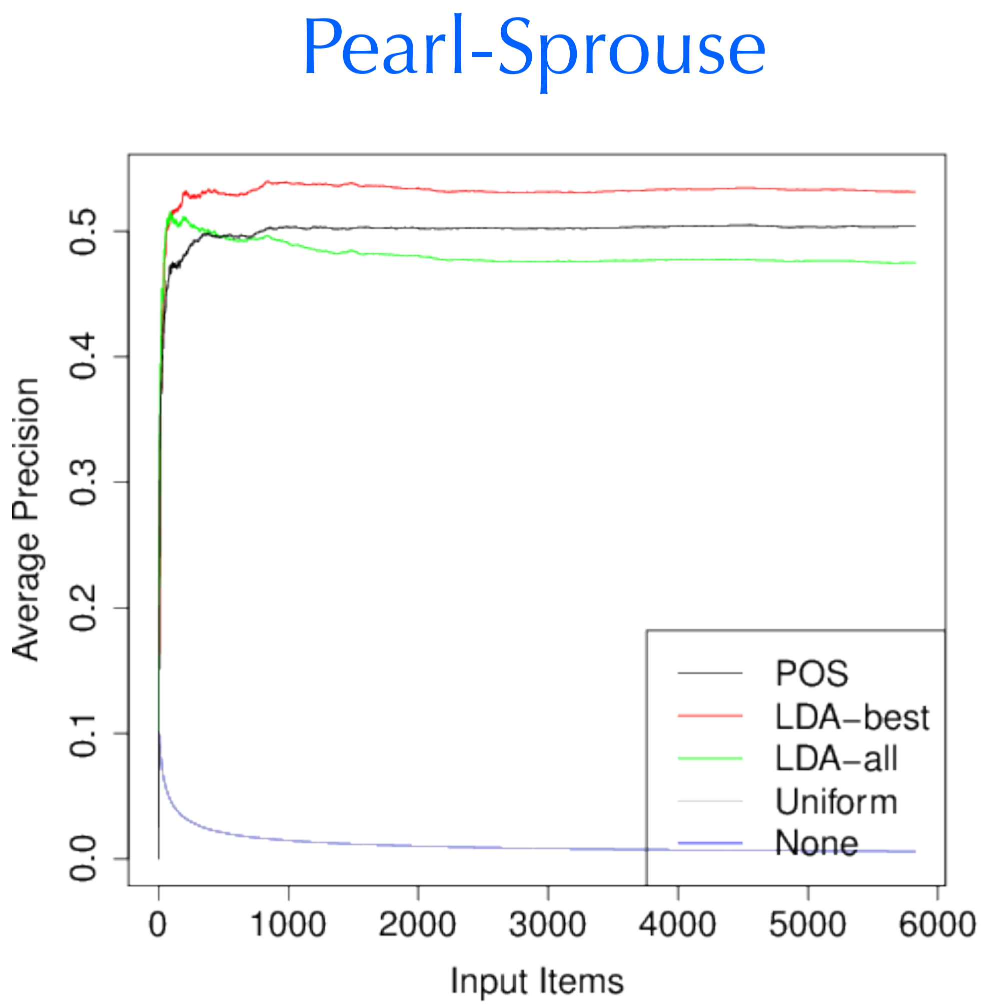


Guessing Meaning from Context



Accuracy of Guessed Meaning



