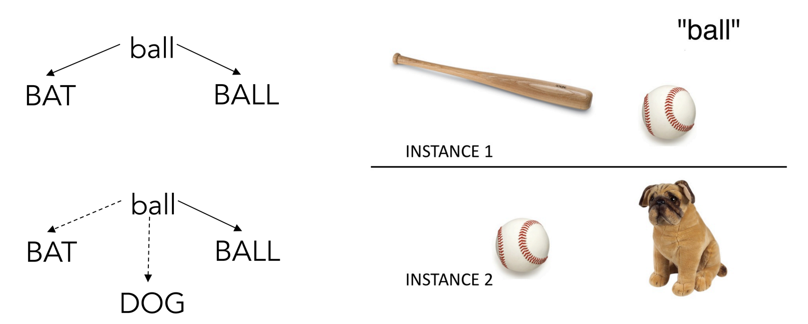
Early Syntactic Bootstrapping in an Incremental Memory-Limited Word Learner

Sepideh Sadeghi, Matthias Scheutz
Tufts University

INTRODUCTION

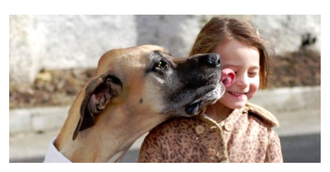
Cross-Situational Word Learning: Infants exploit co-occurrence statistics of words and objects across situations to learn the meaning of words [1].



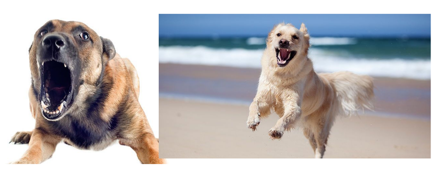
Problem: Cross-situational information is not always reliable as inconsistencies in the word-referent co-occurrence (i.e., when the referent is absent in a scene or when distracting referents are present) inject noise into cross-situational information.

Solution: Syntactic bootstrapping. Infants exploit the syntactic regularities in sentences of language to guide solving the word-to-meaning mapping problem [2-3].

The dog blicked the girl



The dog blicked



Research question: What syntactic regularities are available in early stages of language acquisition?

PRIOR WORK

Previous work on syntactic bootstrapping [4-8]

- studied the problem of joint acquisition in the context of ideal learners, ignoring the memory and computational limitations faced by a learner (i.e., an embodied robot).
- assumed prior access to syntactic concepts such as "subjecthood", lexical categories or more advanced syntactic knowledge (i.e., syntactic parses of the input sentences).
- simulation results from **ideal learners** and models **assuming prior access to structured syntactic knowledge** suggest that it is possible to jointly acquire word order and meanings and that learning is improved as each language capability bootstraps the other.

Our model is an extension of the Bayesian cross-situational model proposed in [9] and an improvement over the model proposed in [10] to handle variable-length utterances.

OBJECTIVES

- We propose that
  - distinguishing the referential words (with an object or event referent in the scene) of the utterance from the non-referential words is an initial step towards learning syntactic regularities.
  - relative order of appearance of the thematic roles associated with referential words in sentences of language provides an initial notion of word order in the absence of any prior syntactic knowledge (i.e., "subjecthood").
- We first present a probabilistic framework for early syntactic bootstrapping in the absence of any syntactic knowledge (concepts of "subjecthood", concepts of NP, VP, adverbs, determiners and other NP/VP modifiers), then we use our framework to study the utility of joint acquisition of word order and word referent and its onset.

PROBABLISTIC FRAMEWORK FOR EARLY SYNTACTIC BOOTSTRAPPING

Input Representation: situation = <utterance, scene>

utterance = W = {doggie, chases, the, cat}

W_R = {doggie, chases, cat} $W_{NR} = W - W_R$ = {the}

scene=E={{<agent,DOG>,<action,CHASE>,<patient,CAT>},
{<agent,CAT>,<action,FLEE>,<patient,DOG>},
{<agent,CAT>,<action,RUN>},
{<agent,DOG>,<action,RUN>}}

