A General-Purpose Algorithm for Constrained Sequential Inference

Daniel Deutsch,* Shyam Upadhyay,* and Dan Roth

*equal contribution

Co-Authors

Shyam Upadhyay

Dan Roth

Structured Prediction

- Structured prediction is everywhere
  - Parsing, tagging, generation
- Inference is hard: exponentially large search space
- Not all outputs are "valid"

Valid Output Structures: Parsing

Input: John kissed Mary

Gold Parse: `(S (NP XX ) (VP XX (NP XX ) ) )`

Valid Output Structures: Parsing

Input: John kissed Mary

Gold Parse: `(S (NP XX ) (VP XX (NP XX ) ) )`

Invalid Parse: `(S (NP ) (VP XX XX (NP XX ) ) )` ✗

Empty phrase

Valid Output Structures: Parsing

Input: John kissed Mary

Gold Parse: `(S (NP XX ) (VP XX (NP XX ) ) )`

Invalid Parse: `(S (NP ) (VP XX XX (NP XX ) ) )` ✗

Empty phrase

Invalid Parse: `(S (VP XX (NP XX ) ) )` ✗

Incorrect number of pre-terminals

Valid Output Structures: Parsing

Input: John kissed Mary

Gold Parse: `(S (NP XX ) (VP XX (NP XX ) ) )`

Invalid Parse: `(S (NP ) (VP XX XX (NP XX ) ) )` ✗

Empty phrase

Invalid Parse: `(S (VP XX (NP XX ) ) )` ✗

Incorrect number of pre-terminals

Invalid Parse: `(S (NP XX ) (VP XX (NP XX ) ) ) )` ✗

Unbalanced parentheses

Valid Output Structures: Semantic Role Labelling

A0 ≈ agent
A1 ≈ patient
…

Input:	Alice	Smith	gave	a	flower	to	Bob	Jones
Gold Tags:	A0	A0	O	O	A1	O	A2	A2

Valid Output Structures: Semantic Role Labelling

Input:	Alice	Smith	gave	a	flower	to	Bob	Jones
Gold Tags:	A0	A0	O	O	A1	O	A2	A2
Invalid Tags:	**A0**	A0	O	O	**A0**	O	A2	A2 ✗

Duplicate A0

Valid Output Structures: Semantic Role Labelling

Input:	Alice	Smith	gave	a	flower	to	Bob	Jones
Gold Tags:	A0	A0	O	O	A1	O	A2	A2

Legal Args: A0, A1, A2 (from Propbank for "gave")

Invalid Tags:	**A0**	A0	O	O	**A0**	O	A2	A2	✗

Duplicate A0

Valid Output Structures: Semantic Role Labelling

Input:	Alice	Smith	gave	a	flower	to	Bob	Jones
Gold Tags:	A0	A0	O	O	A1	O	A2	A2

Legal Args: A0, A1, A2 (from Propbank for "gave")

Invalid Tags:	**A0**	A0	O	O	**A0**	O	A2	A2	✗

Duplicate A0

Invalid Tags:	**A3**	**A3**	O	O	A1	O	A2	A2	✗

Illegal argument A3

Valid Output Structures: Semantic Role Labelling

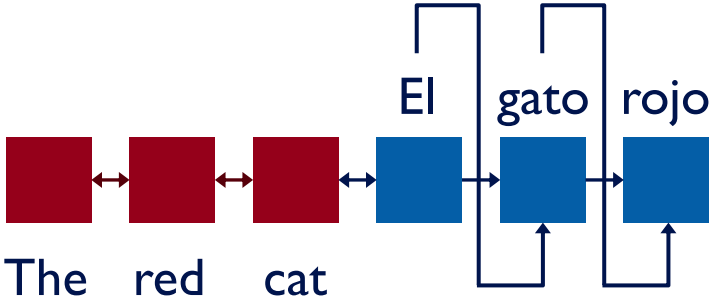
Input:	Alice	Smith	gave	a	flower	to	Bob	Jones	
Gold Tags:	A0	A0	O	O	A1	O	A2	A2	
Legal Args:	A0, A1, A2 (from Propbank for "gave")								
Spans:	[Alice	Smith]	gave	a	flower	to	[Bob	Jones]	
Invalid Tags:	**A0**	A0	O	O	**A0**	O	A2	A2	✗
	Duplicate A0								
Invalid Tags:	**A3**	**A3**	O	O	A1	O	A2	A2	✗
	Illegal argument A3								

Valid Output Structures: Semantic Role Labelling

Input:	Alice	Smith	gave	a	flower	to	Bob	Jones	
Gold Tags:	A0	A0	O	O	A1	O	A2	A2	
Legal Args:	A0, A1, A2 (from Propbank for "gave")								
Spans:	[Alice	Smith]	gave	a	flower	to	[Bob	Jones]	
Invalid Tags:	**A0**	A0	O	O	**A0**	O	A2	A2	✗
	Duplicate A0								
Invalid Tags:	**A3**	**A3**	O	O	A1	O	A2	A2	✗
	Illegal argument A3								
Invalid Tags:	**A0**	**O**	O	O	A1	O	A2	A2	✗
	Spans not respected								

Sequential Inference

- Many tasks have converged on the same solution
  - Assume output structure is a sequence
  - Beam search from left-to-right
  - Seq2Seq
- How are structural constraints enforced?



Enforcing Constraints

- No single way to enforce constraints
  - Not enforced at all (Lample et al., 2016; Choe and Charniak, 2016; Suhr et al., 2018)
  - Post-hoc (Andreas et al., 2013; Vinyals et al., 2015; Upadhyay et al., 2018)
  - Custom inference algorithm (Zhu et al., 2013)

This Work

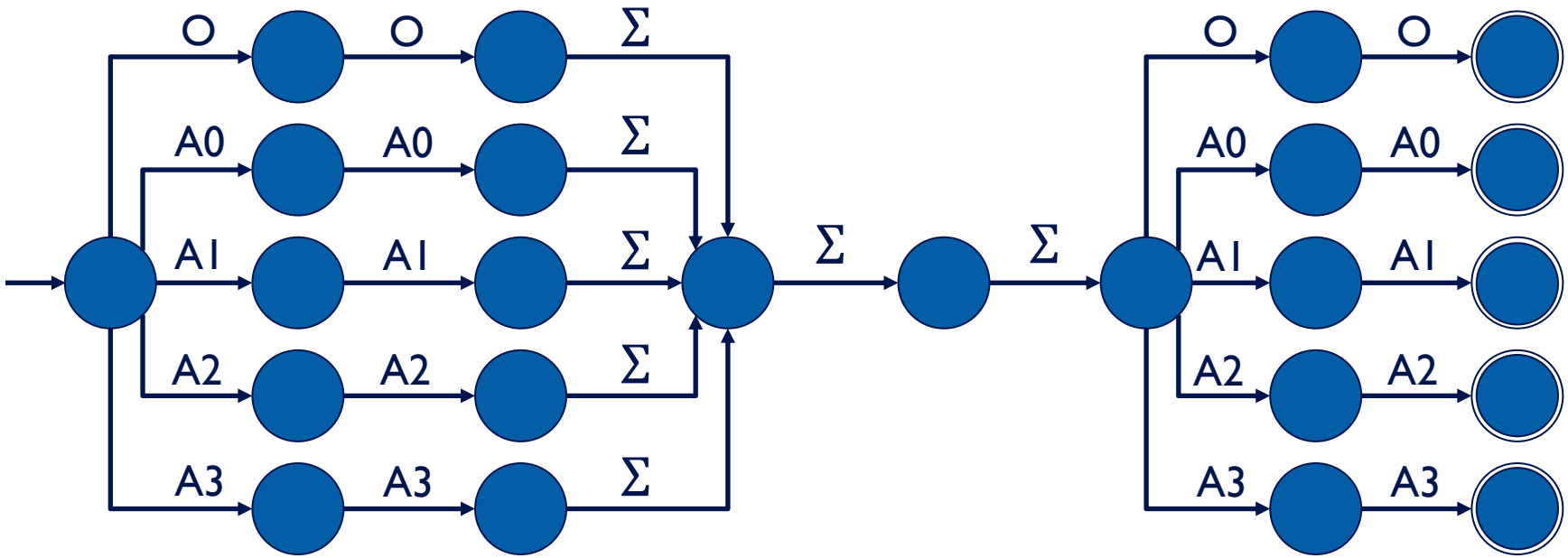
- Propose a generic algorithm for enforcing constraints in sequential inference
- Abstracts over many custom inference procedures in NLP
- Built on expressing constraints as automata
- Automata guides inference to valid outputs

Representing Constraints

- Represent a constraint with an automaton
  - Finite-state automata and push-down automata
- Automaton's language is all of the valid output structures that satisfy the constraint
- Great automaton libraries
  - OpenFST (Allauzen et al., 2007), Pynini (Gorman, 2016)

Representing Constraints as Automata

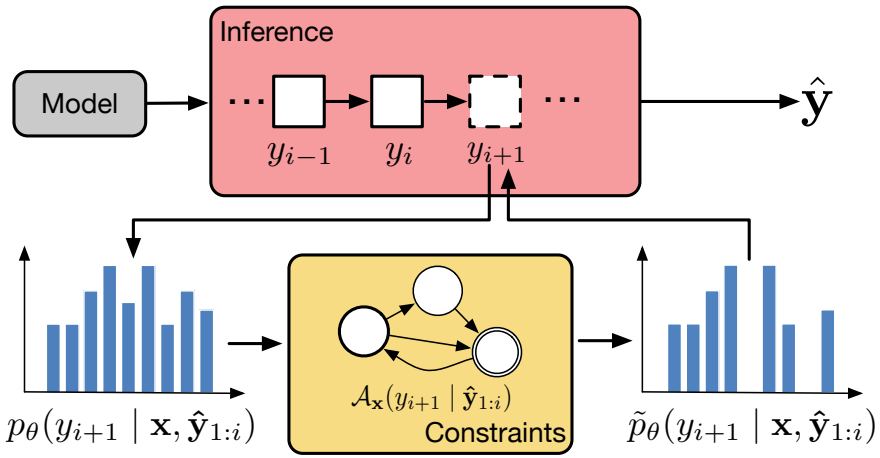
[Alice Smith] gave flowers to [Bob Jones]



Σ = Any tag

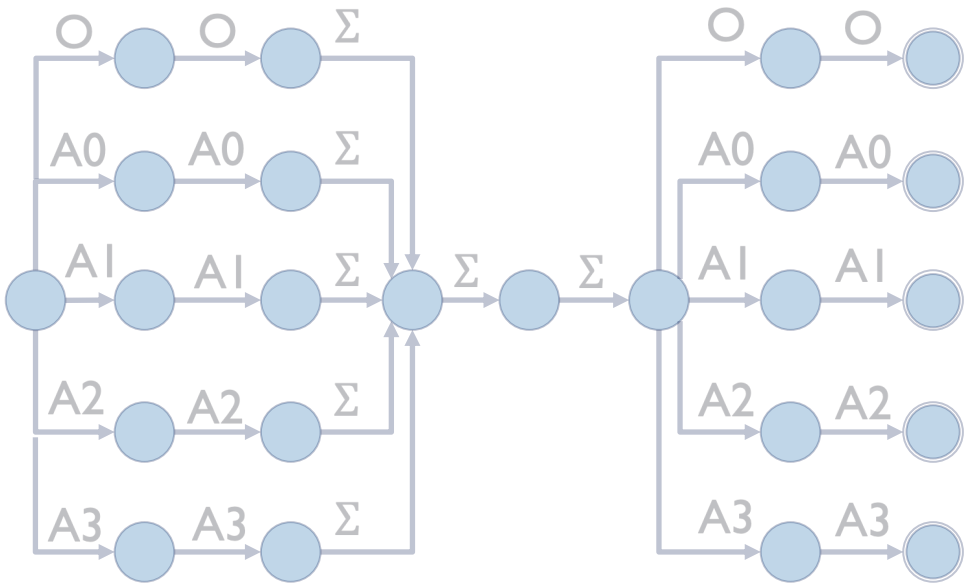
Constrained Inference

- Automaton will be traversed in lock-step with beam search, guiding inference to find a valid structure