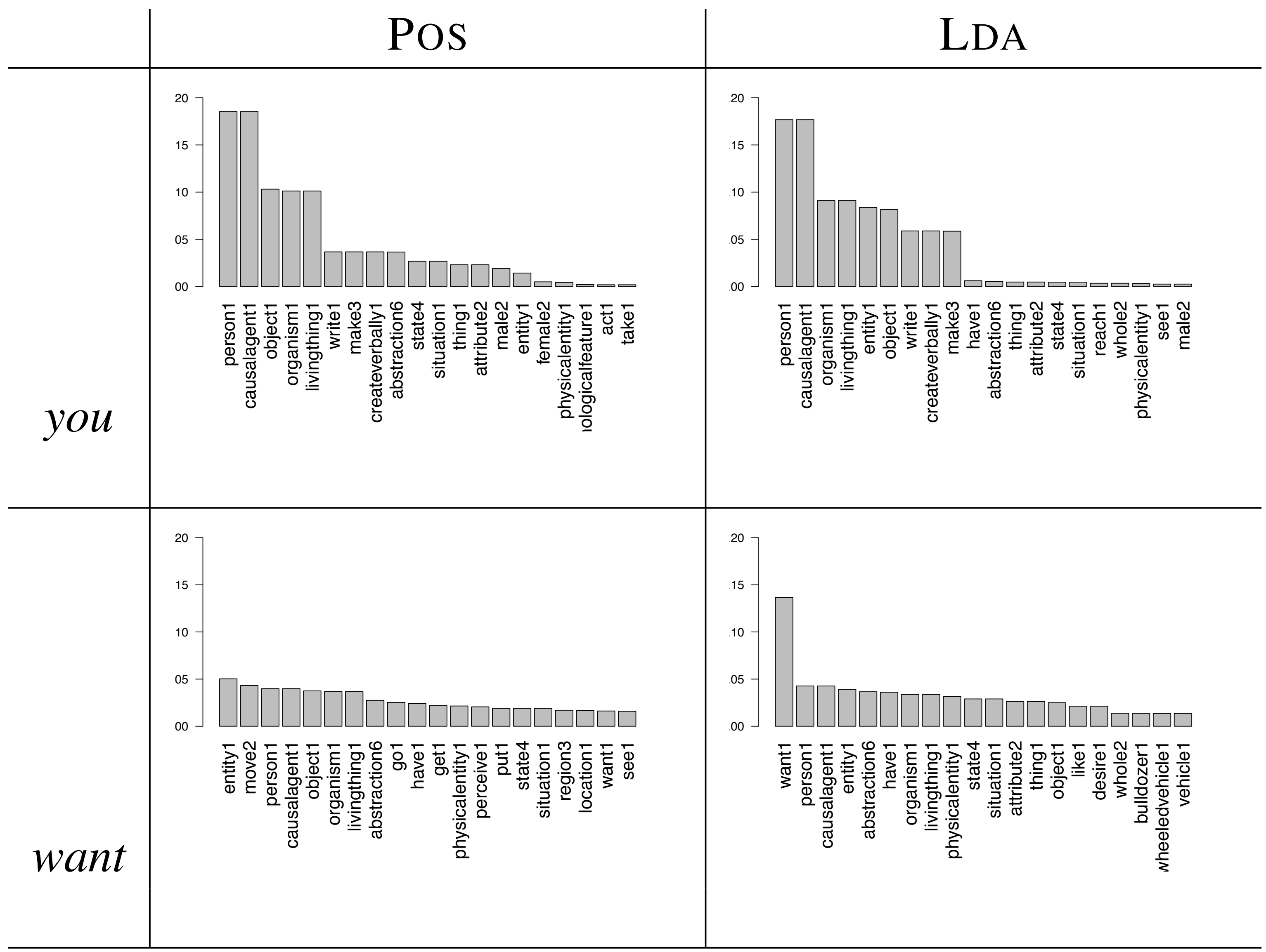


An Example

do you want to read a book?