speaker's intention = I =
{<agent,DOG>,<action,CHASE>,<patient,CAT>}

Word Order Representation: ϴ

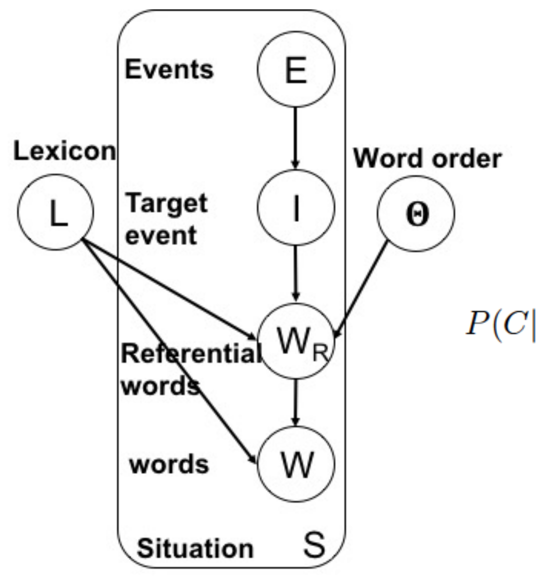
Assuming the learner knows three thematic roles:

$\Theta = \{\theta_1, \theta_2, \theta_3\}$, $\theta_{role\_i} = P(.|role\_i) = multinomial(\pi_1, \pi_2, \pi_3)$, $\pi_1 = P(role\_1|role\_i)$

Models:

- M-WO: joint learner of word order and word referent
- M-B: baseline model - without ϴ

M-WO:



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

$$P(\Theta) \propto 1$$

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

$$P(C|L,\Theta) = \prod_{s\in C}\sum_{I_s\subseteq E_s}\sum_{W_R\subseteq W_s} P(W_s|I_s,W_R,L,\Theta)\cdot P(I_s|E_s)$$

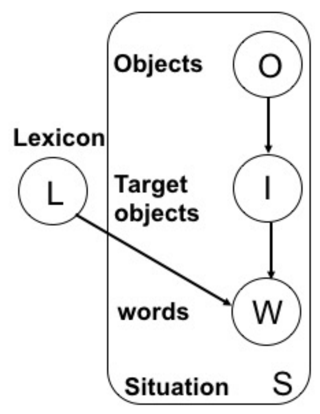
$$P(I_s|E_s) \propto 1$$

$$\boxed{P(W_s|I_s,W_R,L,\Theta) = P(W_R|I_s,L,\Theta)\cdot P(W_{NR}|L)}$$

$$\boxed{P(W_{NR}|L) = \prod_{w_k\in W_{NR}}(1-\gamma)P_{NR}(w_k|L)}$$

$$\boxed{P(W_R|I_s,L,\Theta) = \prod_{w_j\in W_R}[\gamma\sum_{x_i\in I_s}[\frac{1}{|I_s|}P_R(w_j|x_i,L)\cdot\sum_{y_k\in I_s}\frac{1}{|I_s|}P_R(w_{j-1}|y_k,L)\pi_{(role(x_i)|role(y_k))}]]}$$

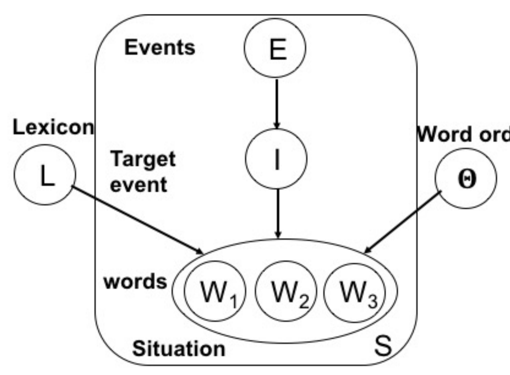
Previous Model Designs



1. Frank, Goodman & Tenenbaum 2009

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

2. Sadeghi, Scheutz 2017.



$$\boxed{P(W_s|I_s,L,\Theta) = \prod_{w_j\in W_s}[\gamma\cdot\sum_{x_i\in I_s}\frac{1}{|I_s|}P_R(w_j|x_i,L)\cdot P(pos(w_j)|role(x_i),\Theta) + (1-\gamma)P_{NR}(w_j|L)]}$$

INCREMENTAL AND MEMORY-LIMITED LEARNING ALGORITHM

Model's memory consists of the knowledge in its lexicon and current situation.

Incremental learning algorithm sees each situation only once (no iteration over data).