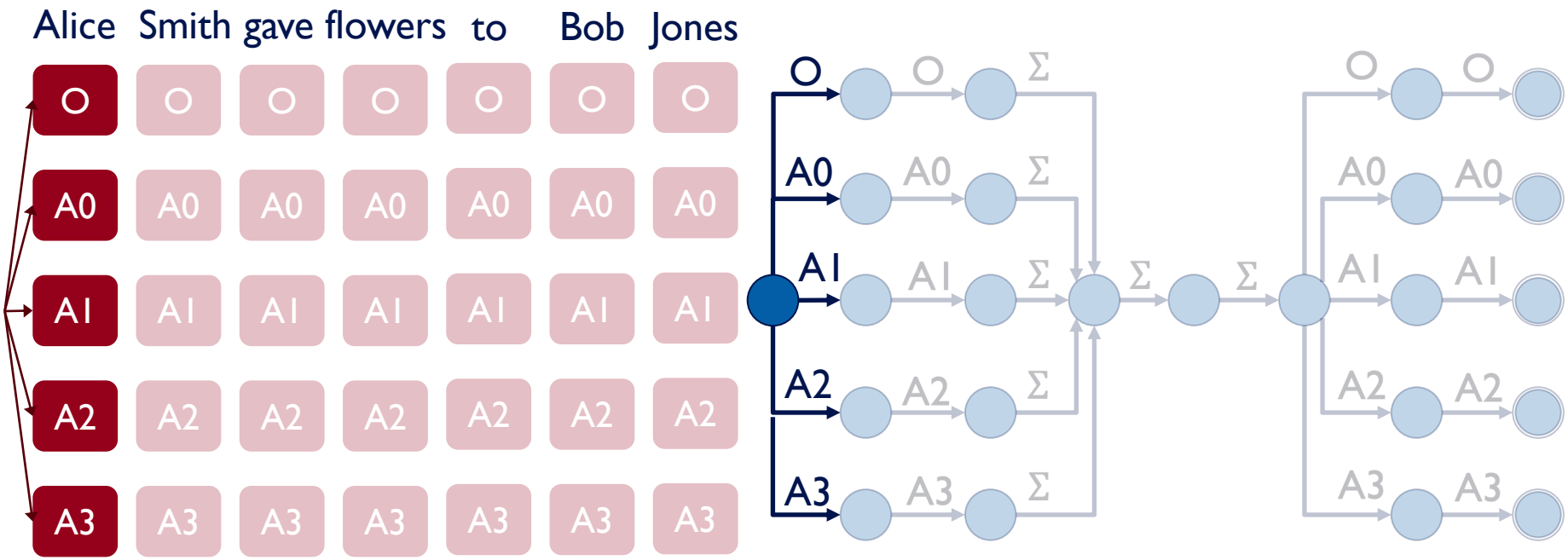
Constrained Inference

Alice	Smith	gave	flowers	to	Bob	Jones
O	O	O	O	O	O	O
A0	A0	A0	A0	A0	A0	A0
A1	A1	A1	A1	A1	A1	A1
A2	A2	A2	A2	A2	A2	A2
A3	A3	A3	A3	A3	A3	A3



