Cognitive and application-driven ML for natural language

Christos Christodoulopoulos

OSU - April 5, 2016

COGNITIVE COMPUTATION GROUP

UNIVERSITY OF ILLINOIS AT URBANA - CHAMPAIGN

Credit: Josh Lee (@wtrsld)

Natural Language (Processing) tasks

Credit: Josh Lee (@wtrsld)

Translation

Sentiment Analysis

Question Answering

Following Instructions

Human Interaction

Natural Language (Processing) tasks

Credit: Josh Lee (@wtrsld)

Machine Translation

Sentiment Analysis

Question Answering

Following Instructions

Human Interaction

ASR/OCR

Morphological Analysis

PoS Tagging

Syntactic Parsing

Semantic Parsing

Reasoning

Generation

Natural Language (Processing) tasks

Credit: Josh Lee (@wtrsld)

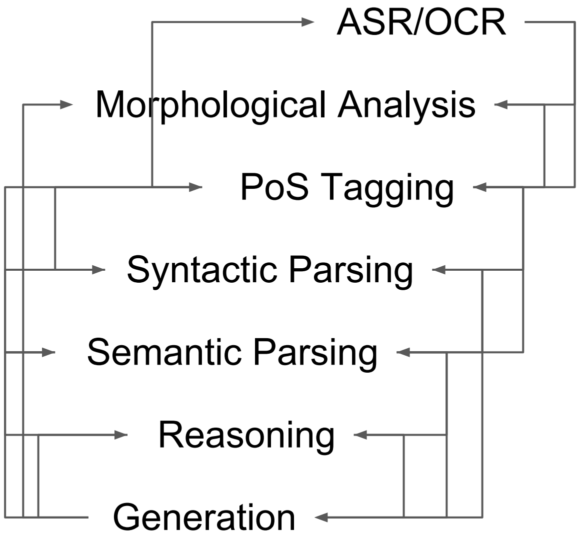
Machine Translation

Sentiment Analysis

Question Answering

Following Instructions

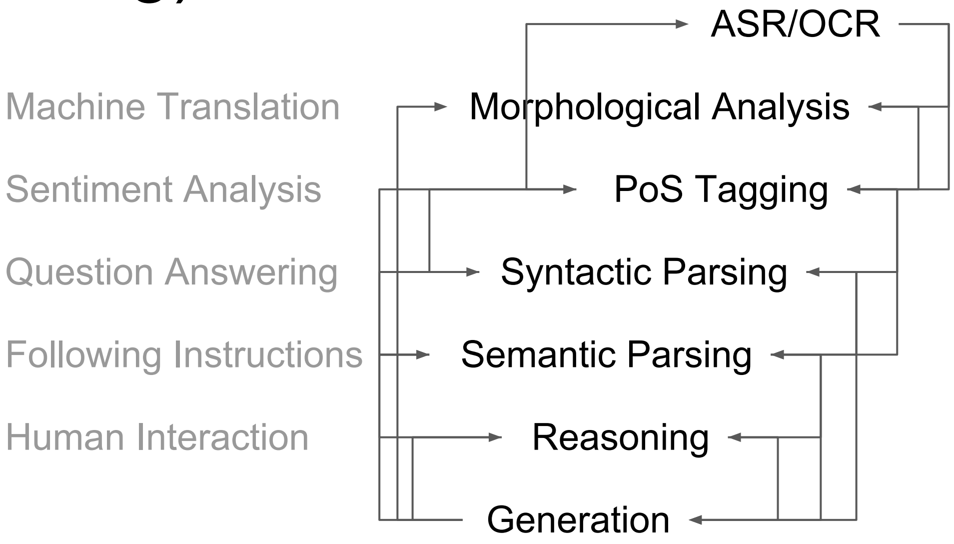
Human Interaction



Not really a pipeline!

Natural Language (Processing) tasks

Credit: Josh Lee (@wtrsld)



Arguments for less supervision

- Expensive (time/money)
- Not always reliable or “correct”
- **Not how humans learn**

Natural Language (Processing) tasks

ASR/OCR

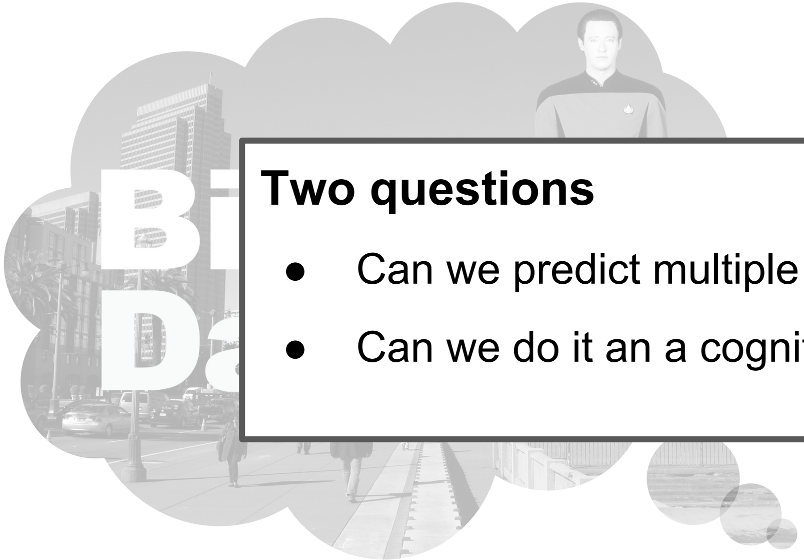
Machine Translation

Morphological Analysis

Sentiment Analysis

PoS Tagging

Generation



Credit: Josh Lee (@wtrsld)

Two questions

- Can we predict multiple NL phenomena jointly and efficiently?
- Can we do it an a cognitively inspired unsupervised way?

Arguments for less supervision

- Expensive (time/money)
- Not always reliable or “correct”
- **Not how humans learn**

Outline

- Application-Driven ML: **Extended SRL**
  - Modular semantic structures
  - Inference over independently-trained models
- Cognitive ML: **Baby SRL**
  - Modeling early stages of language acquisition
  - Testbed for psycholinguistic theories and unsupervised ML
- Unifying the two: **Unsupervised CCG induction**
  - Structured lexical-syntactic-semantic representation
- Future directions

Extended SRL (Joint work with Dan Roth)

- Semantic role labeling covers only specific predicate-argument relations
- Need for comprehensive semantic representations (SRL++)
  - Events, Entailment, Winograd schemas
- Difficult to produce hand-annotated resources (e.g. AMR, λ-calculus)
  - Especially for other languages/genres†
- **Lots** of independently annotated data for semantic tasks
  - Ontonotes, PDTB, Semeval/CoNLL/*SEM shared tasks

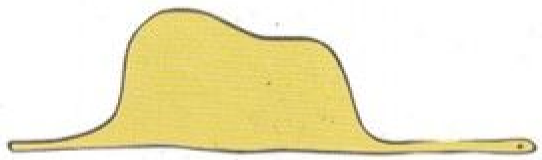
†AMR annotation of The Little Prince
http://amr.isi.edu/download/amr-bank-v1.6.txt

Extended SRL (Joint work with Dan Roth)

- Semantic role labeling covers only specific predicate-argument relations
- Need t
  - Ev
- Difficu … s)
  - Es
- **Lots c**
  - Or

It was a picture of a boa constrictor in the act of swallowing an animal .

```
(p / picture
  :domain (i / it)
  :topic (b2 / boa
          :mod (c2 / constrictor)
          :ARG0-of (s / swallow-01
                    :ARG1 (a / animal))))
```



†AMR annotation of The Little Prince

http://amr.isi.edu/download/amr-bank-v1.6.txt

Extended SRL (Joint work with Dan Roth)

- Semantic role labeling covers only specific predicate-argument relations
- Need for comprehensive semantic representations (SRL++)
  - Events, Entailment, Winograd schemas
- Difficult to produce hand-annotated resources (e.g. AMR, λ-calculus)
  - Especially for other languages/genres†
- **Lots** of independently annotated data for semantic tasks
  - Ontonotes, PDTB, Semeval/CoNLL/*SEM shared tasks

†AMR annotation of The Little Prince
http://amr.isi.edu/download/amr-bank-v1.6.txt

Extended SRL

Modular semantic representation

Inference with independent models

Extended SRL

Modular semantic representation

Inference with independent models

Tokens	Five	prisoners	are	held	captive	in	Guantanamo

Extended SRL

Modular semantic representation

Inference with independent models

Tokens	Five	prisoners	are	held	captive	in	Guantanamo
Quantities	[#5] [unit: prisoner]						

Extended SRL

Modular semantic representation

Inference with independent models

Tokens	Five	prisoners	are	held	captive	in	Guantanamo
Quantities	[#5] [unit: prisoner]						
NER	Card.						Location