- utilizes a Bayesian approach for joint acquisition of the word-object mappings and speaker's intention in each situation.
- utilizes association strength of the word-object mappings for incremental integration of the newly inferred word-object mappings (mini-lexicon) in the previous lexicon (full lexicon) and performing conflict resolution over alternative mappings.

Updating Lexicon

- Inferring the MAP mini-lexicon in each situation
  - generating mini-lexicon proposal (hypothesis generation) … semi-stochastic search
  - scoring (hypothesis evaluation) … relative posterior probability
- Merging the new mini-lexicon with the current lexicon
  - applying soft mutual-exclusivity constraints to produce a preference for one-to-one mappings in the output lexicon.

Updating Word Order

Using a symmetric Dirichlet Prior with parameter β:

$$\pi_{(role_j|role_i)} = \frac{Count(role_j|role_i)+\beta}{\sum_k Count(role_k|role_i)+n\beta}$$

EVALUATION DATA

- We evaluated M-WO and M-B in different ambiguous contexts using the datasets described in Table 1 (each dataset consists of 500 trials). These datasets differ from each other in the source and level of their ambiguity.

Table 1: Sources of ambiguity in evaluation data.

Data	Distracting Events	Non-Ref Words	Prepositions
D1	No	No	5
D2	No	1 per utterance on average	5
D3	1 for verbs in {drop,feed,take,give}	No	5

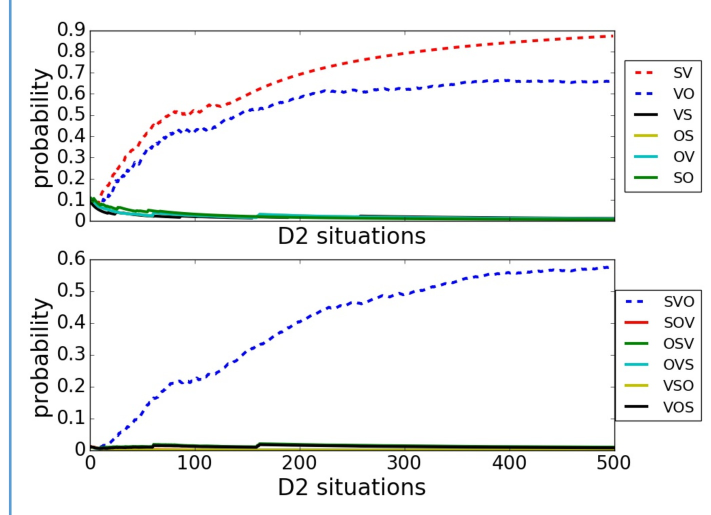
Table 2: Example verb frames for "eat". Except for terminals, the rest of the variables are place-holders for a set of other terminals, variables or a combination of both.

ID	Frame
1	Modifier N-animate (terminal 'eat)
2	Modifier N-animate (terminal 'eat) Modifier N-edible
3	Modifier N-animate (terminal 'eat) Modifier N-edible (terminal 'in) LOC

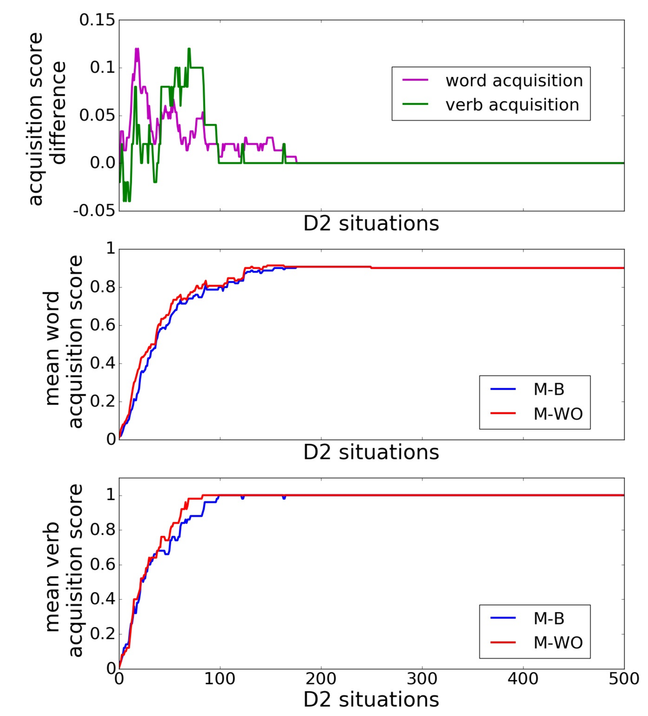
- We used a probabilistic generative process to randomly generate 500 utterances with 10 verbs= {falls,drops,pushes,pulls,takes,gives,eats,feeds,drinks,reads} and 20 objects.
- Our data-generation lexicon also includes five prepositions, ten adjectives and three determiners. Overall, we used 48 frames (with SVO word order), a subset of which is depicted in Table 2.