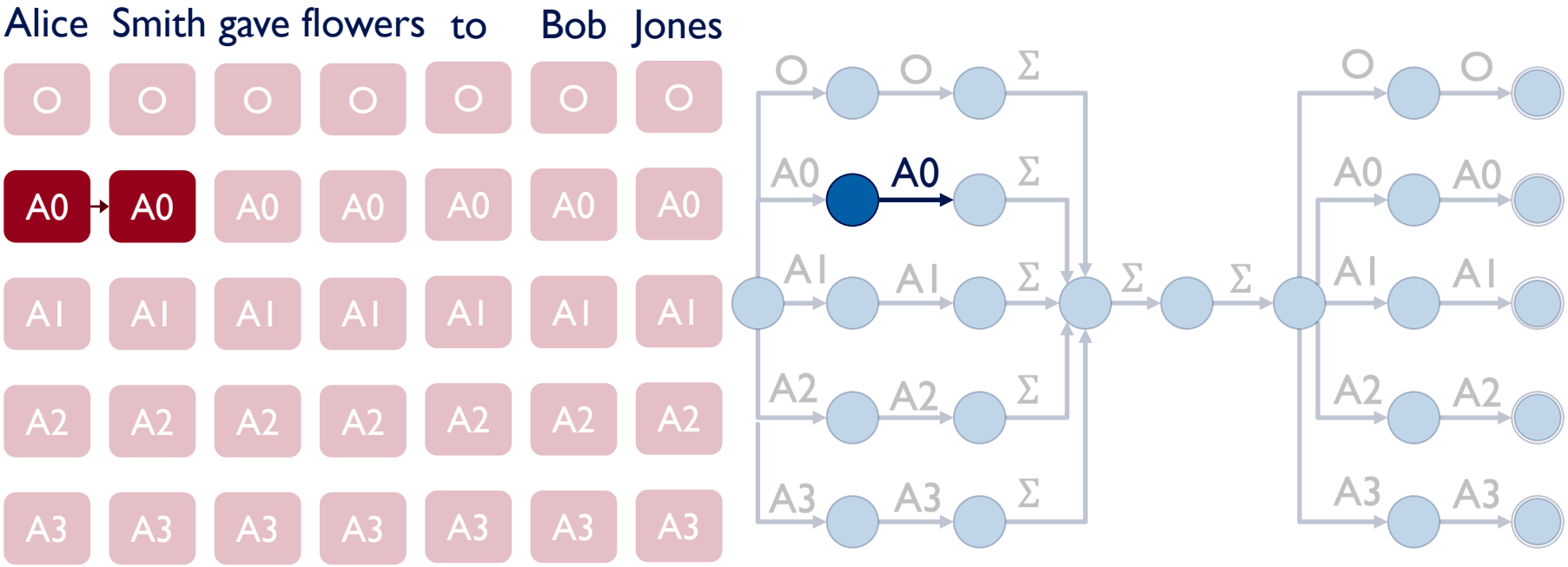
Σ = Any tag

Constrained Inference



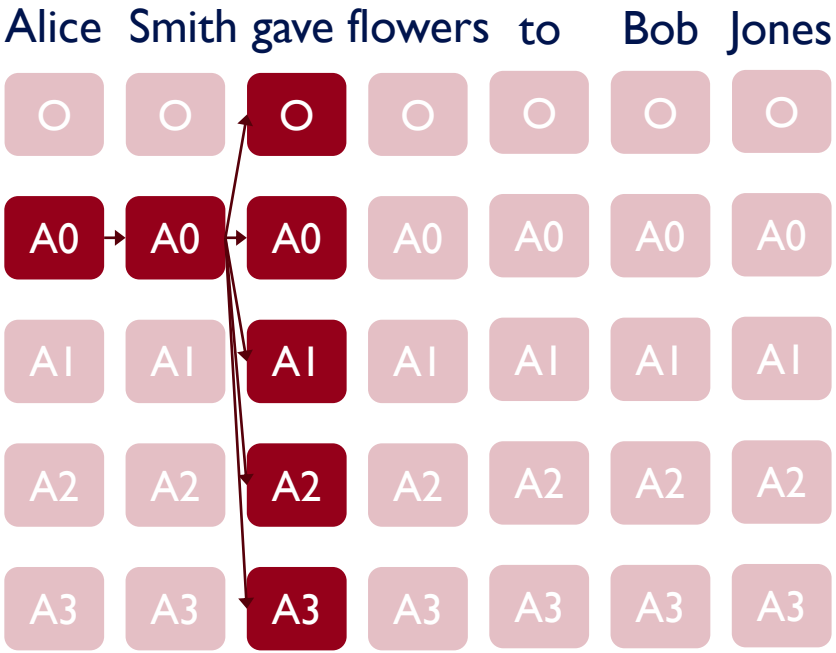
Σ = Any tag

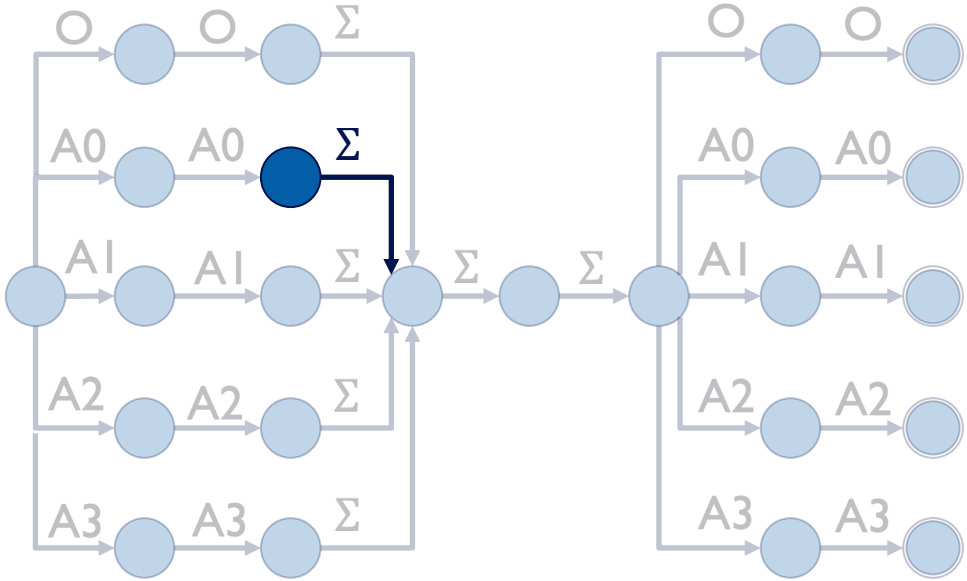
Constrained Inference



Σ = Any tag

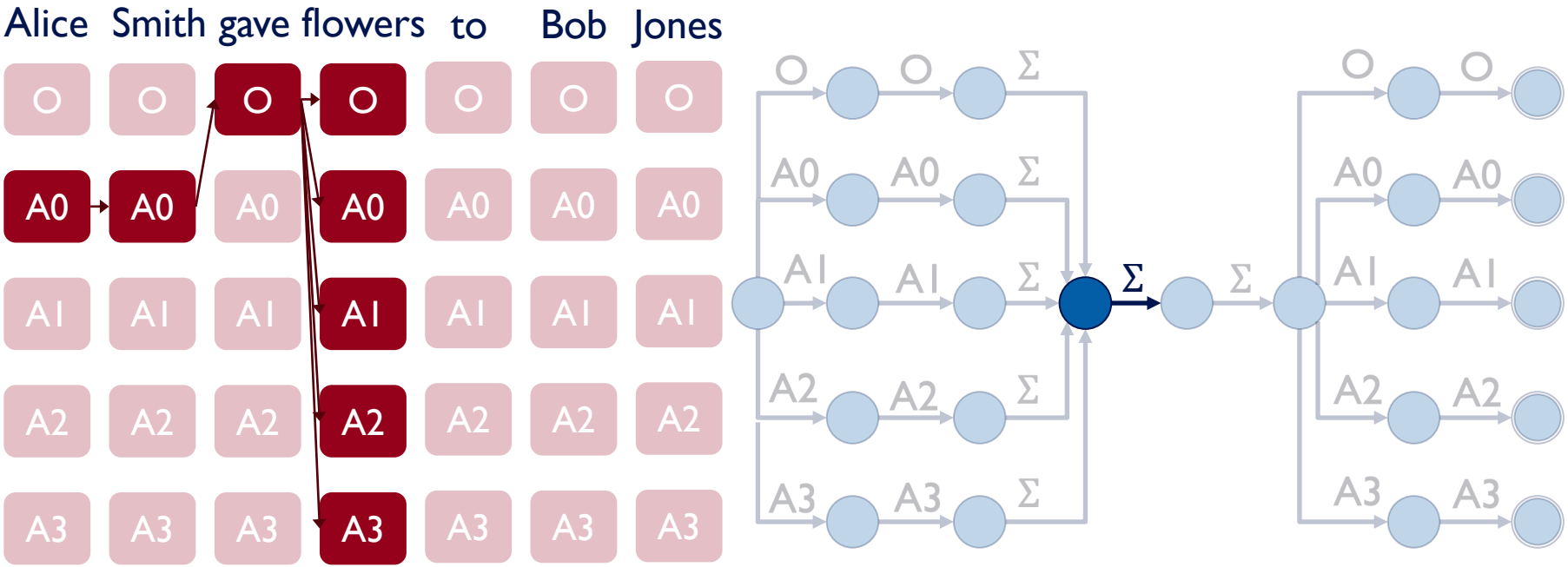
Constrained Inference





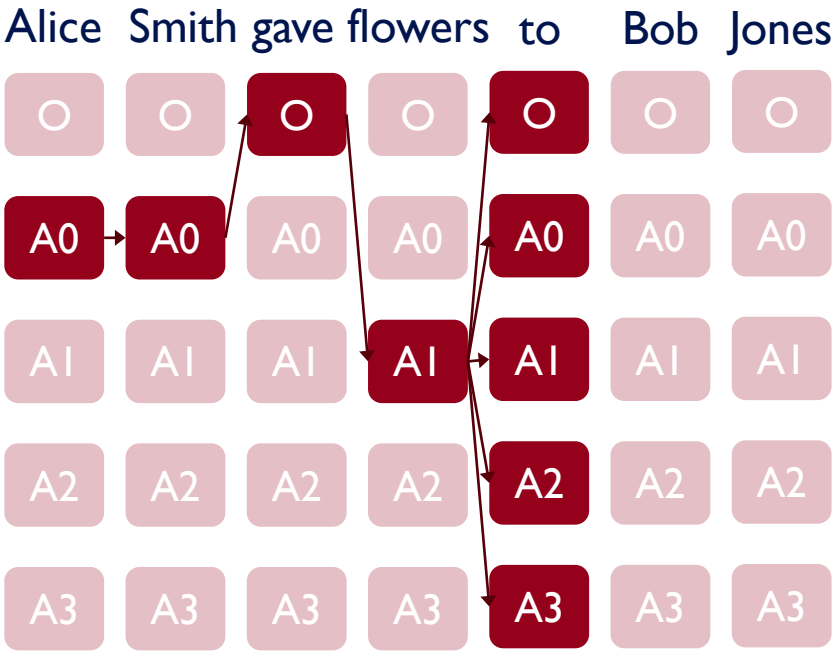
Σ = Any tag

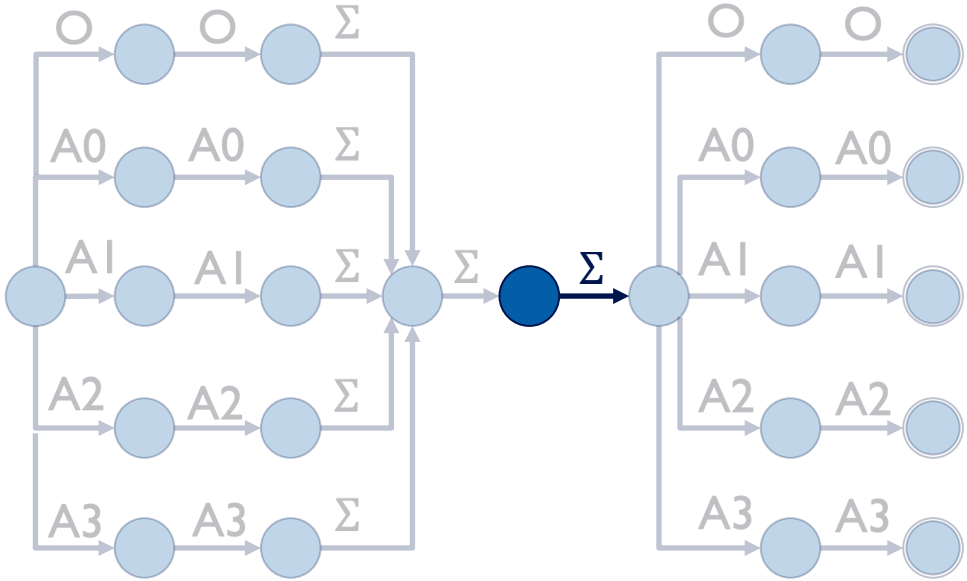
Constrained Inference



Σ = Any tag

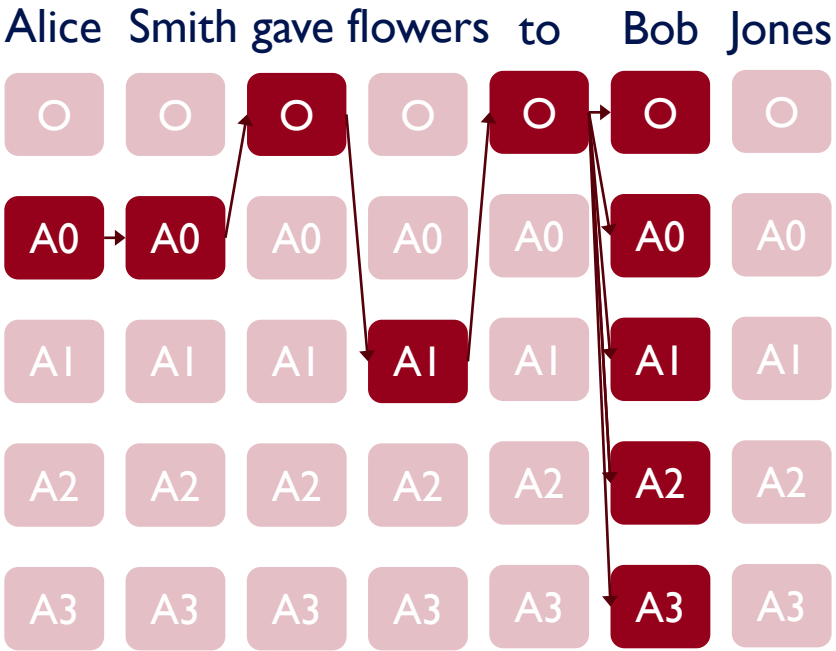
Constrained Inference

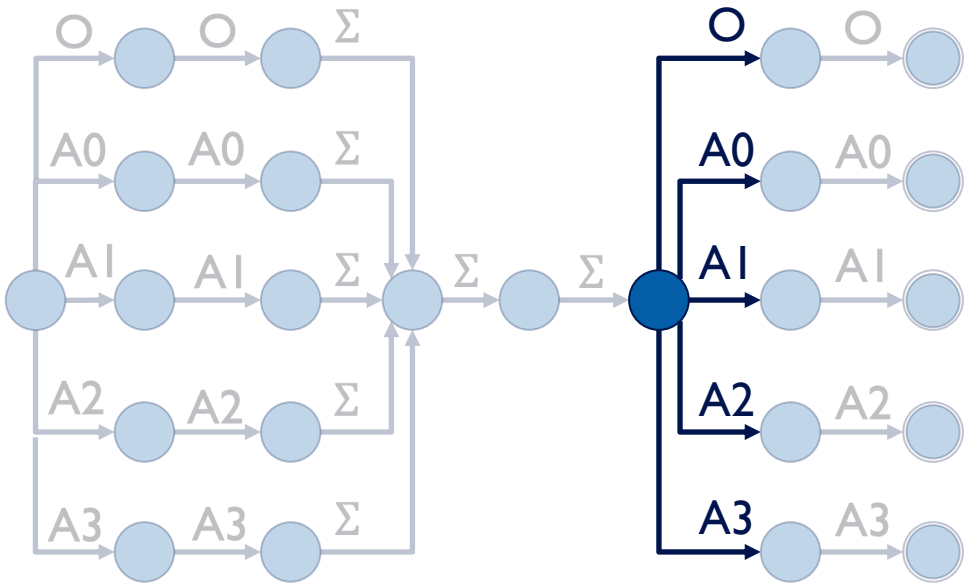




Σ = Any tag

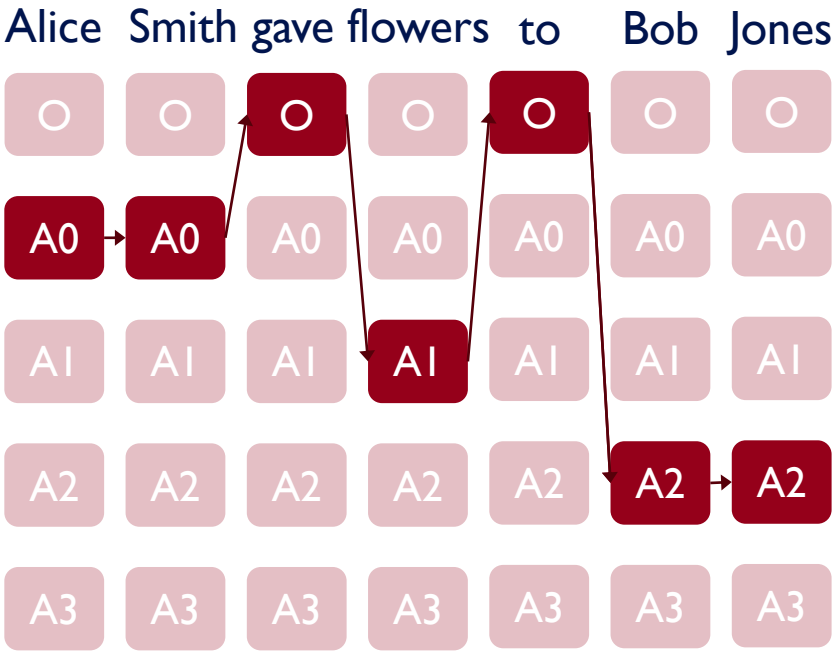
Constrained Inference

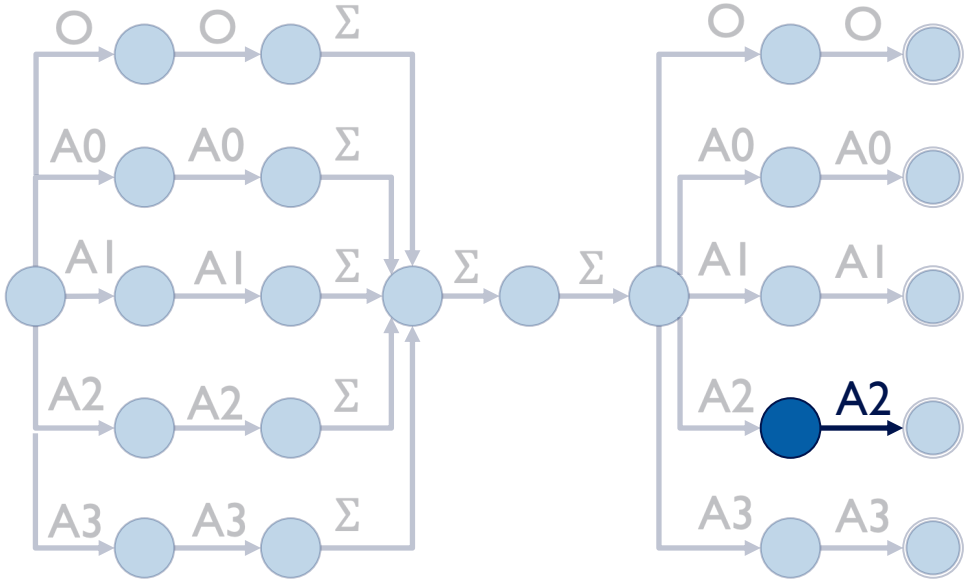




Σ = Any tag

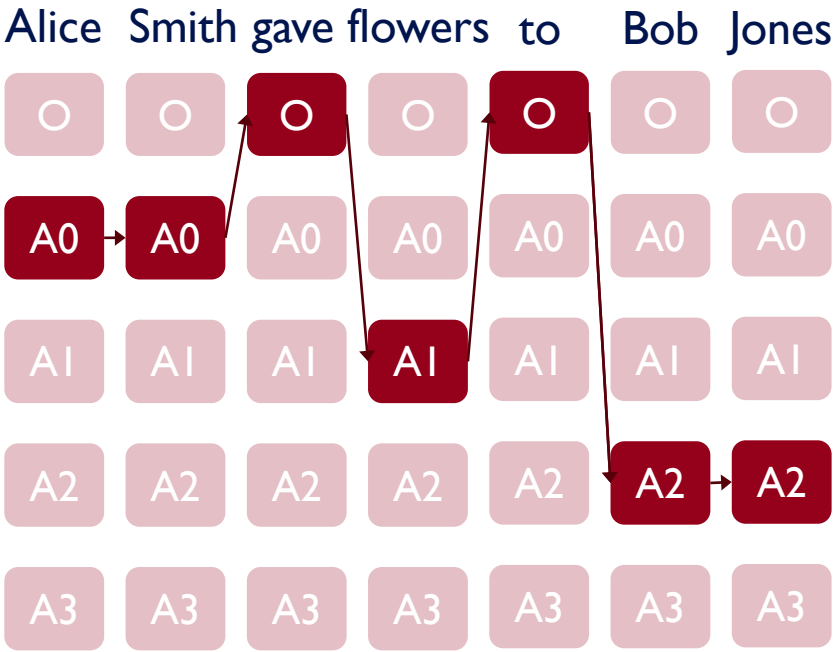
Constrained Inference

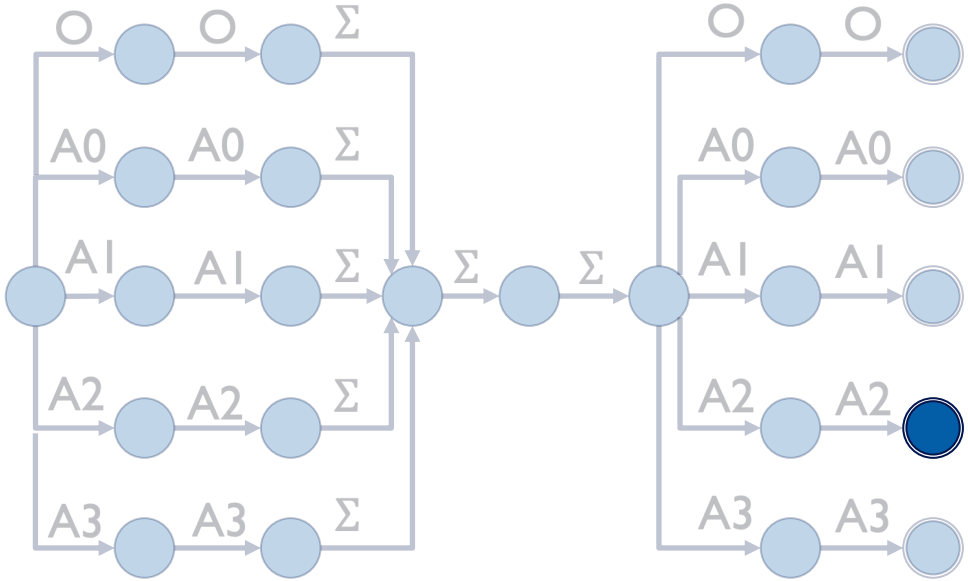




Σ = Any tag

Constrained Inference





Σ = Any tag

Handling Multiple Constraints

- What about multiple constraints?

Handling Multiple Constraints

- What about multiple constraints?
- Can't traverse jointly -- Need to intersect all of the automata

Handling Multiple Constraints

- What about multiple constraints?
- Can't traverse jointly -- Need to intersect all of the automata
- Intersecting automata can be expensive

Handling Multiple Constraints

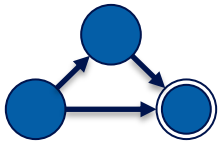
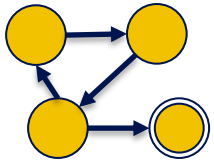
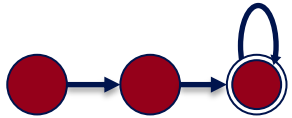
- What about multiple constraints?
- Can't traverse jointly -- Need to intersect all of the automata
- Intersecting automata can be expensive
- Idea: Intersect automata only as necessary (Tromble and Eisner, 2006)
- Maintain an *active set* of enforced constraints

Active Set Algorithm

Input

John kissed Mary

Active Set Algorithm

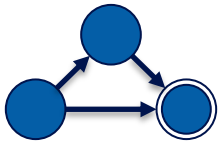
Input	Balanced Parentheses	Non-Empty Phrases	Correct Num. Pre-Terminals
John kissed Mary			

Active Set Algorithm

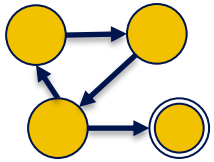
Input

John kissed Mary