Extended SRL

Modular semantic representation

Inference with independent models

Tokens	Five	prisoners	are	held	captive	in	Guantanamo
Quantities	[#5] [unit: prisoner]						
NER	Card.						Location
VerbSRL	Arg1			hold	Arg2		
NomSRL	Arg1				captive	ArgM-LOC	
PrepSRL				Gov		Loc	Obj

Extended SRL

Modular semantic representation

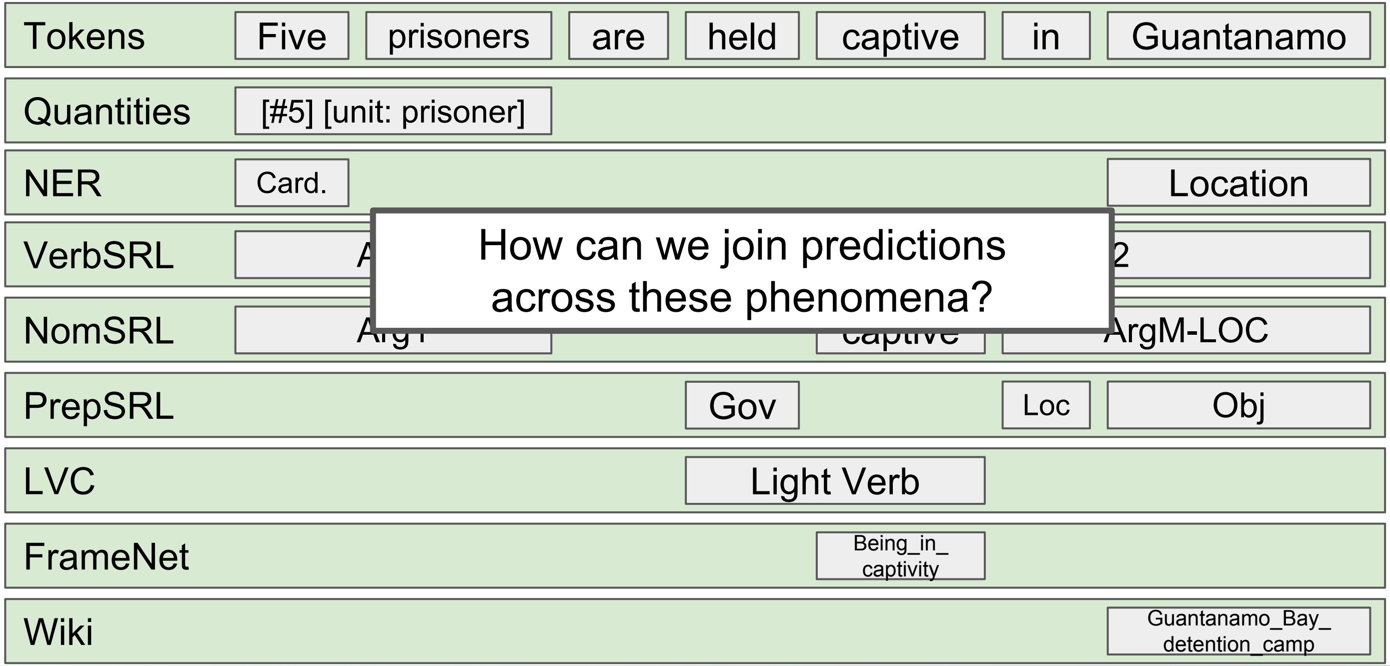
Inference with independent models

Tokens	Five	prisoners	are	held	captive	in	Guantanamo
Quantities	[#5] [unit: prisoner]						
NER	Card.						Location
VerbSRL	Arg1			hold	Arg2		
NomSRL	Arg1				captive	ArgM-LOC	
PrepSRL				Gov		Loc	Obj
LVC				Light Verb			
FrameNet					Being_in_captivity		
Wiki							Guantanamo_Bay_detention_camp

Extended SRL

Modular semantic representation

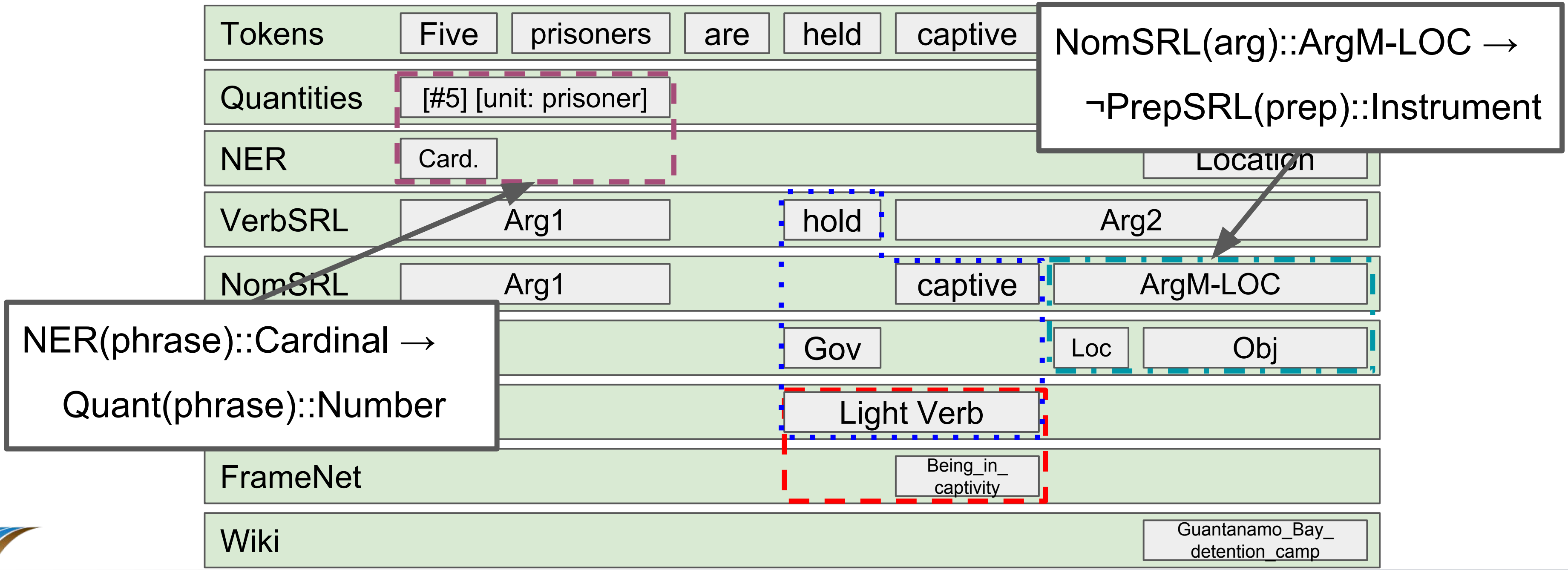
Inference with independent models



Extended SRL

Modular semantic representation

Inference with independent models



Constraint-driven inference

- Constrained Conditional Models [Chang et al., 2012]
  - First-order constraints
  - ILP inference
- Multi-view model combination [Burkett et al., 2010]

Constraint-driven inference

- Constrained Conditional Models [Chang et al., 2012]
  - First-order constraints
  - ILP inference
- Multi-view model combination [Burkett et al., 2010]

$$\boxed{(\hat{y}_1, \hat{y}_2)} = \arg\max_{\boxed{y_1^\star, y_2^\star}} \boxed{\lambda_1} y_1^\star + \boxed{\lambda_2} y_2^\star - \sum_{k=1}^{m} \boxed{\rho_k} \boxed{C_k(\mathbf{x}, \mathbf{y}^\star)}$$

joint prediction

independent predictions

model weights

violation penalty (here: ∞)

constraints

Constraint-driven inference

- Constrainec
  - First-ord
  - ILP infere
- Multi-view r

Algorithm 1: Weight update algorithm

Input: Models $\mathbf{m} = \{m_1, m_2, \ldots, m_n\}$
Input: Labeled Corpora $\{\mathbf{D_1}, \mathbf{D_2}, \ldots, \mathbf{D_n}\}$
Output: Model weights

for $m_i \in \mathbf{m}$ **do**
  $\eta_i \leftarrow 0.1;$
  for $(x_j, y_j) \in \mathbf{D_i}$ **do**
    if $\hat{y}_{m_i}(x_j) \neq y_j \wedge \sum_{k=1}^{m} C_k(\mathbf{x}, \mathbf{y}^\star) > 0$ **then**
      $\lambda_i \leftarrow \lambda_i + \eta_i / \sqrt{j}$
    end
    train;
  end
end