RESULTS

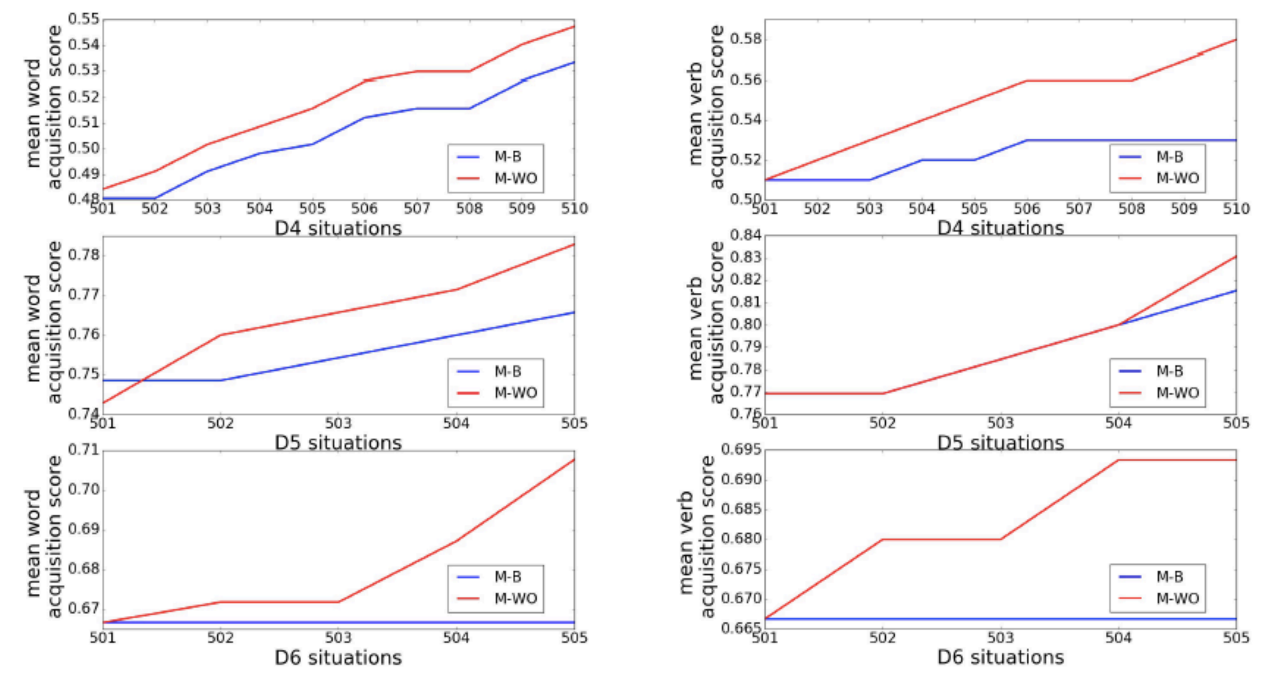
Incremental Acquisition of Word Order



Incremental Acquisition of Meaning

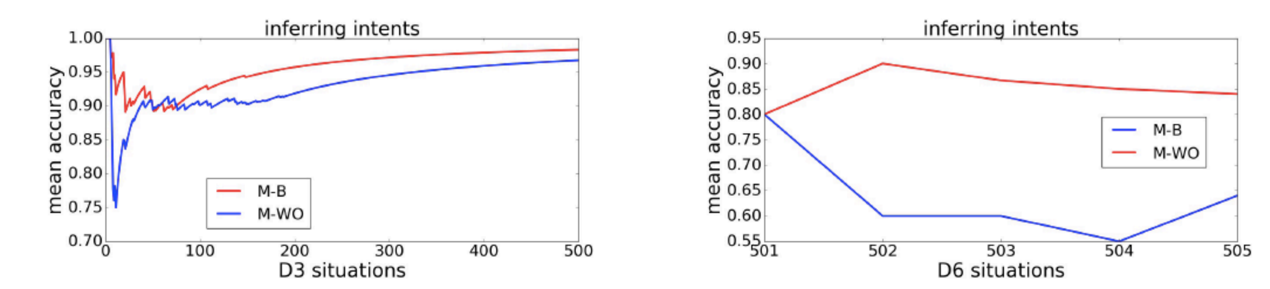


One-Shot Word Learning



(a) word acquisition (b) verb acquisition

Inferring Speaker's Intentions in Ambiguous Visual Contexts



(a) averaged over D3 trials. (b) averaged over D6 trials.

CONCLUSIONS

- Transitional probabilities of the thematic roles associated with the words referring to event participants (concrete objects) and events (actions) can guide early acquisition of the notion of word order before syntactic concepts are available to the learner.
- Our results regarding the general utility of joint acquisition in improving the word learning results aligns with previous results, but suggest that there is a time lag for the emergence of this advantage, during which the acquired word order is being improved.
- The proposed model
  - allows for learning synonyms.
  - handles variable-length real-world utterances.
  - handles online open-world word learning.

FUTURE WORK

- Extend the framework to accommodate learning the structural rules of NPs (i.e., to learn that color modifiers cannot be followed by size modifiers but the opposite is likely as in "the large red box").
- Computational experiments with with different sets of thematic roles, varying the specificity versus generality of the roles, to shed light on whether human-like notions of thematic roles are required for word order acquisition.

ACKNOWLEDGEMENTS: This work was funded by Vienna Science and Technology Fund project ICT15-045.

REFERENCES: [1] Quine 1960, Word and Object. MIT Press. [2] Arunachalam & Waxman 2010. Meaning from syntax: Evidence from 2-year-olds. Cognition. [3] Yuan, Fisher, & Snedeker 2012. Counting the nouns. Child Dev 83. [4] Alishahi & Chrupala 2012. Concurrent acquisition of word meaning and lexical categories. EMNLP.