

# Concurrent Acquisition of Word Meaning and Lexical Categories

Afra Alishahi Grzegorz Chrupała



## Introduction

- ▶ Learning the meaning of words from ambiguous and noisy context is a challenging task for language learners.
- ▶ Children draw on cross-situational statistics as well as syntactic cues to constrain potential referents of words in a complex scene.
- ▶ We model this process by integrating an LDA-based word-class learning module with a probabilistic word learning model.
- ▶ Incrementally induced word classes significantly improve word learning, comparable to manually assigned PoS categories.

## Cross-situational learning

- ▶ Each sentence paired with a simulated scene representation, as a union of semantic features for words in the sentence:
 

**Utterance:** { *mommy, ate, broccoli* }

**Scene:** { ANIMATE, HUMAN, ..., CONSUMPTION, ACTION, ... BROCCOLI, VEGETABLE, ... PLATE, OBJECT, ... }
- ▶ Word meaning is defined as a probability distribution over semantic features
- ▶ Word meanings are acquired using an incremental probabilistic alignment algorithm

## Word class induction

- ▶ Latent Dirichlet Allocation-based model: word types correspond to documents, context words correspond to words in documents
- ▶ Use an incremental version of the collapsed Gibbs sampler
 

```

for t = 1 → ∞ do
  for i = 1 → lt do
    sample zti ~ P(zti|zt-1, wti, dti)
    increment ntizti, wti and ntizti, dti
            
```
- ▶ Only condition on previous word tokens:
 
$$P(z_t | \mathbf{z}_{t-1}, \mathbf{w}_t, \mathbf{d}_t) \propto \frac{(n_{t-1}^{z_t, d_t} + \alpha) \times (n_{t-1}^{z_t, w_t} + \beta)}{\sum_{j=1}^{V_{t-1}} n_{t-1}^{z_t, w_j} + \beta}$$

## Word classes accelerate learning of meaning

- ▶ Alignment between words and semantic features is split into word-based and category-based components:

$$a(w|f, U^{(t)}, S^{(t)}) = \lambda^{(t)}(w) \times a_w(w|f, U^{(t)}, S^{(t)}) + (1 - \lambda^{(t)}(w)) \times a_c(w|f, U^{(t)}, S^{(t)})$$

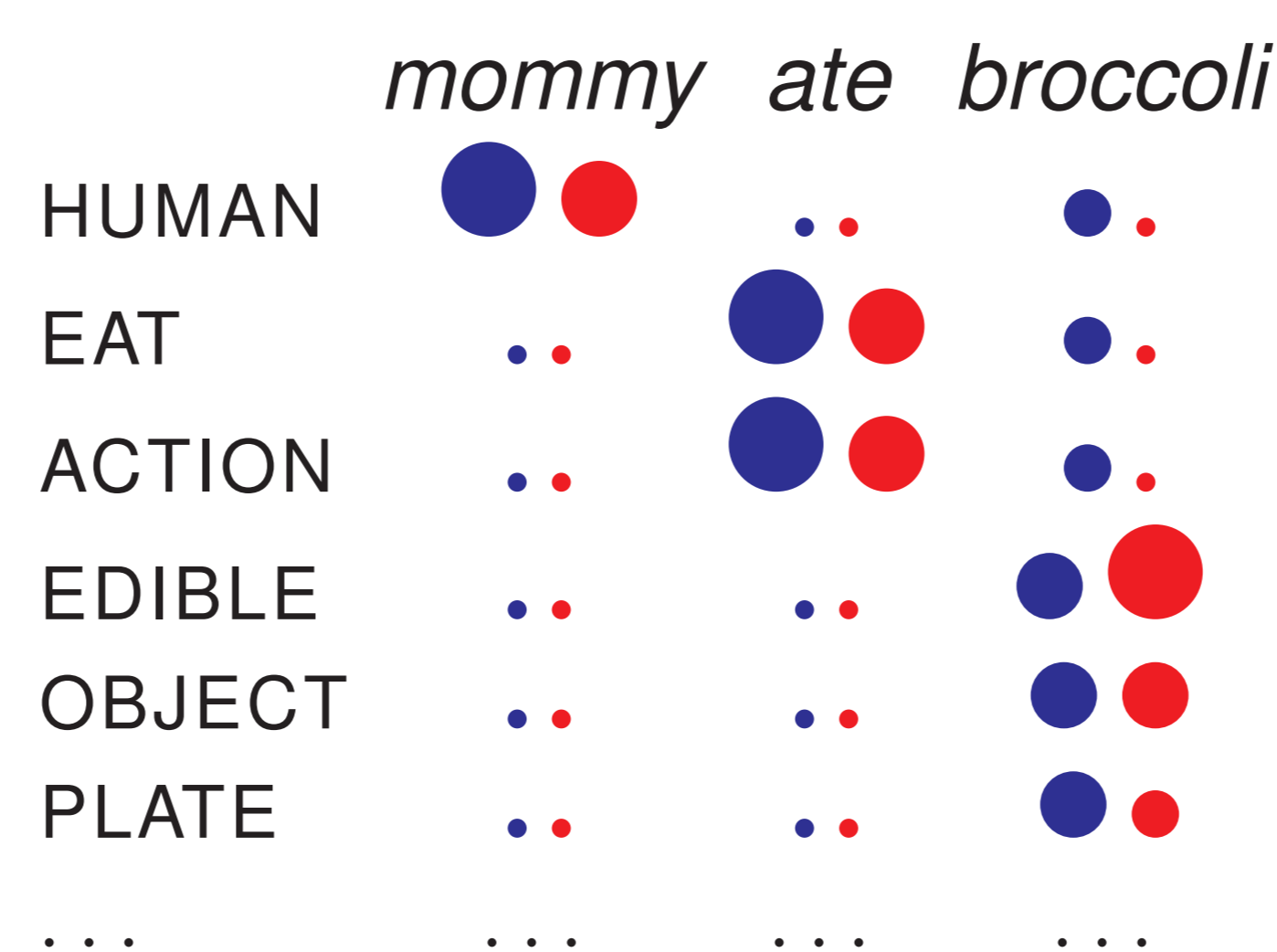
- ▶ Word-based alignment at time  $t$ :