$(\hat{y}_1, \hat{y}_2$ ... $\boxed{\rho_k}\boxed{C_k(\mathbf{x}, \mathbf{y}^\star)}$

joint prediction

ation penalty
e: ∞)

constraints

Extended SRL: Systems

- Clauses
- *Comma SRL
- *Coreference resolution
- FrameNet
- Light-verb constructions
- Metonymy
- Multi-word expressions
- NER
- *Nominal SRL
- Phrasal-verb constructions
- PP attachment
- *Preposition SRL
- Quantities
- Sentiment analysis (aspect-based)
- Sentence specificity
- Temporal extraction
- *Verb SRL
- VP ellipsis
- *Wikification

* denotes previously implemented system

Detour: CogComp software

- LBJava [Rizzolo & Roth, 2010]
  - Rapid development of ML software

- EDISON + TextAnnotation [Sammons et al., 2016]
  - Data structures and feature extraction

Detour: CogComp software

- LBJava [Rizzolo & Roth, 2010]
  - Rapid development of ML software

- EDISON + TextAnnotation [Sammons et al., 2016]
  - Data structures and feature extraction

NER system definition:

```
import edu.illinois.cs.cogcomp.esrl.core.features.*;
import edu.illinois.cs.cogcomp.core.datastructures.textannotation.Constituent;

discrete NERLabel(Constituent phrase) <- { return phrase.getLabel(); }

discrete NERClassifier(Constituent phrase) <-
  learn NERLabel
  using Capitalization, WordBigrams, POSBigrams, WordContextBigrams,
        POSContextBigrams, ChunkContextBigrams

  with SparseNetworkLearner {
    SparseAveragedPerceptron.Parameters p =
        new SparseAveragedPerceptron.Parameters();
    p.learningRate = .1;
    p.thickness = 2;
    baseLTU = new SparseAveragedPerceptron(p);
  }
end
```

Detour: CogComp software

- LBJava [Rizzolo & Roth, 2010]
  - Rapid development of ML software
- EDISON + TextAnnotation [Sammons et al., 2016]
  - Data structures and feature extraction

NER system definition:

Output label

Classifier definition

```
import edu.illinois.cs.cogcomp.esrl.core.features.*;
import edu.illinois.cs.cogcomp.core.datastructures.textannotation.Constituent;

discrete NERLabel(Constituent phrase) <- { return phrase.getLabel(); }

discrete NERClassifier(Constituent phrase) <-
  learn NERLabel
  using Capitalization, WordBigrams, POSBigrams, WordContextBigrams,
          POSContextBigrams, ChunkContextBigrams
  with SparseNetworkLearner {
    SparseAveragedPerceptron.Parameters p =
          new SparseAveragedPerceptron.Parameters();
    p.learningRate = .1;
    p.thickness = 2;
    baseLTU = new SparseAveragedPerceptron(p);
  }
end
```

Extended SRL: Constraints

- $[\text{LVC(phrase)::true} \rightarrow \neg\text{PVC(phrase)::true}] \wedge$
  $[\text{PVC(phrase)::true} \rightarrow \text{phrase contains } \{\text{IN}|\text{PRT}|\text{RB}\}]$
    - LVC accuracy: 81.2 → 82.2
    - Same PVC accuracy (89.9)
    - Candidate selection prevented constraint violation

Extended SRL: Constraints

- [LVC(phrase)::true → ¬PVC(phrase)::true] ∧ [PVC(phrase)::true → phrase contains {IN|PRT|RB}]
    - LVC accuracy: 81.2 → 82.2
    - Same PVC accuracy (89.9)
    - Candidate selection prevented constraint violation
- Quantities(phrase)::Date → Temporal(phrase)::Date
    - Quantities overall F1: 78.7 → 78.9 (Date acc: 79.2 → 85.5)
    - Temp overall F1: 75.7 → 75.9
    - Double implication doesn't work

Extended SRL: Constraints

- NER(phrase)::Cardinal → Quant(phrase)::Number
- NER(phrase)::Date → Temp(phrase)::Date
- **NER(phrase)::Date → Temp(phrase)::Date ∧ ¬Quant(phrase)::Number**
- VerbSRL(arg)::ArgM-LOC → ¬PrepSRL(prep)::Instrument
- **Metonymy(x)::true ∧ NER(x)::y → ∃z ∀x' Met(x')::false ∧ NER(x')::z ∧ z≠y**

Extended SRL: Summary

- Combination of multiple phenomena
  - No need for joint annotations
- Joint inference via first-order constraints
  - Offer linguistic insights
- Flexible interface
  - Only requirement is list of *k*-best predictions

Extended SRL: Summary

- Combination of multiple phenomena
  - No r
- Joint inf
  - Offe
- Flexible
  - Only

But can we do it with less (direct) supervision?

- Other languages
- “Better-than-gold” performance [e.g. Spitkovsky et al., 2011]

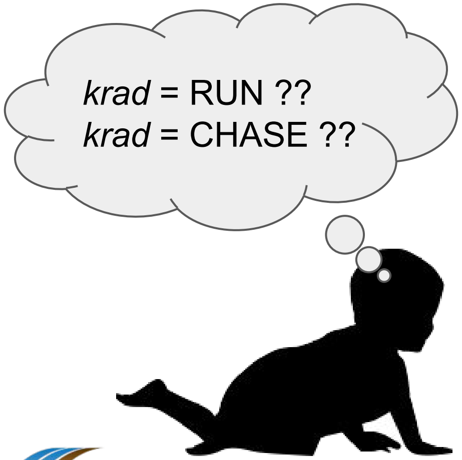
One solution: Look into human language acquisition

- Cognitive insights for ML models
- Testbed for cognitive theories

A (simple) model of language acquisition

“The girl ***krads*** the boy”

“The boy ***krads***”



Baby SRL (Joint work with Cindy Fisher and Dan Roth)

- Syntactic bootstrapping
  - Using the structure of the utterances to predict the semantic
- An account of how syntactic bootstrapping can begin
  - **Connor et al. (2010)**
  - Fisher et al. (2010)
  - Gutman et al. (2014)
  - van Schijndel & Elsner (2014)
- Framed as an SRL problem
  - Learn Agent/Patient roles for novel-verb utterances

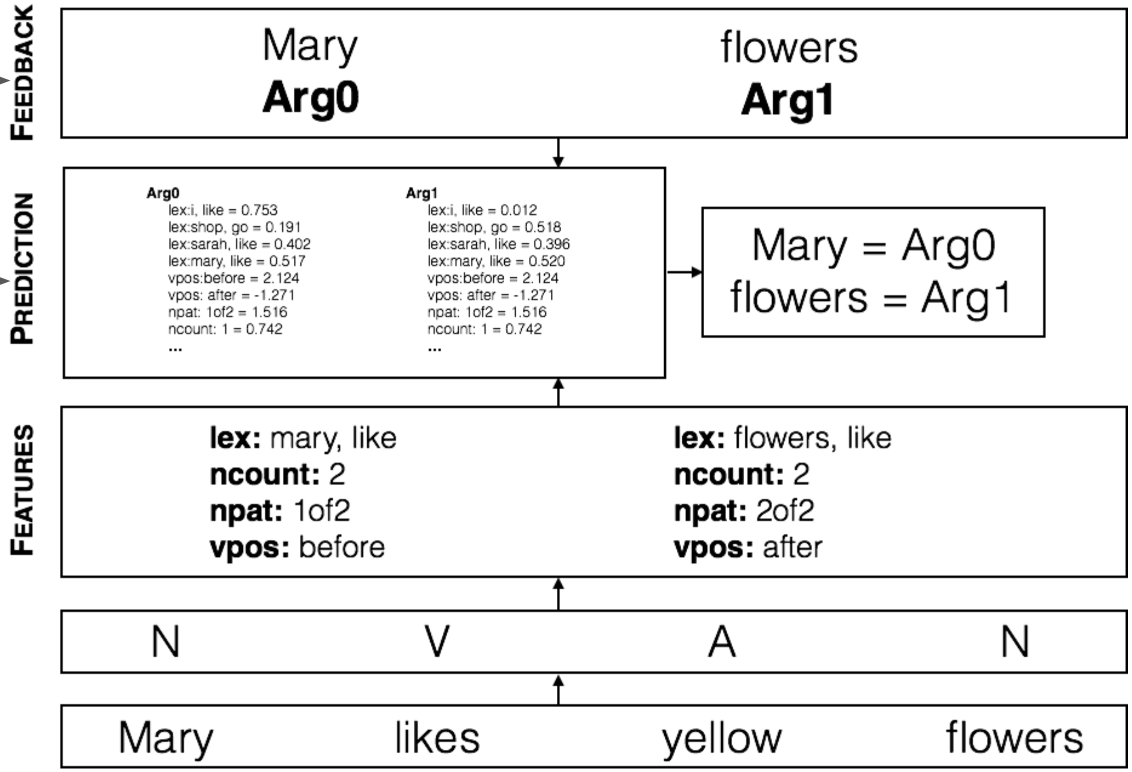
Experiment 1: Supervised model

Given **veridical feedback** (“mind reading”), do **low-level** syntactic features capture anything useful about semantic roles/verb preferences?