[5] Alishahi & Fazly 2010. Integrating syntactic knowledge into a model of cross-situational word learning. CogSci. [6] Yu, C. 2006. Learning syntax–semantics mappings to bootstrap word learning. CogSci. [7] Maurits, Perfors & Navarro 2009. Joint acquisition of word order and word reference. CogSci. [8]Abend et al. 2017. Bootstrapping language acquisition. Cognition. [9] Frank, Goodman & Tenenbaum 2009. Using speakers' referential intentions to model early crosssituational word learning. Psychological Science. [10] Sadeghi, Scheutz 2017. Joint acquisition of word order and word referent in a memory-limited and incremental learner. CogInfoCom.

Early Syntactic Bootstrapping in an Incremental Memory-Limited Word Learner

Sepideh Sadeghi, Matthias Scheutz

- How does the usage of words in sentences of language guide word learning in early stages of language acquisition in the absence of prior syntactic knowledge?

    a) Dog blicked

    b) Dog blicked Sarah

    c) Sarah blicked the dog

(c)

(b)

Early Syntactic Bootstrapping in an Incremental Memory-Limited Word Learner

- Previous work on syntactic bootstrapping (Yu 2006; Maurits, Perfors, and Navarro 2009; Alishahi and Fazly 2010; Alishahi and Chrupała 2012; Abend et al. 2017)
  - Studied the problem of joint acquisition in the context of ideal learners, ignoring the memory and computational limitations faced by a learner (e.g., an embodied robot).
  - Assumed prior access to syntactic concepts such as “subjecthood”, lexical categories or more advanced syntactic knowledge such as syntactic parses of the input sentences.

- Our contribution: A probabilistic framework for early syntactic bootstrapping
  - Bayesian model of cross-situational word learning
  - Limited memory of past observation (evidence)
  - Incremental learning algorithm
  - Relies on no prior syntactic knowledge
  - Jointly learns the word referent, speaker’s intentions, and language word order