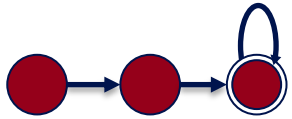
Balanced Parentheses



Non-Empty Phrases



Correct Num. Pre-Terminals

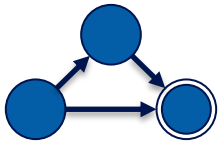
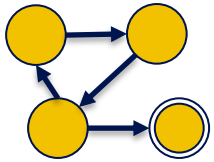
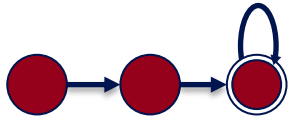


Constraint Automaton

Inference Output

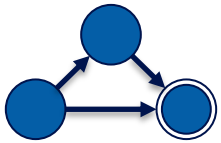
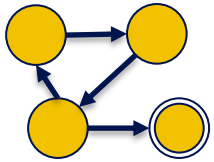
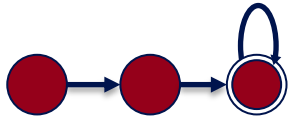
Violated Constraints

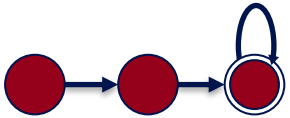
Active Set Algorithm

Input	Balanced Parentheses	Non-Empty Phrases	Correct Num. Pre-Terminals
John kissed Mary			

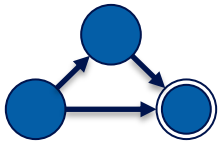
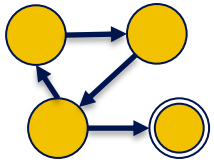
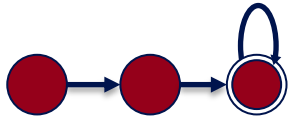
Constraint Automaton	Inference Output	Violated Constraints
n/a	(S (VP XX (NP XX))) ✗	Correct Num. Pre-Terminals

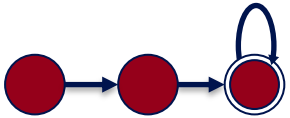
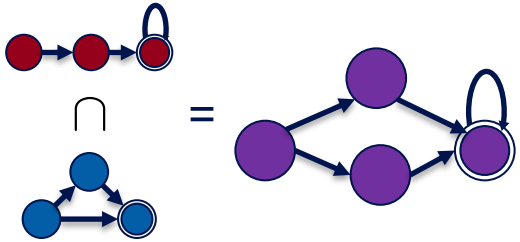
Active Set Algorithm

Input	Balanced Parentheses	Non-Empty Phrases	Correct Num. Pre-Terminals
John kissed Mary			

Constraint Automaton	Inference Output	Violated Constraints
n/a 	(S (VP XX (NP XX))) ✗	Correct Num. Pre-Terminals
	(S (NP XX) (VP XX (NP XX)))) ✗	Balanced Parentheses

Active Set Algorithm

Input	Balanced Parentheses	Non-Empty Phrases	Correct Num. Pre-Terminals
John kissed Mary			

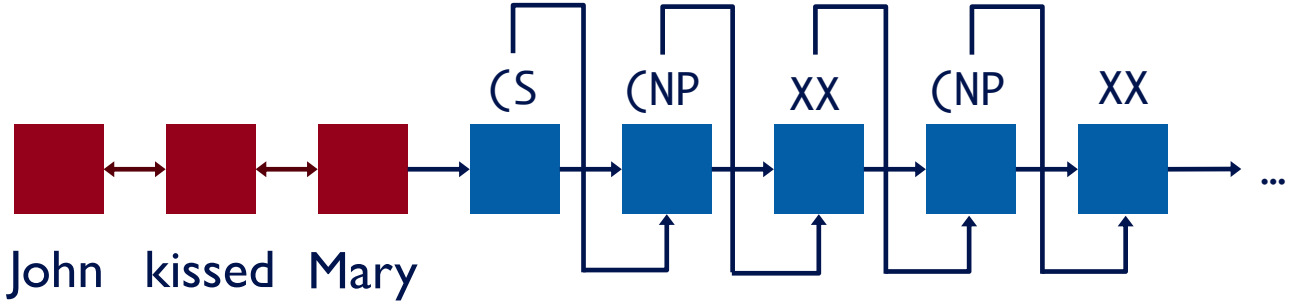
Constraint Automaton	Inference Output	Violated Constraints
n/a 	(S (VP XX (NP XX))) ✗	Correct Num. Pre-Terminals
	(S (NP XX) (VP XX (NP XX)))) ✗	Balanced Parentheses
	(S (NP XX) (VP XX (NP XX))) ✓	None

Generality of Algorithm

- Abstract over constraint-specific inference algorithms:
  - BIO tagging
  - Shift-reduce parsing
  - Require/disallow n-grams
  - Constraining verbalizations in text-to-speech (Zhang et al., 2019)

Constituency Parsing Experiments

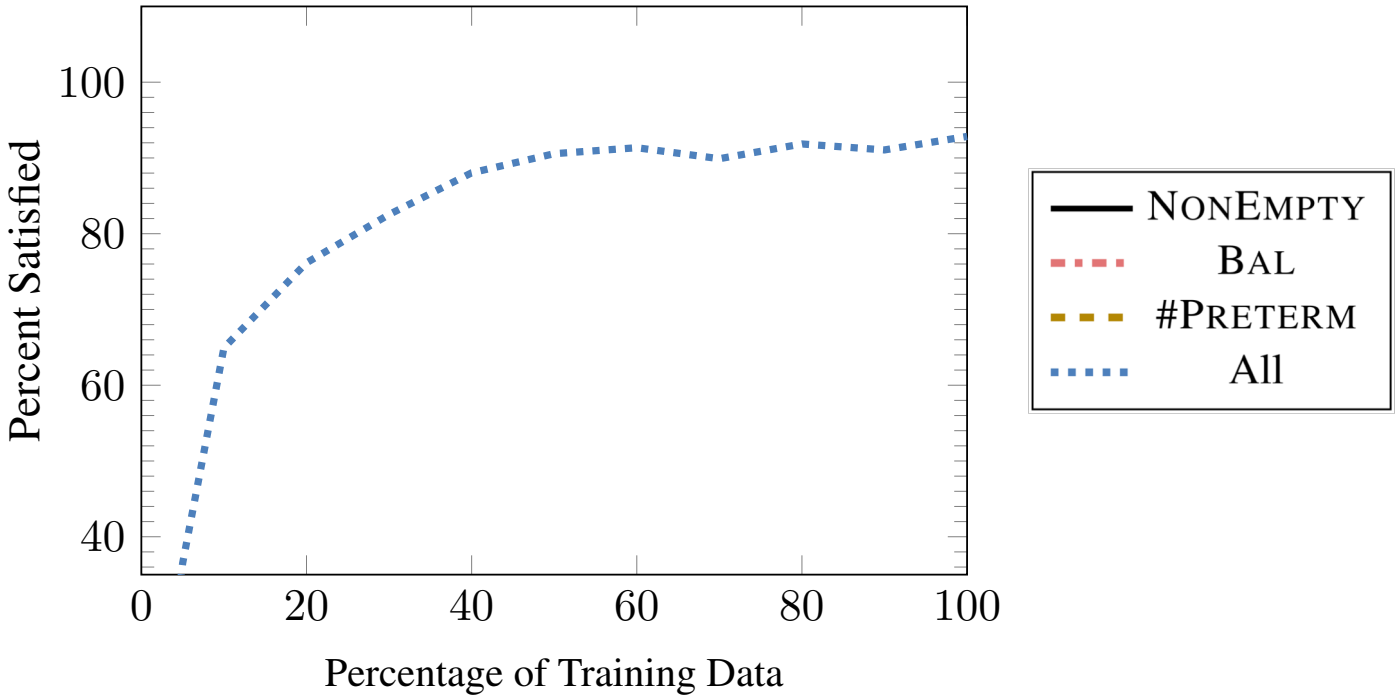
- Seq2Seq constituency parsing on the Penn Treebank