An Example

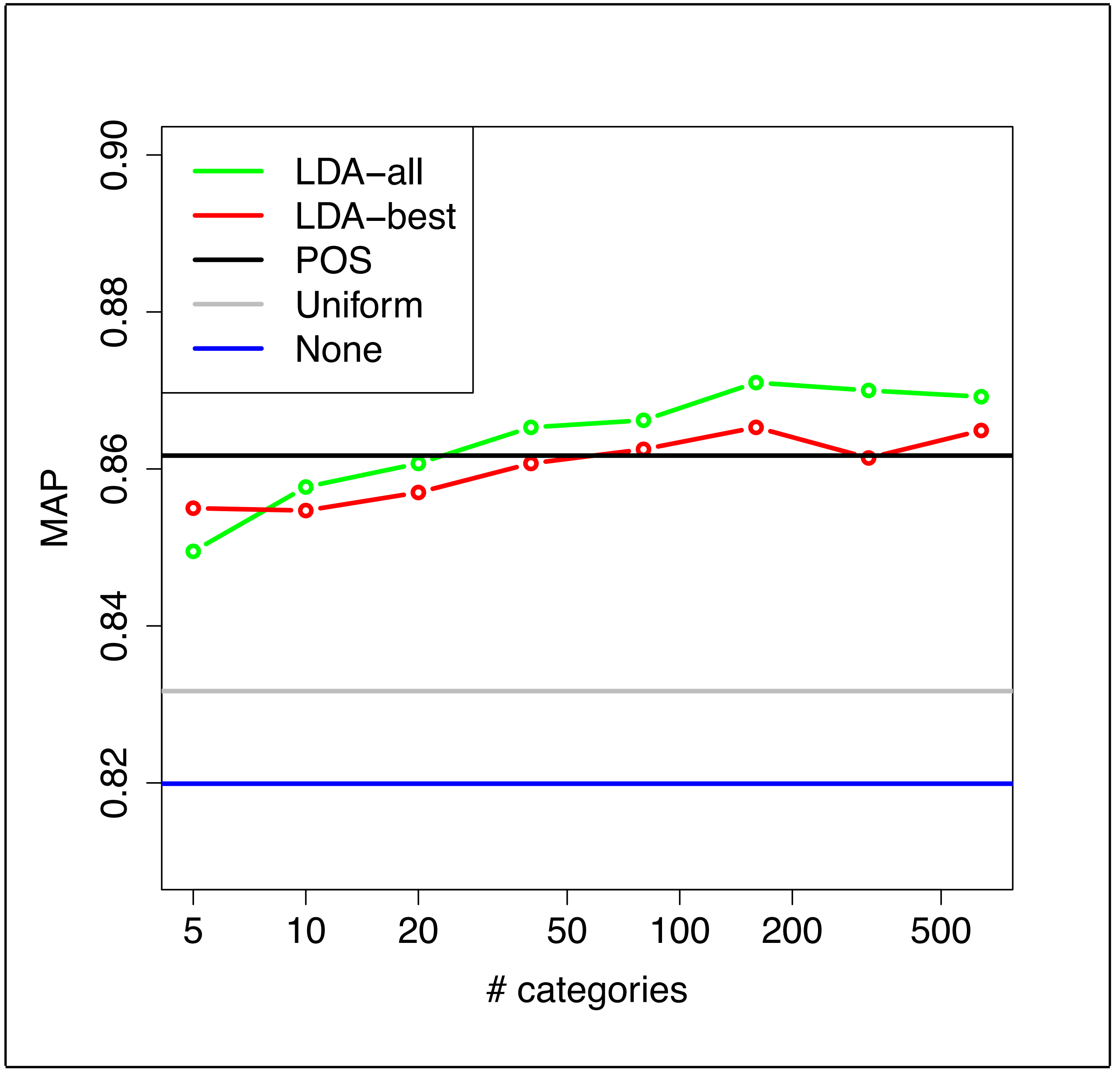
do <u>you</u> <u>want</u> to read a book?