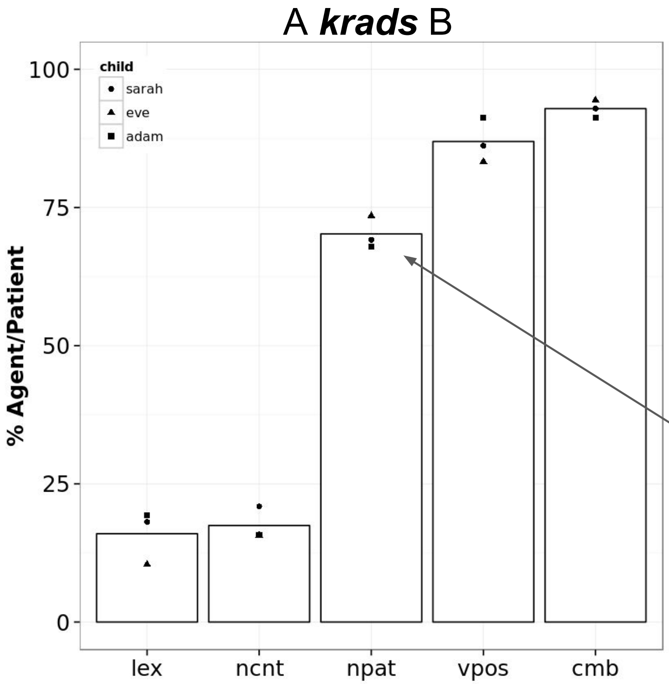
Experiment 1: Supervised model

BabySRL corpus
[Connor et al. 2008]

Online Percerptron
α=0.1, θ=4.0

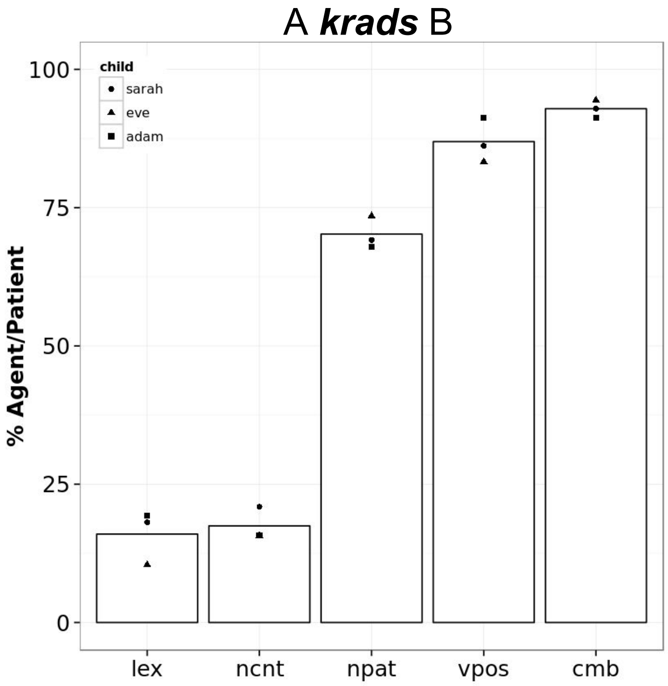


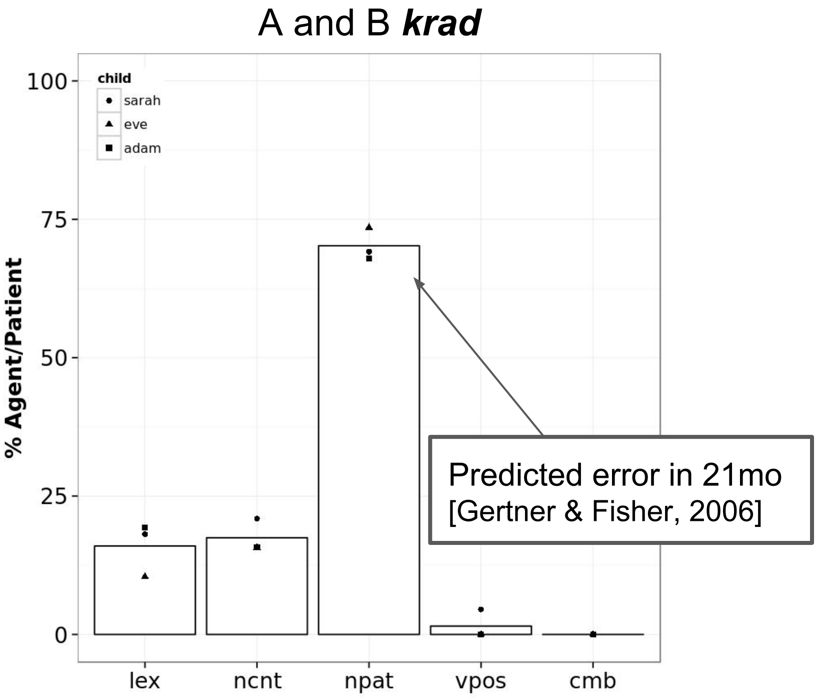
Results on novel-verb sentences



Structure-Mapping hypothesis
[Fisher, 1996]

Results on novel-verb sentences





Experiment 1: Supervised model

Given **veridical feedback** (“mind reading”), do **low-level** syntactic features capture anything useful about semantic roles/verb preferences?

YES, but verb knowledge is crucial

Experiment 2: Removing bottom-up supervision

Can we predict nouns/verbs using

distributional **clusters** and a few **seed nouns**?

Experiment 2: Predicting nouns

- HMM (80 states) + Variational Inference
  - Trained on 2.2M tokens of CDS
  - List of function words to separate clusters
- List of seed nouns
  - MacArthur-Bates CDI production norms [Dale & Fenson, 1996]
  - 75 nouns+pronouns (cutoff 50% at 21-mo)

Noun identification:
state *x* contains > ***k*** seed nouns

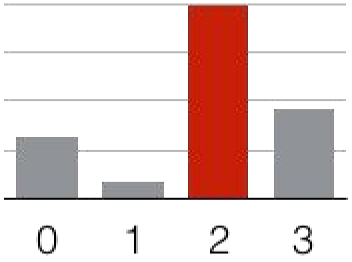
where ***k*** = 4 [or dynamic]

Experiment 2: Predicting verbs

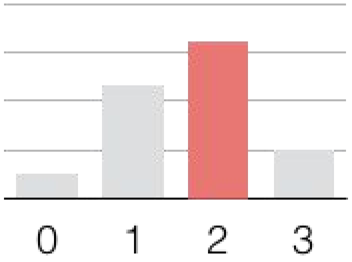
Step 1: Argument histograms

for each sentence:
 count the number of nouns
 collect histograms for each non-noun state

38: have, like, do, has, did, ought, remembers,…
…
47: little, other, new, good, big, next,…,yellow, nice, merry,…
…
59: **baby**, **horsie**, **dog**, bath,…, **flowers**, song, secret, **cat**,…
60: now, here, today, too, again, later, yesterday, first,…
61: it, **you**, **me**, her, nana, them,…, **daddy**, Mary,…



likes (HMM: 38)



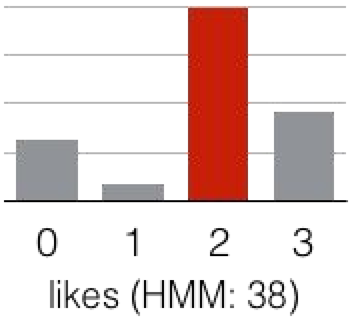
yellow (HMM: 47)