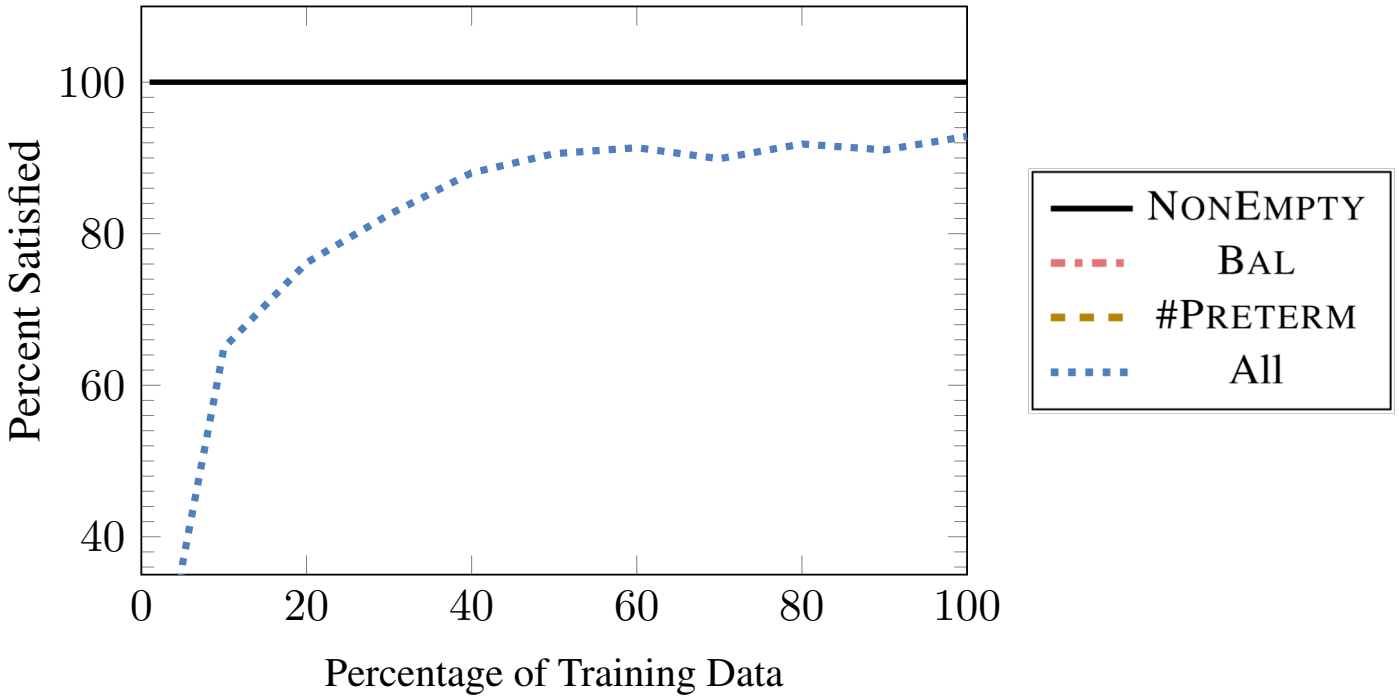
Necessity of Constraints

- Do we need constraints at all?
- Enforce three different constraints
  - Balanced parentheses
  - Correct number of pre-terminals
  - No empty phrases

