

An incremental model of syntactic bootstrapping

Christos Christodoulopoulos, Dan Roth, Cynthia Fisher

University of Illinois at Urbana-Champaign

{christod,danr,clfishe}@illinois.edu

introduction

- Language learners must map lexical units to underlying semantic representations
- Syntactic bootstrapping:** learners use syntactic structure to acquire the meanings of novel verbs
- Structure Mapping:**
 - children have an innate one-to-one mapping between nouns and semantic arguments
 - they use this information to identify verbs and assign semantic roles to their arguments
- Connor et al. (2012): computational model of syntactic bootstrapping via structure mapping
- Current work: two improvements
 - Incremental learning:** predictions are made as soon as the learner receives input
 - Aggregated verb predictions:** distributional clusters behave in a grammatically consistent fashion



The girl **krads** the boy

girl: noun
boy: noun
krads: **verb**

syntactic
bootstrapping

(BabySRL features)

girl: agent
boy: patient
krads: **chase**

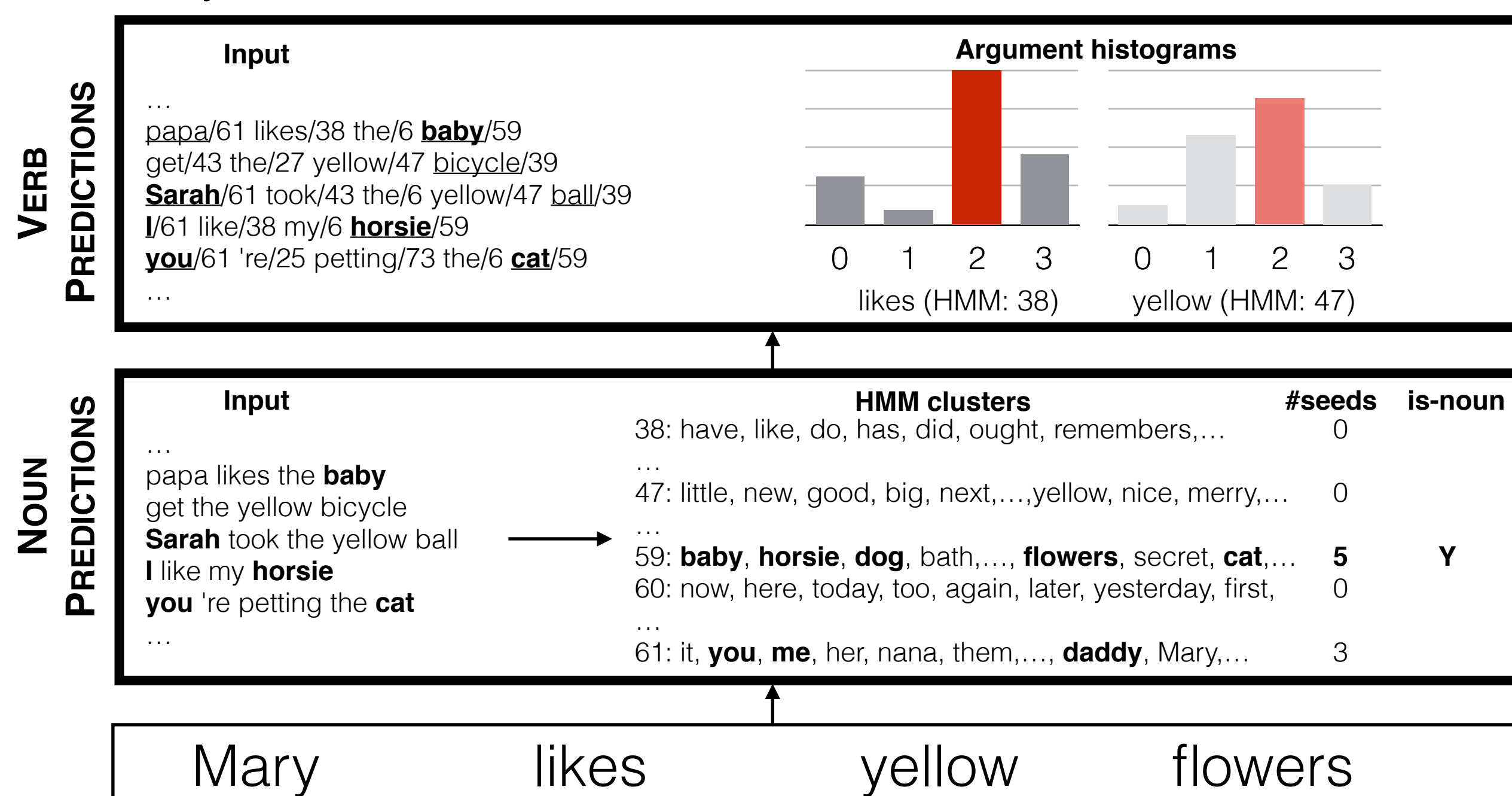
input → syntax semantics

seed nouns

Assumptions:

- there exist verbs
- nouns are not verbs
- once a verb, always a verb

incremental heuristics



example

1: papa/57 wants/58 an/6 **apple**/39

2: get/43 the/27 red/79 **bicycle**/39

3: come/75 and/21 move/43 **horsie**/39

4: i/50 forgot/58 a/6 **spoon**/39

5: you/50 're/25 eating/73 the/6 **broom**/39

System state

state	seeds	frags	is-noun	is-verb	ratio
39	1		Y*	1.0	
57		1		0.0	
58					
43					
79					
75					
59					
73					

state	seeds	frags	is-noun	is-verb	ratio
39	2			1.0	
57				0.0	
58				0.0	
43	1		Y*	1.0	
79		1		0.0	
75					
59					
73					

state	seeds	frags	is-noun	is-verb	ratio
39	3			1.0	
57				0.0	
58				0.0	
43	1		Y	1.0	
79		1		0.0	
75					
59					
73					

state	seeds	frags	is-noun	is-verb	ratio
39	4		Y	1.0	
57				0.5	
58	2		Y	1.0	
43				0.0	
79				0.0	
75					
59	1			0.0	
73					

state	seeds	frags	is-noun	is-verb	ratio
39	4		Y	1.0	
57				0.5	
58				1.0	
43				1.0	
79				0.0	
75					
59	2			0.0	
73		2	Y	1.0	

Predictions

apple: noun (seed)
papa: verb (random)

bicycle: noun (seed)
get: verb (random)

horsie: noun (seed)
move: verb
(st 43 - ratio 1.0 vs st 75 - 0.0)

spoon: noun (seed)
i: noun (seed)
forgot: verb (only 1)

broom: noun (state 39)
you: noun (seed)
eating: verb (only 1)

conclusion

- Given enough seed nouns, incremental model achieves **similar performance** within around **2000 sentences** for noun predictions and **3000 sentences** for verb predictions
- Number of seed nouns** shown to be sufficient in the previous work **is not sufficient** under more realistic incremental model
- Adopting more realistic assumptions** about the early stages of language acquisition can **tell us more about what learners require** to bootstrap the acquisition of syntactic categories while **maintaining high performance**

references

Michael Connor, Cynthia Fisher, and Dan Roth. 2012. Starting from scratch in semantic role labeling: Early indirect supervision. In A. Alishahi, T. Poibeau, and A. Korhonen, editors, *Cognitive Aspects of Computational Language Acquisition*. Springer.

results