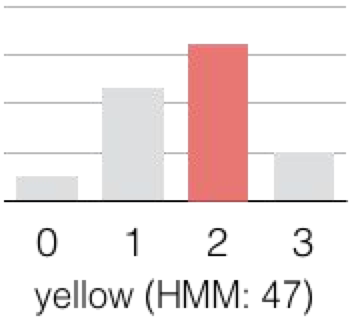
Experiment 2: Predicting verbs

Step 2: Aggregate verb predictions
store histogram-based predictions

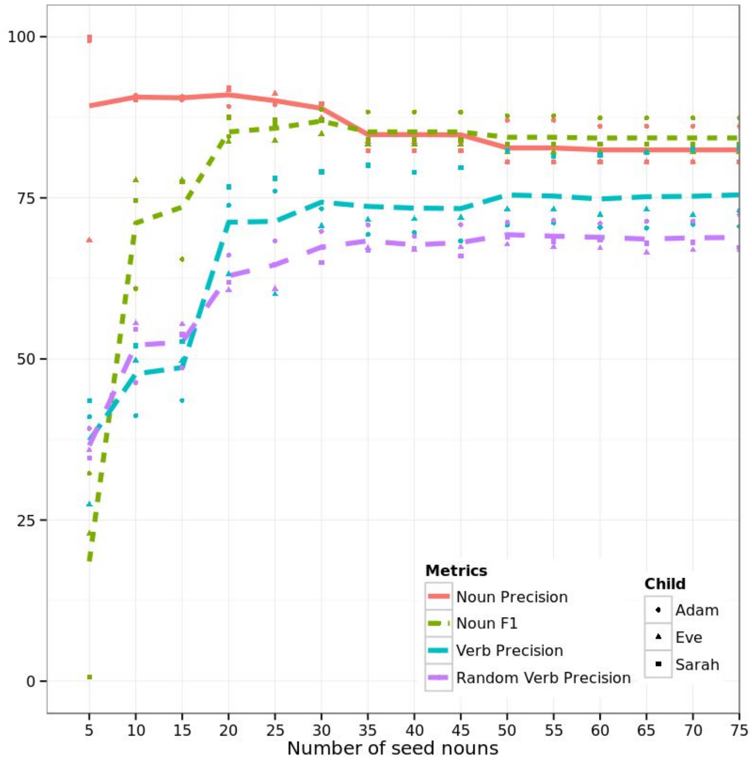
for each sentence:
    pick the HMM state most freq. appearing as verb

38: have, like, do, has, did, ought, remembers,…
…
47: little, other, new, good, big, next,…,yellow, nice, merry,…
…
59: **baby**, **horsie**, **dog**, bath,…, **flowers**, song, secret, **cat**,…
60: now, here, today, too, again, later, yesterday, first,…
61: it, **you**, **me**, her, nana, them,…, **daddy**, Mary,…



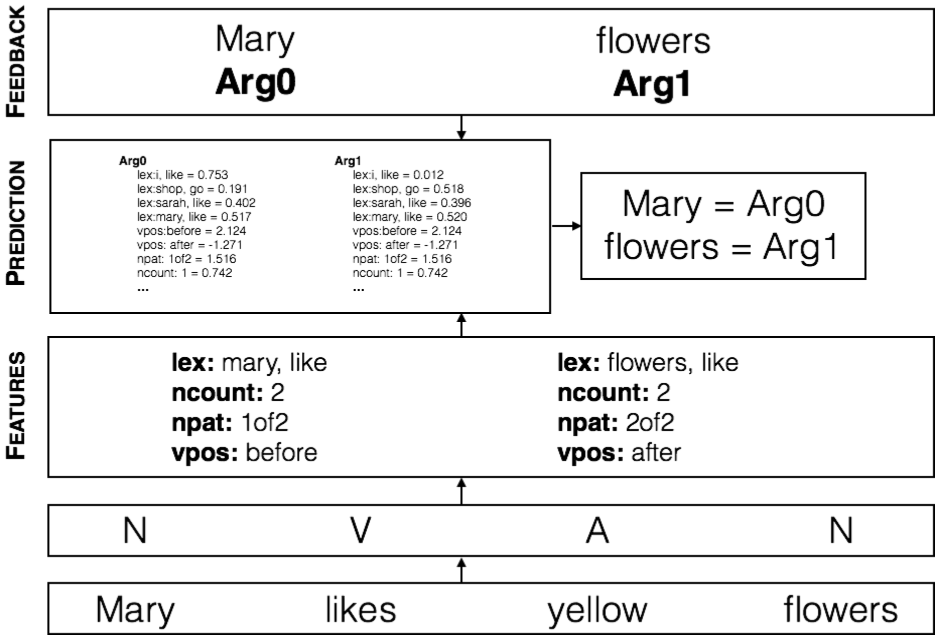


Results of verb/noun heuristics

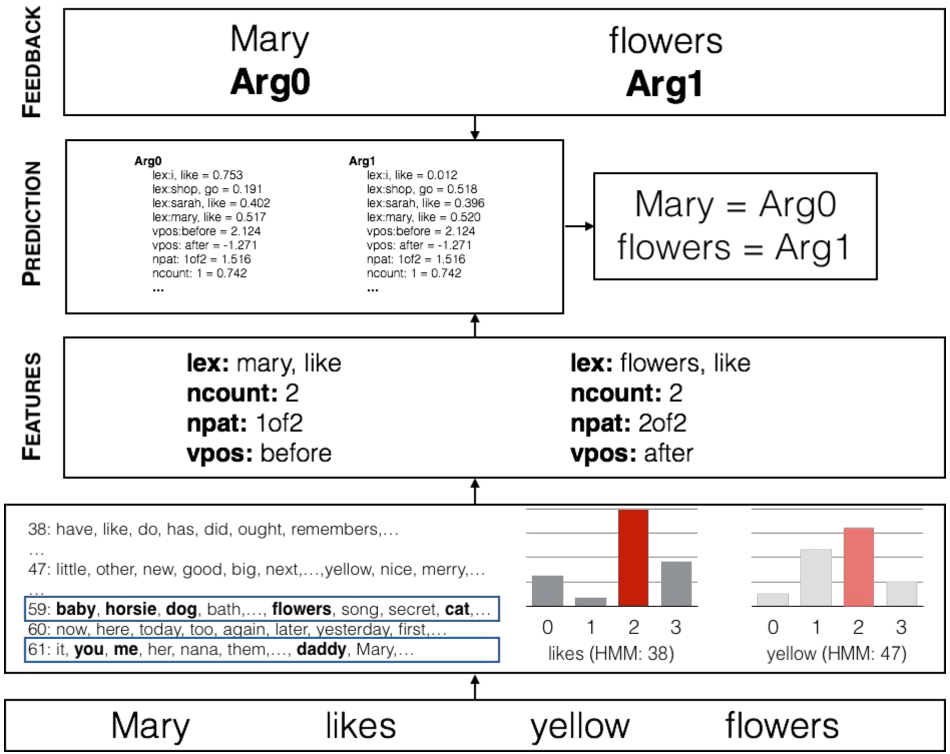


- Noun discovery is very accurate (84.3%)
  - ~30 nouns needed
- Verb accuracy is 6.6% > guessing (75.5%)
  - Using only argument counting
- Verb recall is low (~36%)
  - Multiple-verb sentences