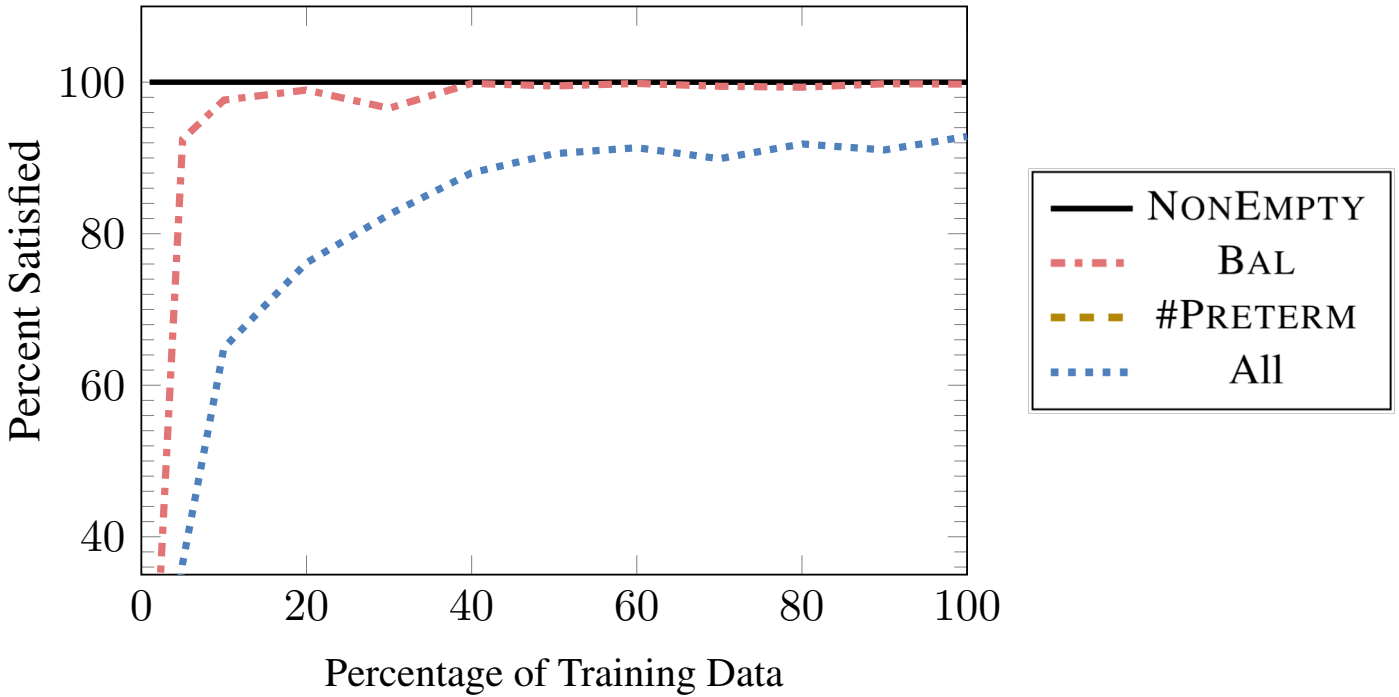
Necessity of Constraints



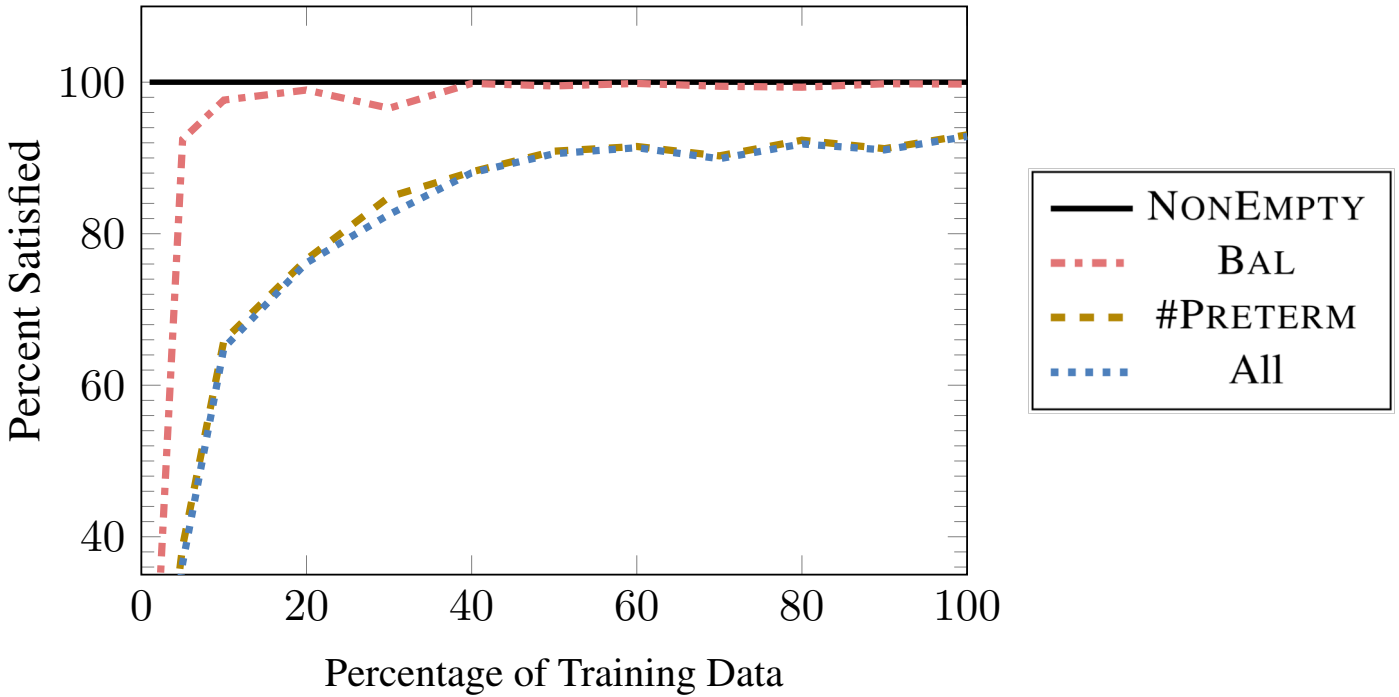
Necessity of Constraints



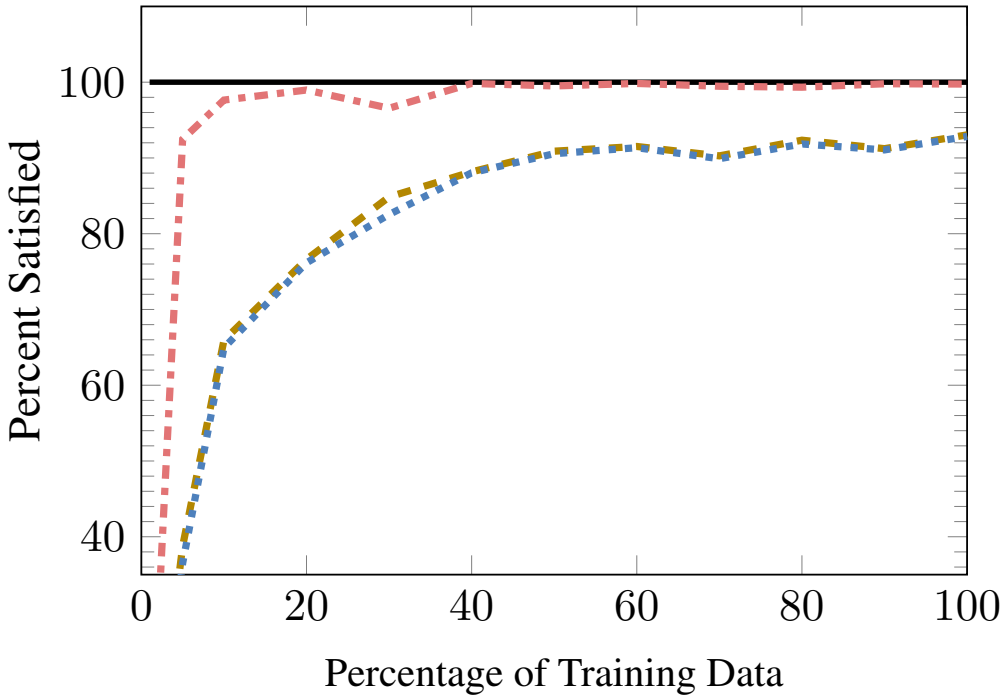
Necessity of Constraints



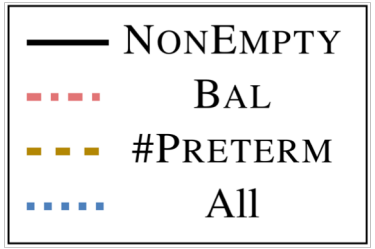
Necessity of Constraints



Necessity of Constraints



Correct number of pre-terminals isn't learned with full Penn Treebank



Inference Algorithm Comparison

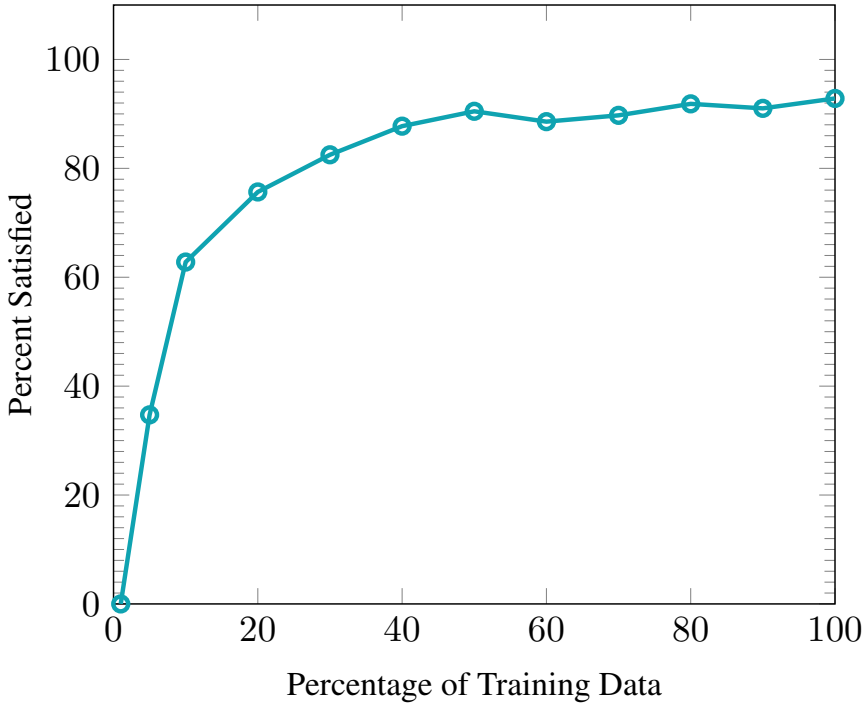
- Compare how well different algorithms satisfy constraints
  - Unconstrained
  - Constrained
  - Post-Hoc

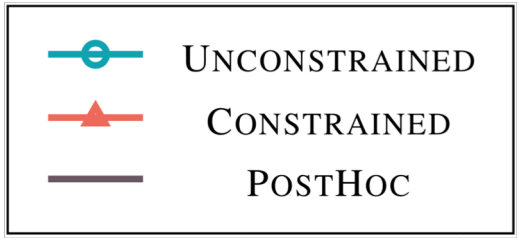
Post-Hoc Inference

- Run unconstrained beam search with a beam of size k
- Return highest-scoring valid output

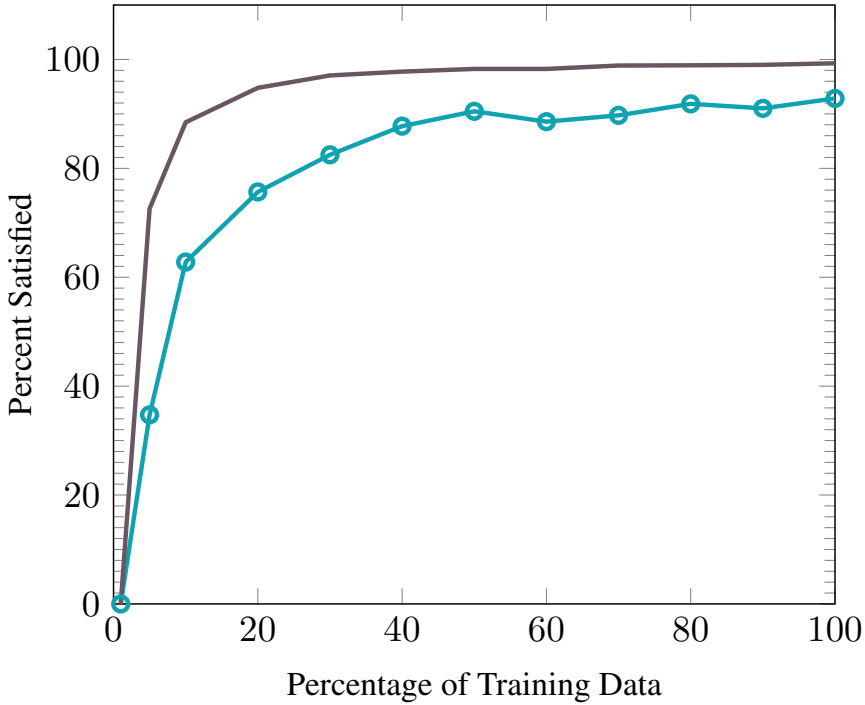
0.32	(S (NP) (VP XX XX (NP XX)))
0.28	(S (VP XX (NP XX)))
0.25	(S (NP XX) (VP XX (NP XX)))
0.10	(S (NP XX) (VP XX (NP XX))))