$$a_w(w|f, U^{(t)}, S^{(t)}) = \frac{p^{(t-1)}(f|w)}{\sum_{w' \in U^{(t)}} p^{(t-1)}(f|w')}$$

- ▶ Category-based alignment at time  $t$ :

$$a_c(w|f, U^{(t)}, S^{(t)}) = \frac{p^{(t-1)}(f|\text{cat}(w))}{\sum_{w' \in U^{(t)}} p^{(t-1)}(f|\text{cat}(w'))}$$

- ▶ Interaction of cross-situational and distributional evidence:



- ▶ Computing category likelihoods from word meanings:

$$p^{(t)}(f|\text{cat}(w)) = \frac{1}{|\text{cat}(w)|} \sum_{w' \in \text{cat}(w)} p^{(t)}(f|w')$$

- ▶ The interpolation parameter depends on word frequency  $n^{(t)}(w)$ :

$$\lambda^{(t)}(w) = \frac{n^{(t)}(w)}{n^{(t)}(w) + 1}$$

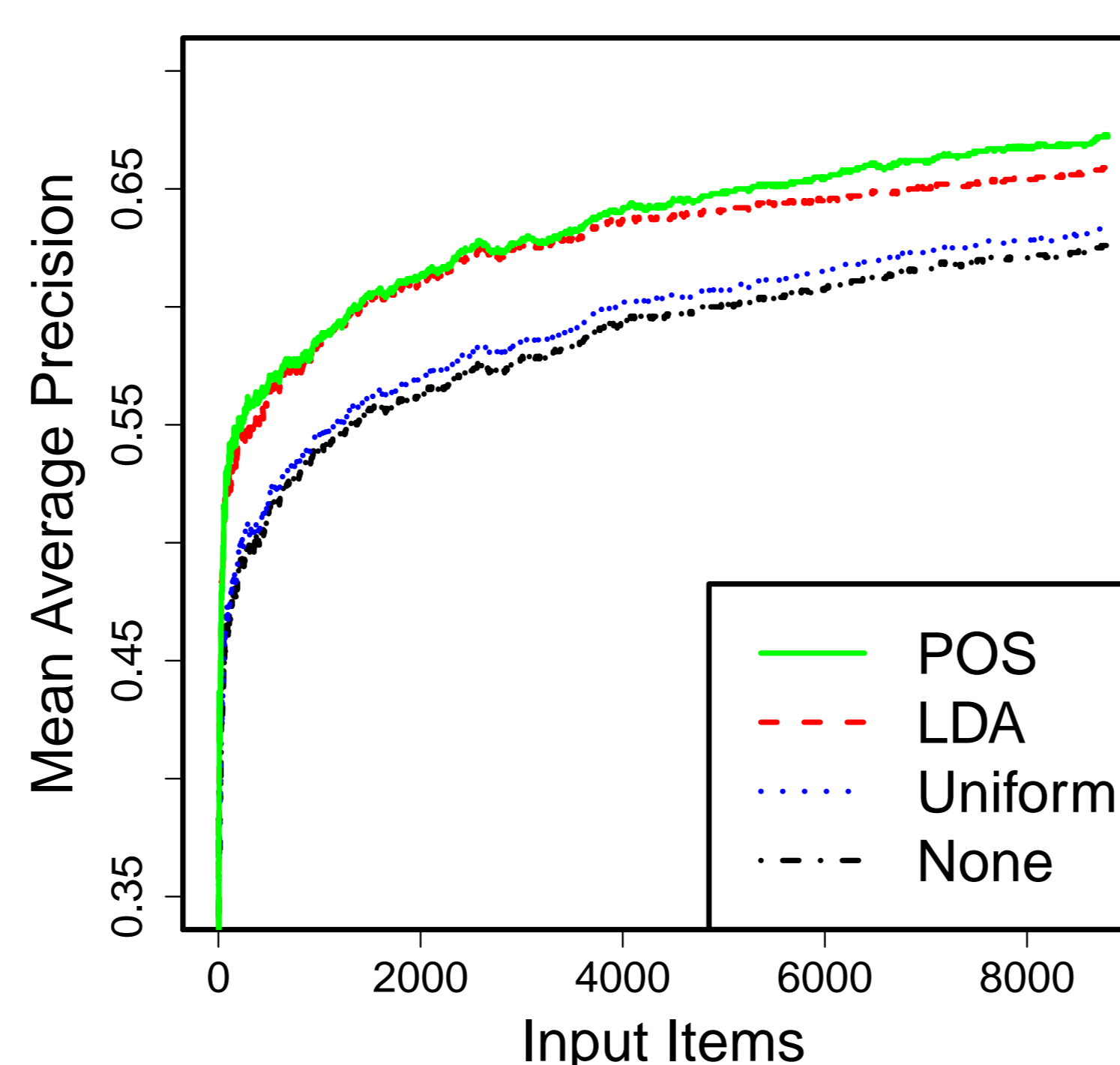
- ▶ Finally, word meanings are updated based on accumulated alignments:

$$p^{(t)}(f|w) = \frac{\sum_{i=1}^t a(w|f, U^{(i)}, S^{(i)})}{\sum_{f' \in F} \sum_{i=1}^t a(w|f', U^{(i)}, S^{(i)})}$$

## Sample categories

do are have can not go put did get play  
 is that it what not there he was where put  
 you not I the we what it they your a  
 to you we and I will not can it on  
 it a that the not he this right got she  
 are do is have on in can want did going  
 one I not shall there then you are we it  
 is in are on oh with and of have do  
 the a your of that it this some not very  
 going want bit go have look got will at little

## Learning curves



## Conclusion

- ▶ Categories induced from distributional cues improve cross-situational word learning
- ▶ Contribution of our categories is comparable to gold, manually assigned PoS tags
- ▶ LDA word class induction gives **soft** categories. In future work, we plan to exploit the whole distribution over categories for each word.

## References

- ▶ Alishahi, A. and Fazly, A. (2010). Integrating Syntactic Knowledge into a Model of Cross-situational Word Learning. CogSci'10.
- ▶ Chrupała, G. (2011). Efficient induction of probabilistic word classes with LDA. IJCNLP.
- ▶ Fazly, A., Alishahi, A., and Stevenson, S. (2010). A Probabilistic Computational Model of Cross-Situational Word Learning. Cognitive Science, 34(6):1017–1063.
- ▶ Yu, C. (2006). Learning syntax–semantics mappings to bootstrap word learning. CogSci'06.