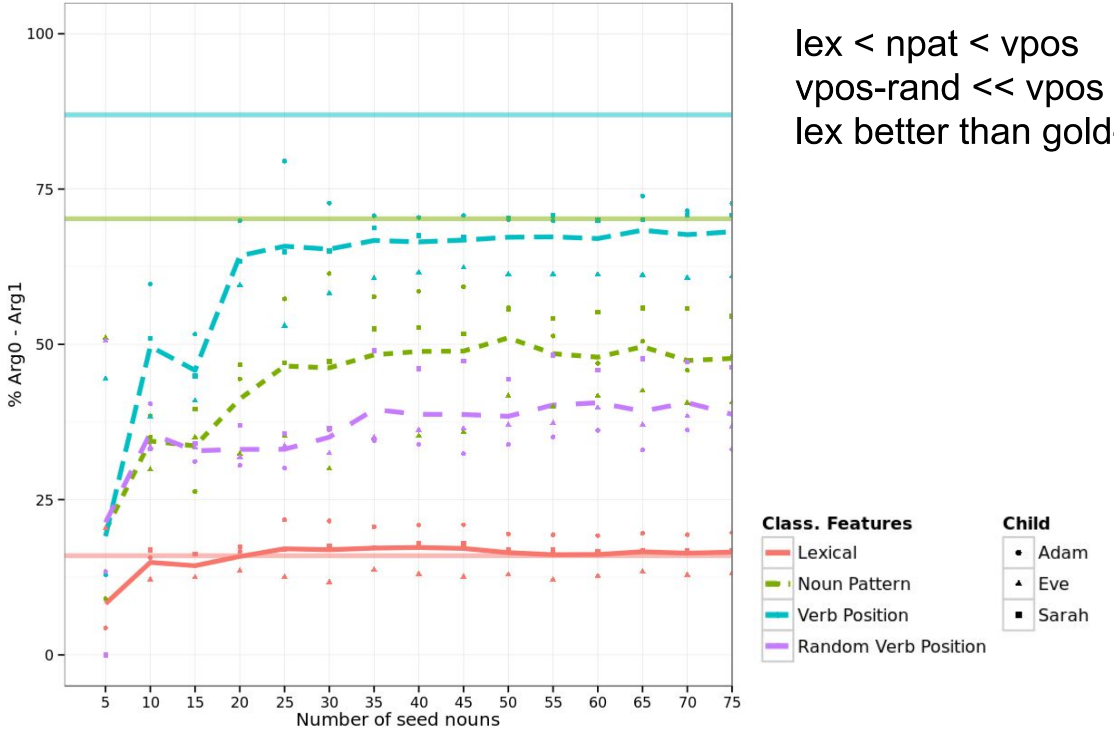
Experiment 2: SRL predictions



Experiment 2: SRL predictions

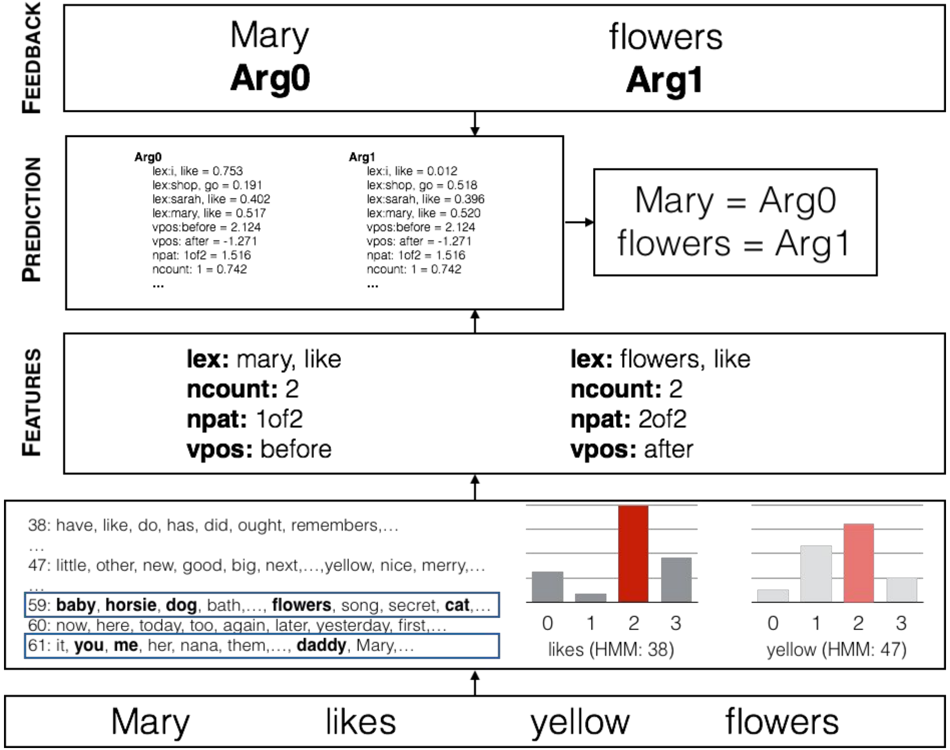


Results on novel-verb sentences (transitive)

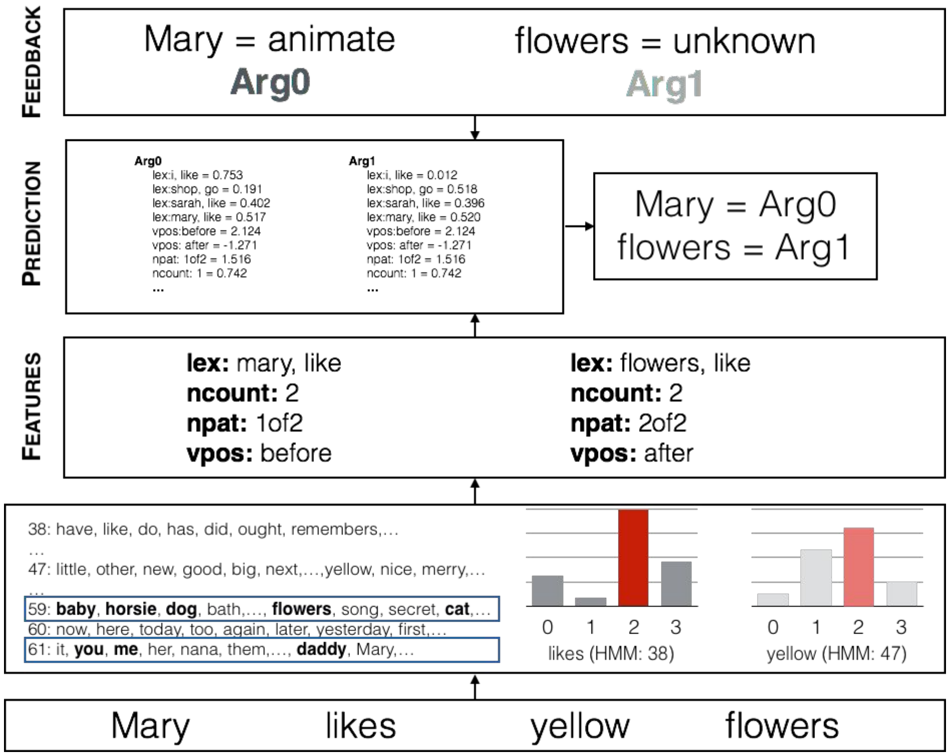


lex < npat < vpos
vpos-rand << vpos (~30%)
lex better than gold-PoS

Experiment 3: Reducing top-down supervision



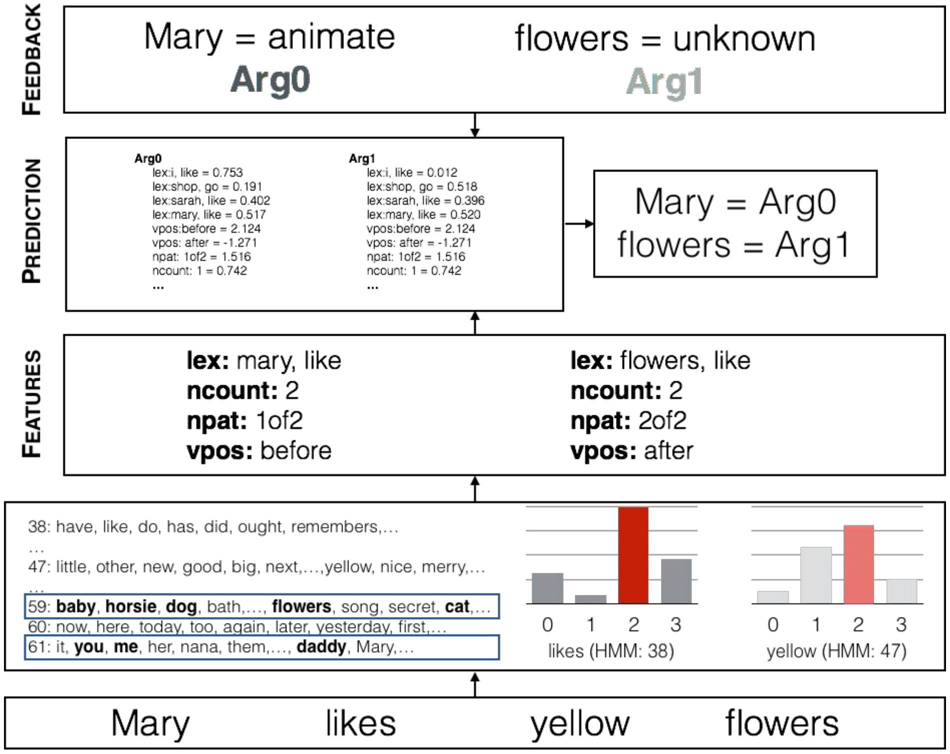
Experiment 3: Reducing top-down supervision



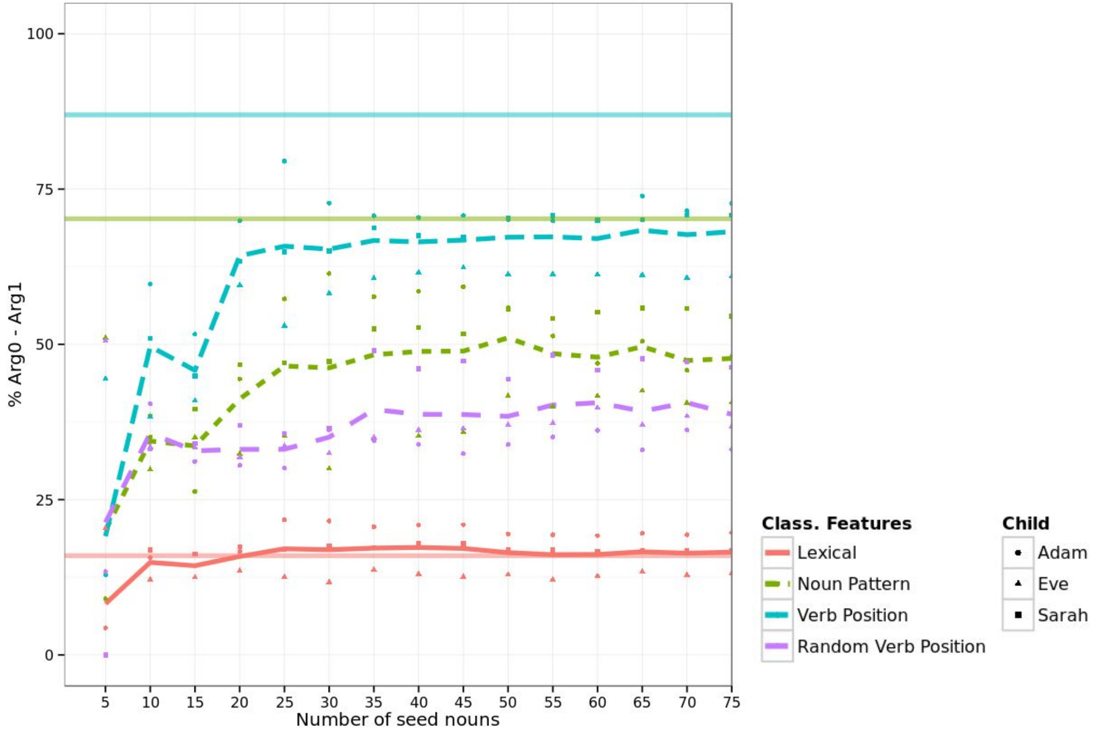
Experiment 3: Reducing top-down supervision