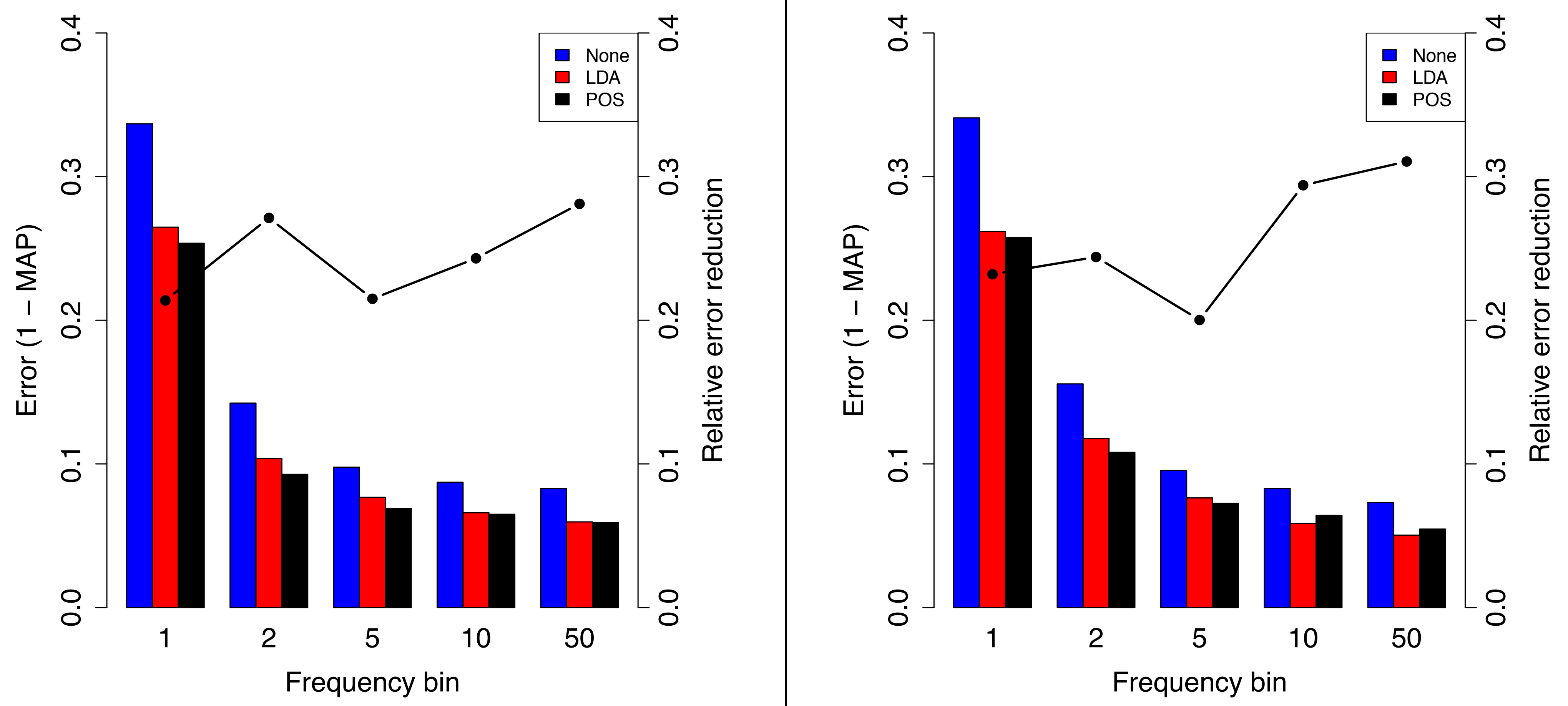
Summary

- Syntactic information can be seamlessly integrated into cross-situational learning
- Appropriate categories can improve the overall word learning performance
- Automatically induced, data-driven categories are as effective as the manually-annotated POS categories

Class Granularity



Impact of Word Frequency



(a) Manchester **(b)** Pearl-Sprouse

Automatically Induced Categories

- An incremental version of the collapsed Gibbs sampler:

$$\begin{boxed}{\begin{array}{l}\textbf{for } t = 1 \rightarrow \infty \ \textbf{ do} \\ \quad \textbf{for } i = 1 \rightarrow I_t \textbf{ do} \\ \quad\quad \text{sample } z_{t_i} \sim P(z_{t_i} | \mathbf{z}_{t_i - 1}, \mathbf{w}_{t_i}, \mathbf{d}_{t_i}) \\ \quad\quad \text{increment } n_t^{z_{t_i}, w_{t_i}} \text{ and } n_t^{z_{t_i}, d_{t_i}}\end{array}}\end{boxed}$$

- Only conditioned on previous word tokens:

$$P(z_t | \mathbf{z}_{t-1}, \mathbf{w}_t, \mathbf{d}_t) \propto \frac{(n_{t-1}^{z_t, d_t} + \alpha) \times (n_{t-1}^{z_t, w_t} + \beta)}{\sum_{j=1}^{V_{t-1}} n_{t-1}^{z_t, w_j} + \beta}.$$