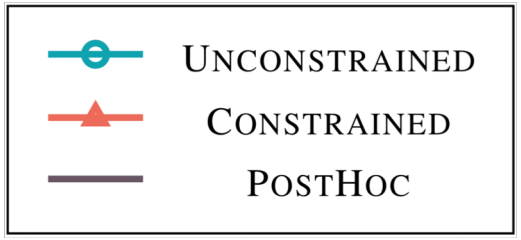
Parsing Learning Curve



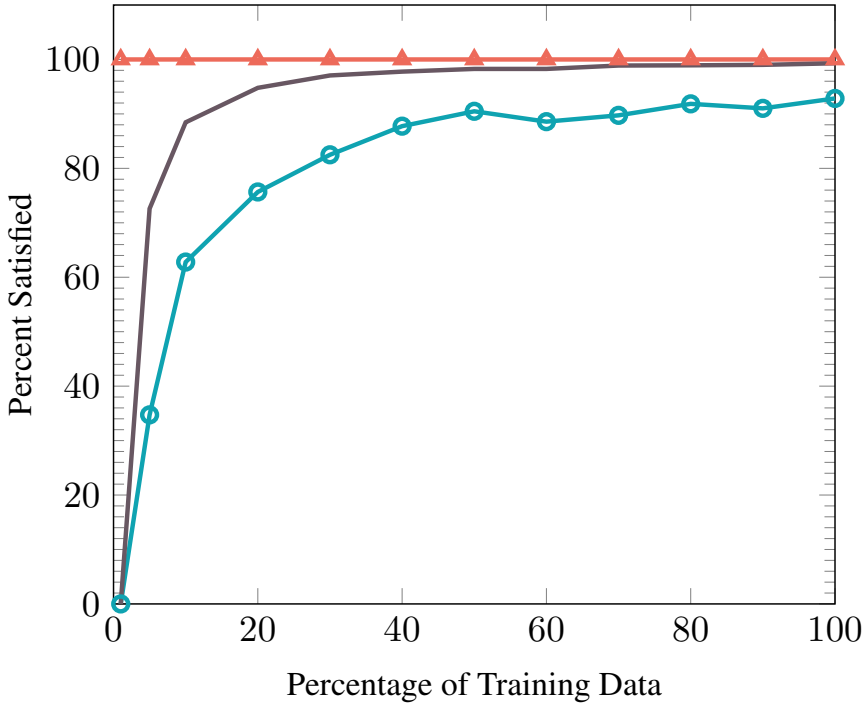


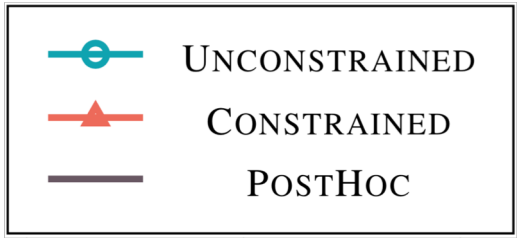
Parsing Learning Curve



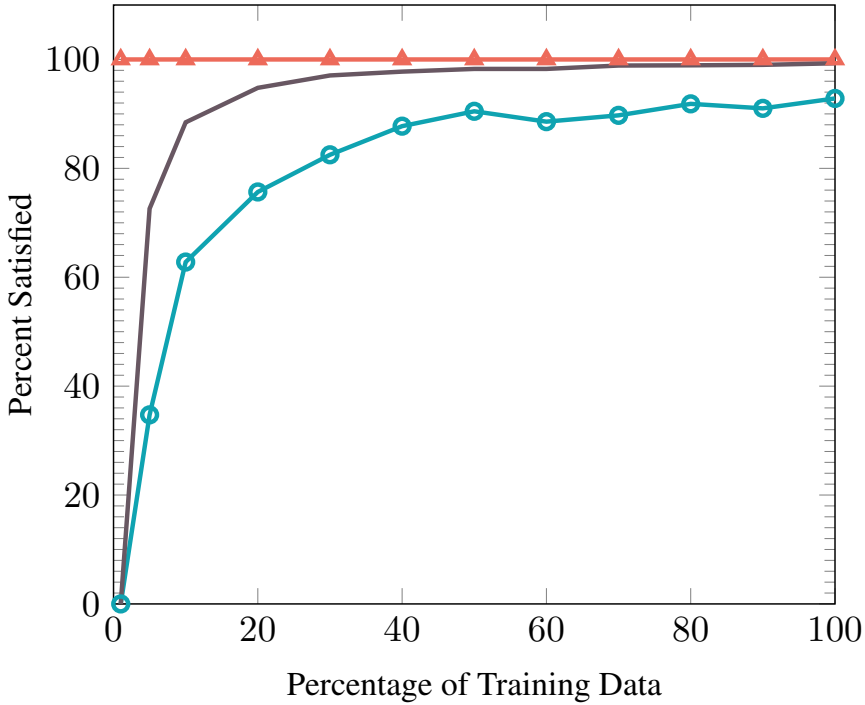


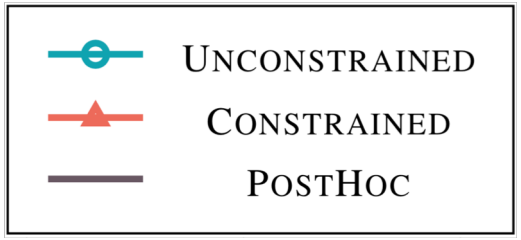
Parsing Learning Curve





Parsing Learning Curve

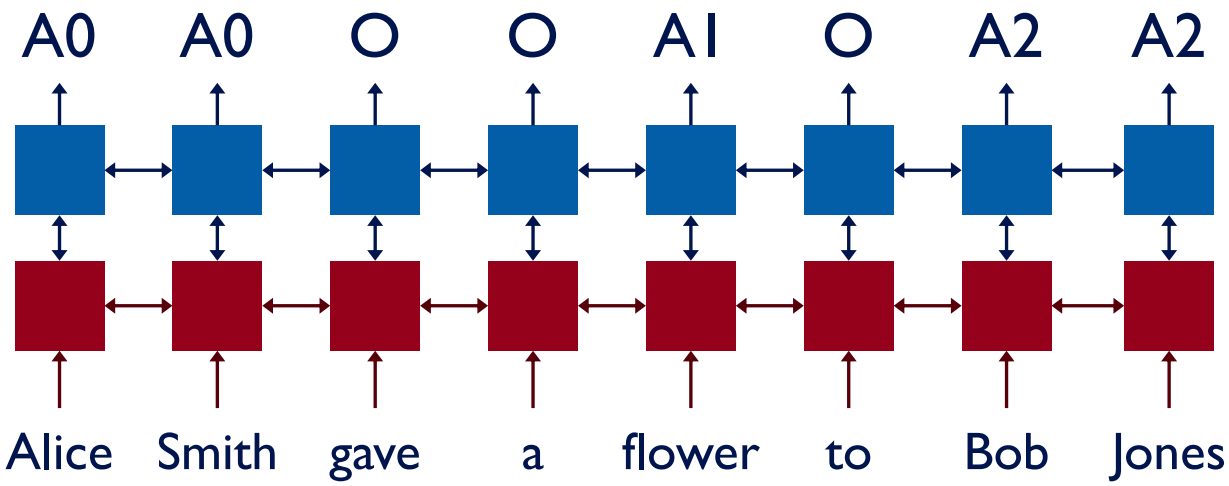




F_1 details available in the paper

Semantic Role Labeling Experiments

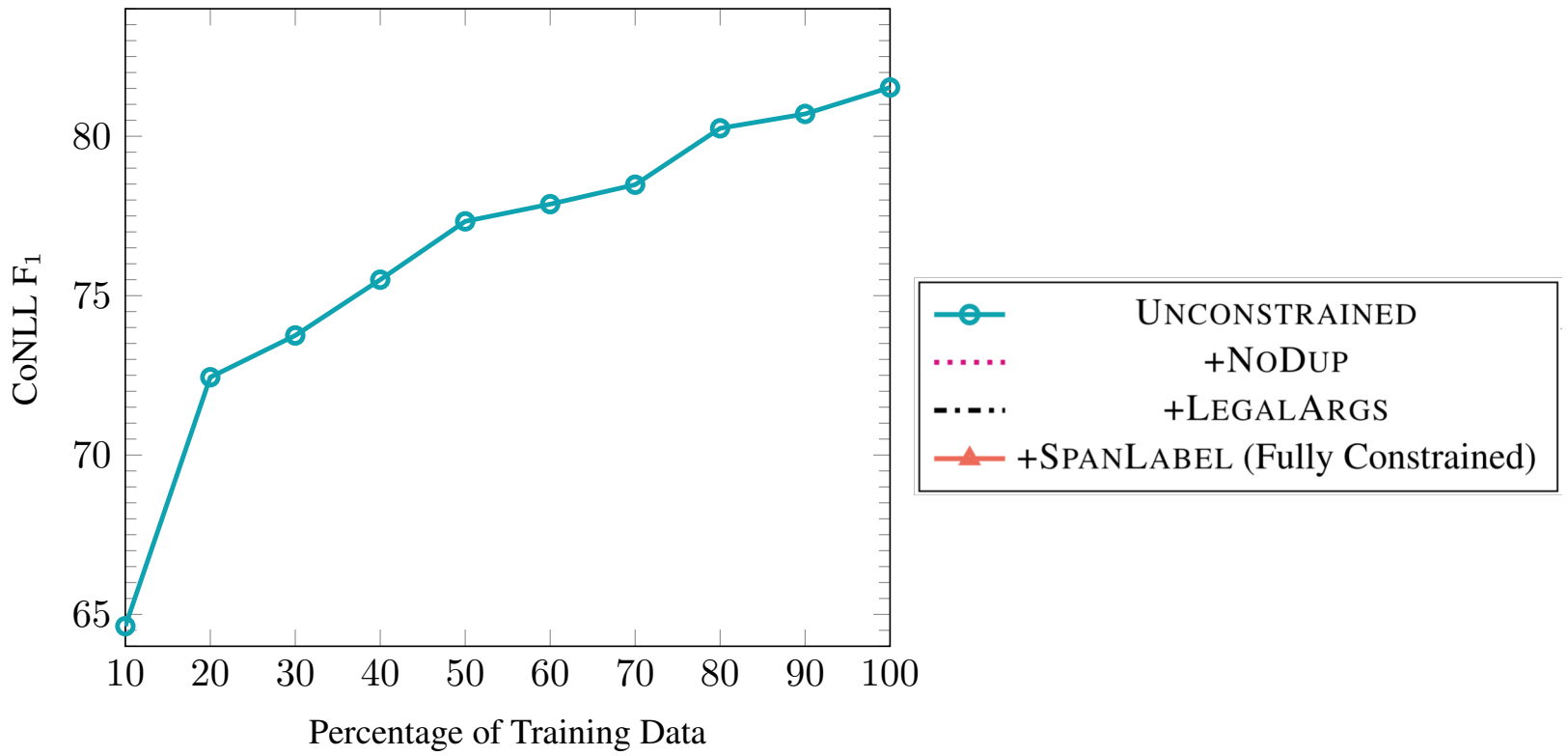
- Model of He et al. (2017) on CoNLL 2005
- Assume gold predicates



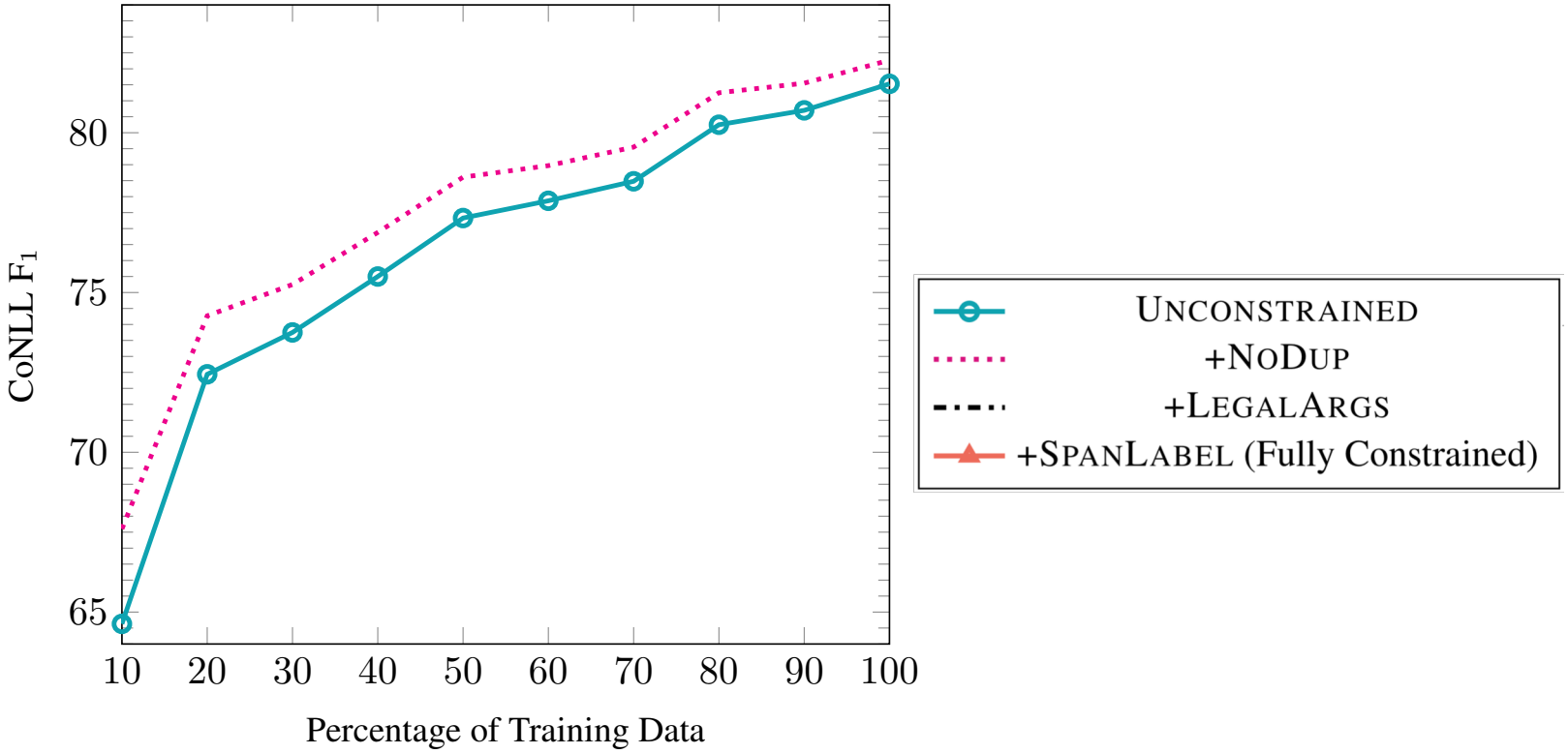
Semantic Role Labeling Learning Curve

- Impose constraints at varying levels of supervision
  - Disallow duplicate arguments
  - Disallow invalid arguments (via Propbank)
  - Require spans to have the same label