Input: BabySRL learner L
Data: CDS corpus C, animate nouns N_a, inanimate nouns N_i
foreach *sentence* $s \in C$ **do**

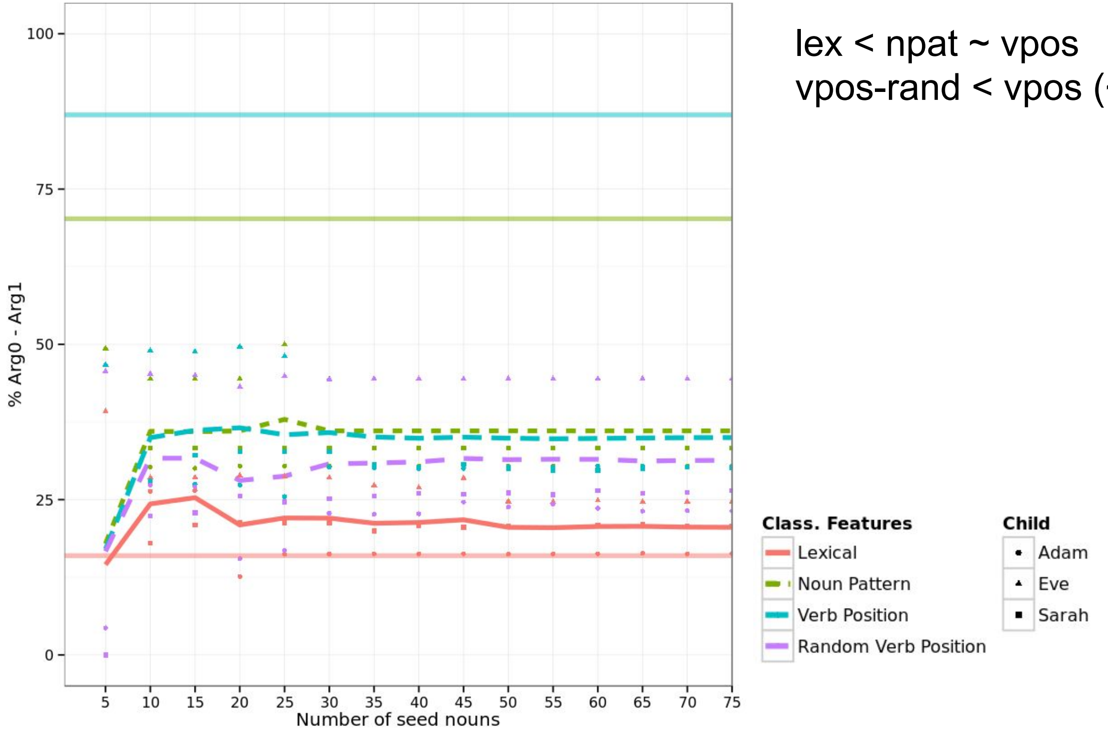
```
    // Mark animate nouns as agents
    if ∃s_i ∈ N_a then
        SRL(s_i) = Arg0;
        // Mark all other nouns as patients
        SRL(s_j)=Arg1 ∀s_j ∈ s, j ≠ i;
    else
        // Mark inanimates as patients
        if ∃s_i ∈ N_i then
            SRL(s_i) = Arg1;
        else
            // Sentence has no animacy information
            skip s;
        end
    end
    // If there are multiple animates
    if |SRL(s_i) = Arg0| > 1 then
        // Let the learner decide
        SRL(argmax_{s_i} L(s_i = Arg0)) = Arg0;
        // Mark all other nouns as agents
        SRL(s_j)=Arg1 ∀s_j ∈ s, j ≠ i;
    end
end
```



Results on novel-verb sentences (transitive)



Results on novel-verb sentences (transitive)



lex < npat ~ vpos
vpos-rand < vpos (~4%)

Baby SRL summary

- Modeling early language acquisition
  - Testbed for psycholinguistic theories
  - Replication of experimental results
- Structure-mapping for syntactic bootstrapping
  - Identifying verbs from noun structure
  - Predicting semantic roles using low-level syntactic features

Baby SRL summary

- Modeling early language acquisition
  - Testbed for psycholinguistic theories
  - Replica
- Structure-
  - Identify
  - Predict

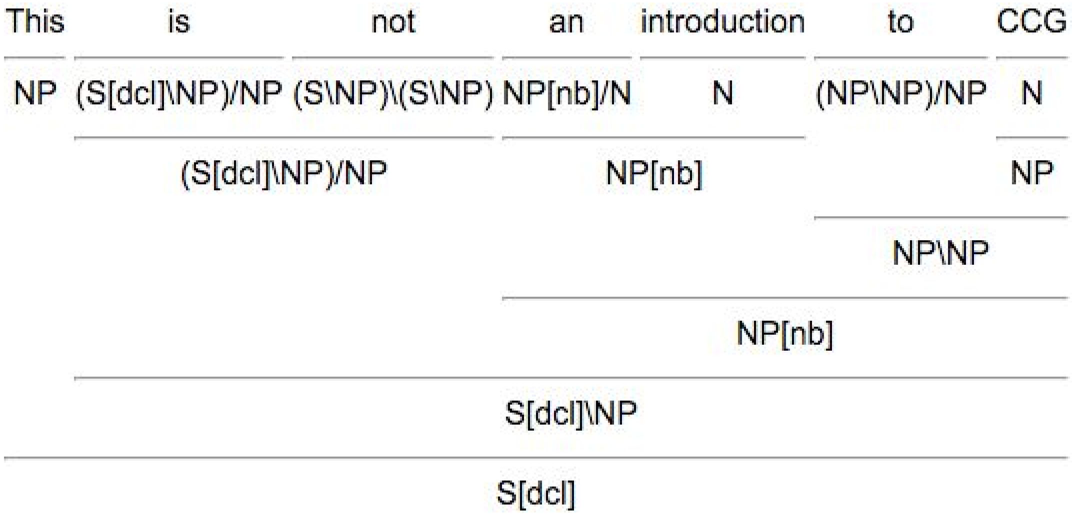
Getting closer to E-SRL

- Phrases instead of words
- Generalised predicate-argument structures
- Direct access to semantics

Bridging the gap: Inducing CCG structures
(Joint work with Yonatan Bisk and Julia Hockenmaier)

Combinatory Categorial Grammar [Steedman, 2000]

CCGbank [Hockenmaier & Steedman, 2007]



Bridging the gap: Inducing CCG structures
(Joint work with Yonatan Bisk and Julia Hockenmaier)

Combinato

CCGbank

- Can we induce a CCG from raw text?
- Bisk & Hockenmaier (2013)
  - Inducing CCG from raw text + PoS
- Can we replace PoS with induced clusters?
  - Bisk, **Christodoulopoulos**, & Hockenmaier, 2015

NP[nb]

S[dcl]\NP

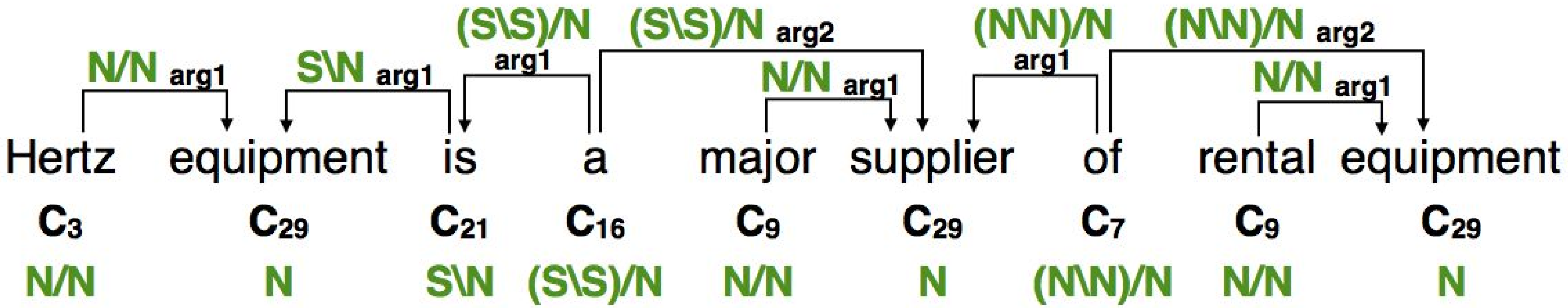
S[dcl]

Unsupervised CCG induction

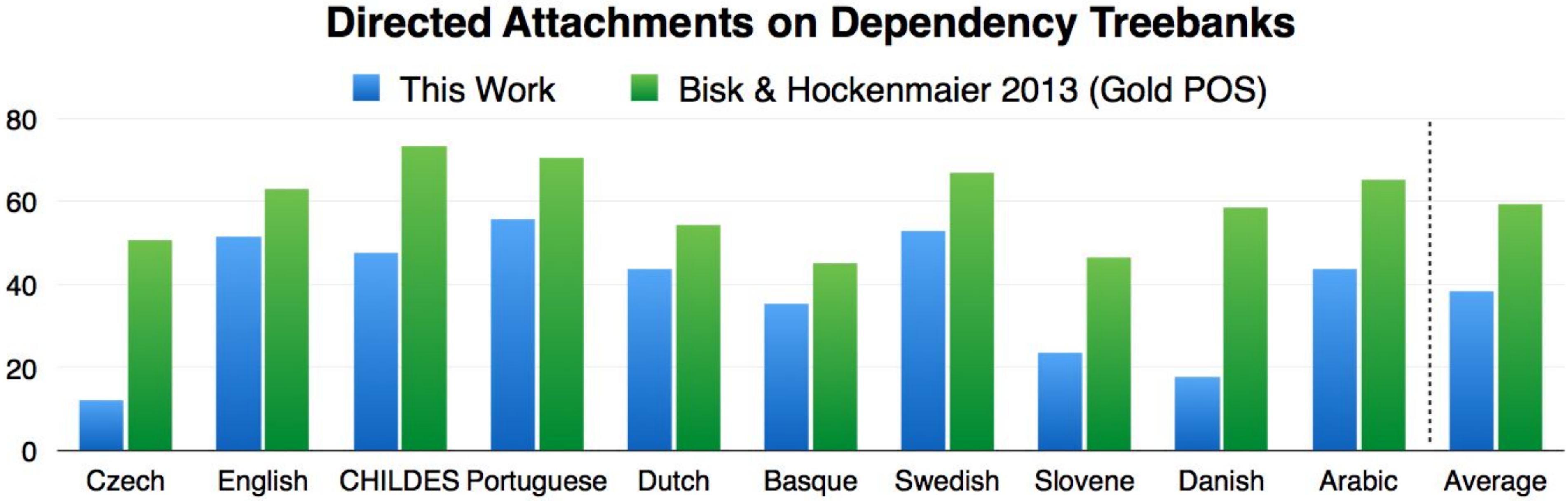
	Hertz	*equipment*	*is*	*a*	*major*	*supplier*	*of*	*rental*	*equipment*
Labels	**the, its, their, his,** these, our, Robert, my, your, every, His, Hurricane, Sir, Their, Freddie, Dean, Du, Tom, Jim, Remic, Roger, Gary, Ronald, Kenneth, Alex, Bruce, Litigation, Jay, Alfred, Ad, CS, Andrew, negotiable, Thrift, Patrick, Allied, Speaker, ...	**shares, sales, business,** companies, prices, investors, them, people, bonds, stocks, earnings, officials, income, rates, markets, analysis, products, funds, operations, growth, banks, issues, costs, concern, traders, him, assets, loans, firms, results, here, ...	**'s, is, was, are, has, were, had,** rose, fell, 're, ended, expects, whose, 've, remains, gained, owns, includes, became, jumped, holds, takes, provides, climbed, grew, gets, operates, sells, tumbled, seeks, becomes, begins, eased, allowed, helps,	**the, its, their, his,** these, our, Robert, my, your, every, His, Hurricane, Sir, Their, Freddie, Dean, Du, Tom, Jim, Remic, Roger, Gary, Ronald, Kenneth, Alex, Bruce, Litigation, Jay, Alfred, Ad, CS, Andrew, negotiable, Thrift, Patrick, Allied, Speaker, ...	**shares, sales, business,** companies, prices, investors, them, people, bonds, stocks, earnings, officials, income, rates, markets, analysis, products, funds, operations, growth, banks, issues, costs, concern, traders, him, assets, loans, firms, results, here, ...	**shares, sales, business,** companies, prices, investors, them, people, bonds, stocks, earnings, officials, income, rates, markets, analysis, products, funds, operations, growth, banks, issues, costs, concern, traders, him, assets, loans, firms, results, here, ...	**the, its, their, his,** these, our, Robert, my, your, every, His, Hurricane, Sir, Their, Freddie, Dean, Du, Tom, Jim, Remic, Roger, Gary, Ronald, Kenneth, Alex, Bruce, Litigation, Jay, Alfred, Ad, CS, Andrew, negotiable, Thrift, Patrick, Allied, Speaker, ...	**shares, sales, business,** companies, prices, investors, them, people, bonds, stocks, earnings, officials, income, rates, markets, analysis, products, funds, operations, growth, banks, issues, costs, concern, traders, him, assets, loans, firms, results, here, ...	**shares, sales, business,** companies, prices, investors, them, people, bonds, stocks, earnings, officials, income, rates, markets, analysis, products, funds, operations, growth, banks, issues, costs, concern, traders, him, assets, loans, firms, results, here, ...
R0		**N**	S		N	**N**		N	**N**
R1	**N/N**	S/S	**S\N** N\N	S\S N/N	**N/N**	N\N	N\N N/N	**N/N**	N\N
R2	(S/S)/(S/S)	(N/N)\(N/N) (N\N)/(N\N) (S/S)/(S\N)	(S/S)\(S/S) (S\S)/(S\S) (N/N)/(N/N) (S\N)\N (N\N)\N	(N/N)/(N/N) (N\N)\(N\N) **(S\S)/N** (S\S)\S (N/N)/N (N/N)\S (N/N)\(S\N) (S\S)\(S\N)	(N\N)/(N\N) (N/N)/N (S\S)\(S\S) (N/N)\(N/N)	(N/N)\(N/N) (N\N)\N (N/N)/(N/N) (N\N)/(N\N)	(N\N)\(N\N) (N/N)/(N/N) **(N\N)/N** (N/N)\N (N/N)/N (N\N)\N	(N\N)/(N\N) (N\N)\(N\N) (N/N)\(N/N) (N/N)/N	(N/N)\(N/N) (N\N)\N

Unsupervised CCG induction

Final round



Unsupervised CCG induction: Experiments



Unsupervised CCG induction: Summary

- Structured syntactic categories
  - Predicate-Argument structure directly from categories
- Transparent to semantics
  - Inverse of Kwiatkowski et al. (2012)

Overall Summary

- **Extended SRL**
  - Modular semantic structures with joint inference
  - Improved performance over independently-trained models
  - Constraint-based theories of language processing [McDonald et al., 1994; Ferreira et al., 2002]
- **Baby SRL**
  - Identifying verbs using only a few seed nouns
  - Low-level syntactic features can guide semantic learning
  - Empirical evidence for structure-mapping account of syntactic bootstrapping
- **Unsupervised CCG induction**
  - Structured lexical-syntactic-semantic representation from raw text

The road ahead

- **Extended SRL**
  - Mine for constraints
  - Entailment via latent structure alignment [Sammons et al., 2009]
- **Baby SRL**
  - Add internal structure (CCG)
  - More predicates, constructions (pro-drop, filler gap)
- **Unsupervised CCG induction**
  - Use weak semantic supervision
  - Add constraints from E-SRL

DARPA grant
FA8750-13-2-0008

NSF grant BCS-
1348522

NIH grant R01-
HD054448

Thank you!

Dan Roth, Cindy Fisher, Mark Sammons, Vivek Srikumar, Parisa Kordjamshidi,
Shyam Upadhyay, Daniel Khashabi, Stephen Mayhew, Mark Steedman,
Yonatan Bisk, Julia Hockenmaier, Catriona Silvey

Source code:
https://gitlab-beta.engr.illinois.edu/cogcomp/illinois-esrl
https://gitlab-beta.engr.illinois.edu/babysrl-group/babysrl