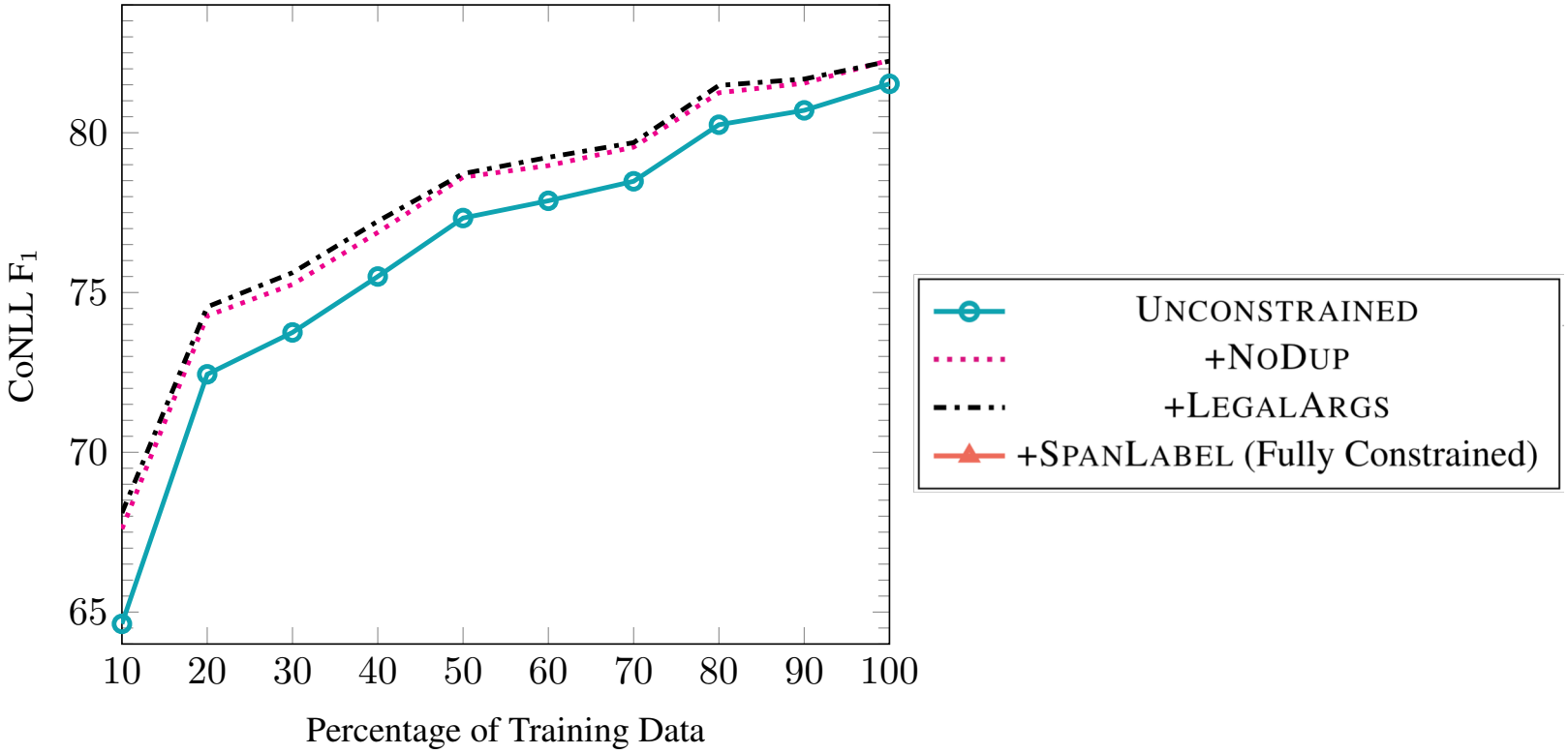
Semantic Role Labeling Learning Curve



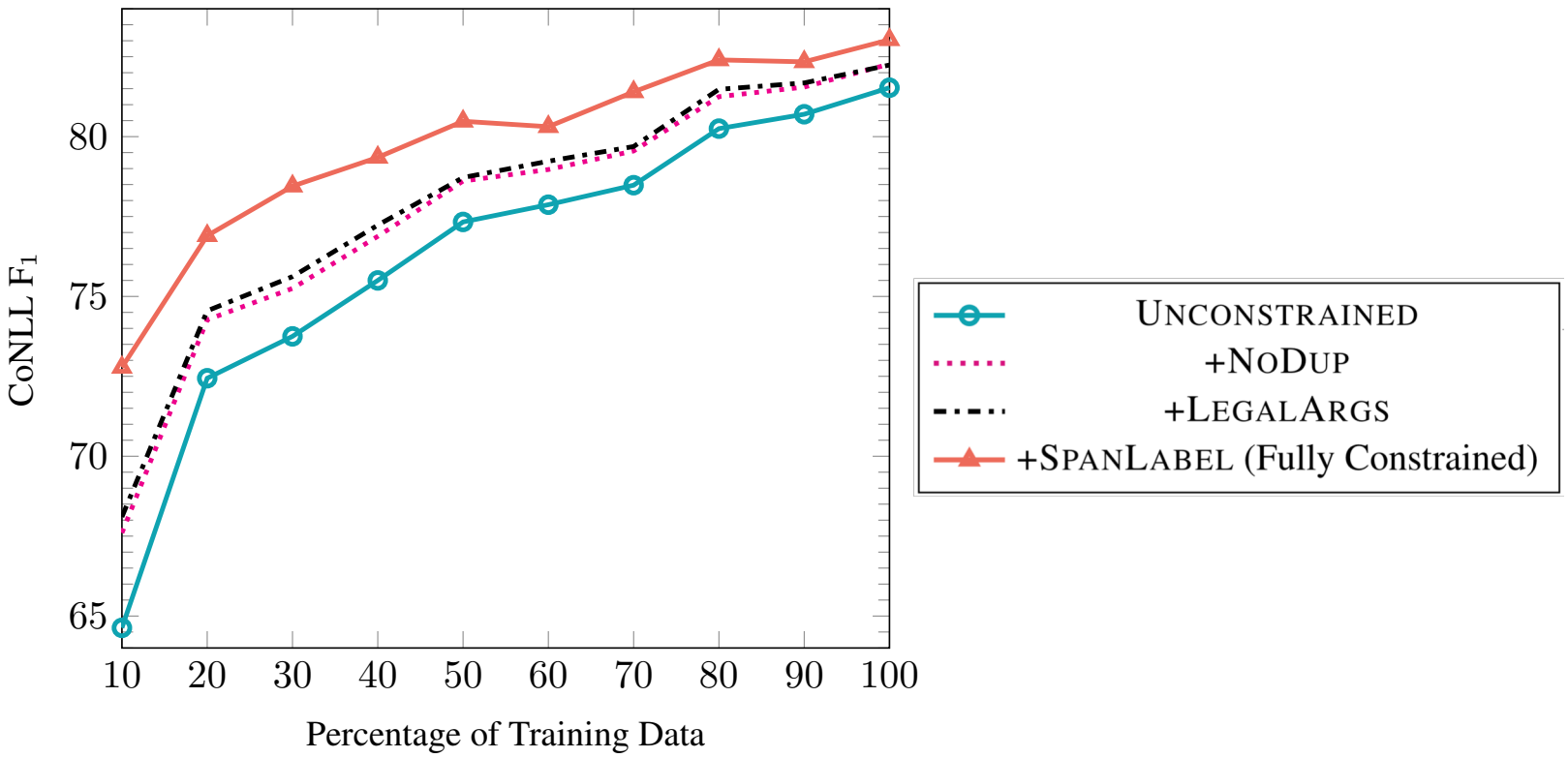
Semantic Role Labeling Learning Curve



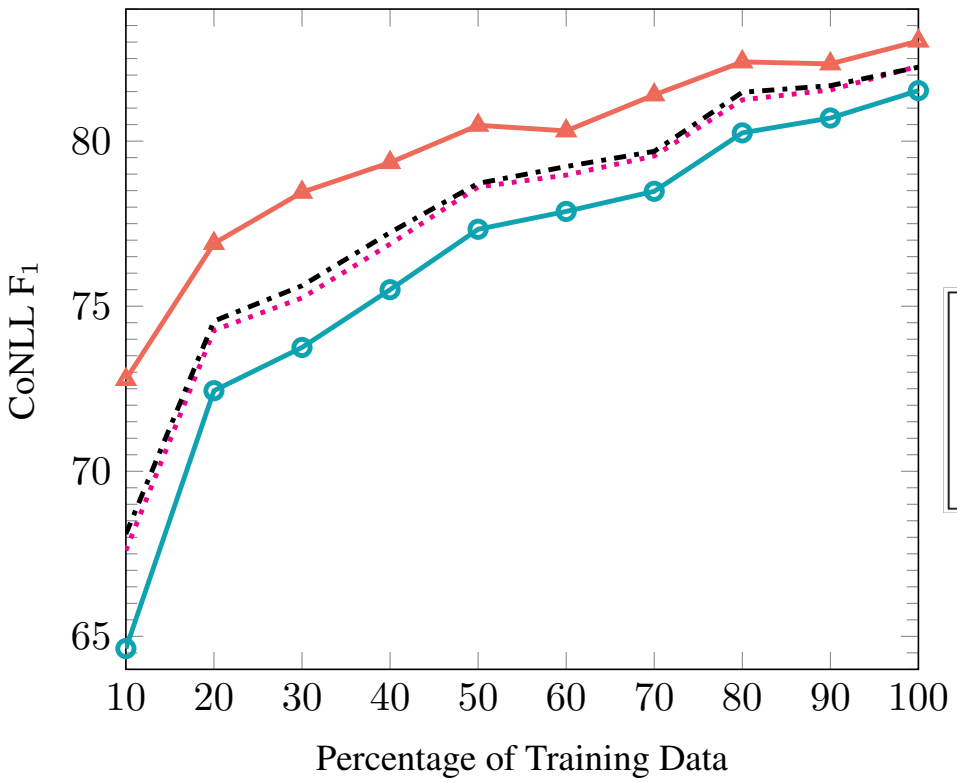
Semantic Role Labeling Learning Curve



Semantic Role Labeling Learning Curve

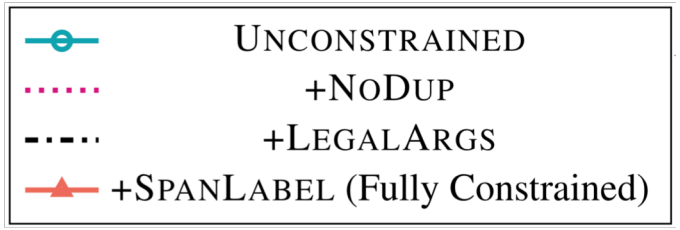


Semantic Role Labeling Learning Curve



Constraints consistently help at all levels of supervision

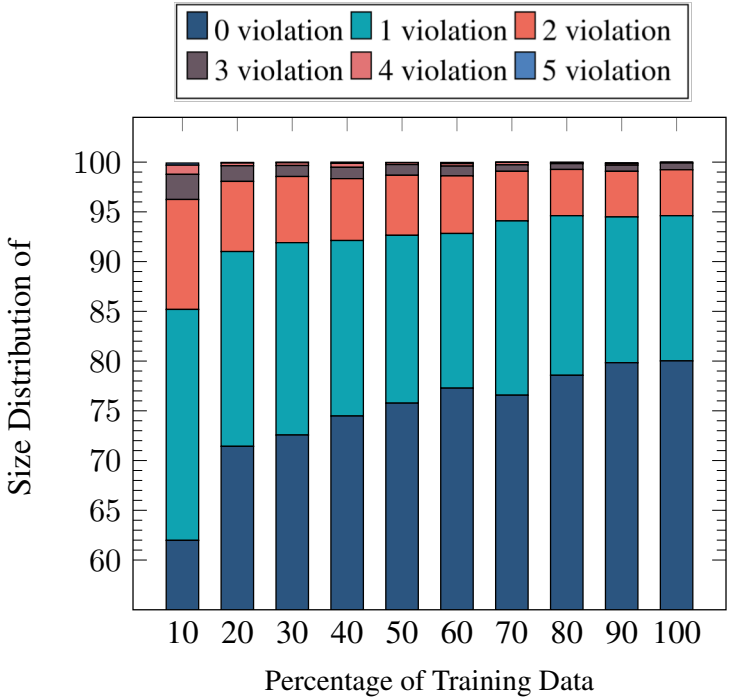
Constraints help most in low-supervision settings

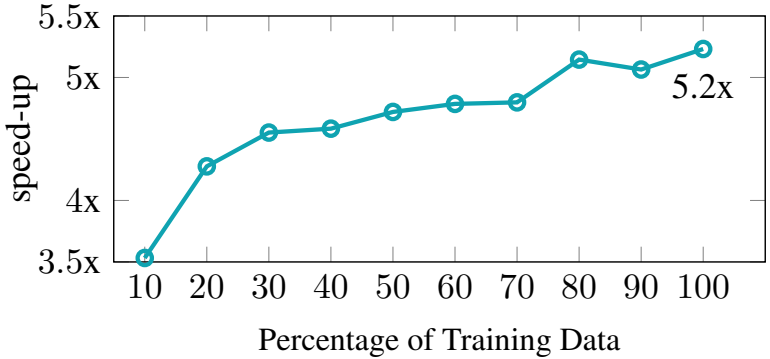


Active Set Size & Speedup

- Measured number of constraints added to the working set until a valid output found
- Calculated speed of inference relative to intersecting all constraints then running inference

Active Set Size & Speedup





Fewer constraints are necessary to impose at higher levels of supervision, which results in larger speedups

Conclusion

- Presented a general-purpose algorithm to enforce constraints in sequential inference
- Demonstrated enforcing constraints is necessary and beneficial at all levels of supervision
- Showed the active set method for imposing multiple constraints results in faster inference

Thank you!

https://github.com/CogComp/gcd

References

- Cyril Allauzen, Michael Riley, Johan Schalkwyk, Wojciech Skut, and Mehryar Mohri. 2007. OpenFst: A general and efficient weighted finite-state transducer library. In *International Conference on Implementation and Application of Automata*.
- Kyle Gorman. 2016. Pynini: A Python Library for Weighted Finite-state Grammar Compilation. In *Proceedings of the SIGFSM Workshop on Statistical NLP and Weighted Automata*.
- Luheng He, Kenton Lee, Mike Lewis, and Luke Zettlemoyer. 2017. Deep Semantic Role Labeling: What works and What's Next. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*.
- Hao Zhang, Richard Sproat, Alex H. Ng, Felix Stahlberg, Xiaochang Peng, Kyle Gorman, and Brian Roark. 2019. Neural Models of Text Normalization for Speech Applications. In *Computational Linguistics*.
- Muhua Zhu, Yue Zhang, Wenliang Chen, Min Zhang, Jingbo Zhu. 2013. Fast and accurate shift-reduce constituency parsing. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*.