Semantic Role Labeling and its Extensions

Vivek Srikumar

past

present

future

Joint work with Dan Roth

February 19, 2014, Nuance Communications

Understanding text

On Feb. 20, 1962, a Marine Corps fighter pilot from small-town America stepped forward in response to the country's need. The astronaut was John Glenn, whom the author Tom Wolfe has called "the last true national hero America has ever had".
On Monday and Tuesday, Glenn will be honored with a dinner and a spaceflight forum at Ohio State University, home of the John Glenn School of Public Affairs.

- Tom Wolfe called John Glenn "the last true national hero America has ever had".
- Tom Wolfe is an author. ~~Glenn wrote about Wolfe.~~
- The spaceflight forum is located in the Ohio State University.
- John Glenn will be honored on Monday.
- John Glenn School of Public Affairs is located at the Ohio State University.

Understanding text

On Feb. 20, 1962, a Marine Corps fighter pilot from small-town America stepped forward in response to the country's need. The astronaut was John Glenn, whom the author Tom Wolfe has called "the last true national hero America has ever had".
On Monday and Tuesday, Glenn will be honored with a dinner and a spaceflight forum at Ohio State University, home of the John Glenn School of Public Affairs.

Who called John Glenn "the last true national hero America has ever had"?

~~A: The astronaut~~

A: The author Tom Wolfe

Semantic Role Labeling (SRL)

Given a sentence, identifies who does what to whom, where and when.

The astronaut was John Glenn, whom the author Tom Wolfe has **calle d**
“the last true national hero America has ever had.”

Predicate

Arguments

Relation: `Call`
Caller[**A0**]: the author Tom Wolfe
Item being labeled [**A1**]: John Glenn
Attribute [**A2**]: the last true national hero America has ever had

“**the author Tom Wolfe**” fulfills the role of *Caller* for the predicate `Call`

Who called John Glenn “the last true national hero America has ever had”?

Verbs do not give the full picture

On Feb. 20, 1962, a Marine Corps fighter pilot from small-town America stepped forward in response to the country's need. The astronaut was John Glenn, whom the author Tom Wolfe has called "the last true national hero America has ever had".
On Monday and Tuesday, Glenn will be honored with a dinner and a spaceflight forum at Ohio State University, home of the John Glenn School of Public Affairs.

- ✓ Tom Wolfe called John Glenn "the last true national hero America has ever had".
- ✗ Tom Wolfe is an author.
- ✗ The spaceflight forum is located in the Ohio State University.
- ✓ John Glenn will be honored on Monday.
- ✗ John Glenn School of Public Affairs is located at the Ohio State University.

Outline

- Semantic Role Labeling (SRL) for verbs
  - Integer linear program for inference
- Extending semantic role labeling: Prepositions
  - Ontology of preposition triggered relations
  - Modeling preposition relations
  - Verb-preposition interplay
- Closing thoughts

Semantic role labeling of verbs

Verb Semantic Role Labeling

- PropBank [Palmer et. al. 05]
  - Large human-annotated corpus of verb semantic relations
  - Provides a semantic layer over the parse trees of Penn Treebank [Marcus et al ‘94]

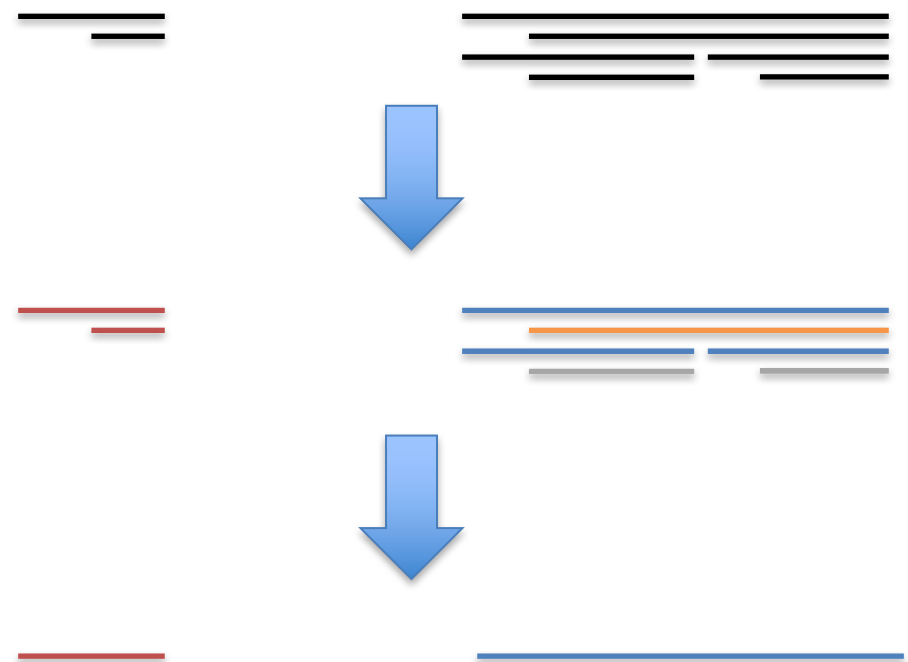
Relation: `Call`
Caller[**A0**]: the author Tom Wolfe
Item being labeled [**A1**]: John Glenn
Attribute [**A2**]: the last true national hero America has ever had

Predicting verb arguments

[Punyakanok, et al '08]

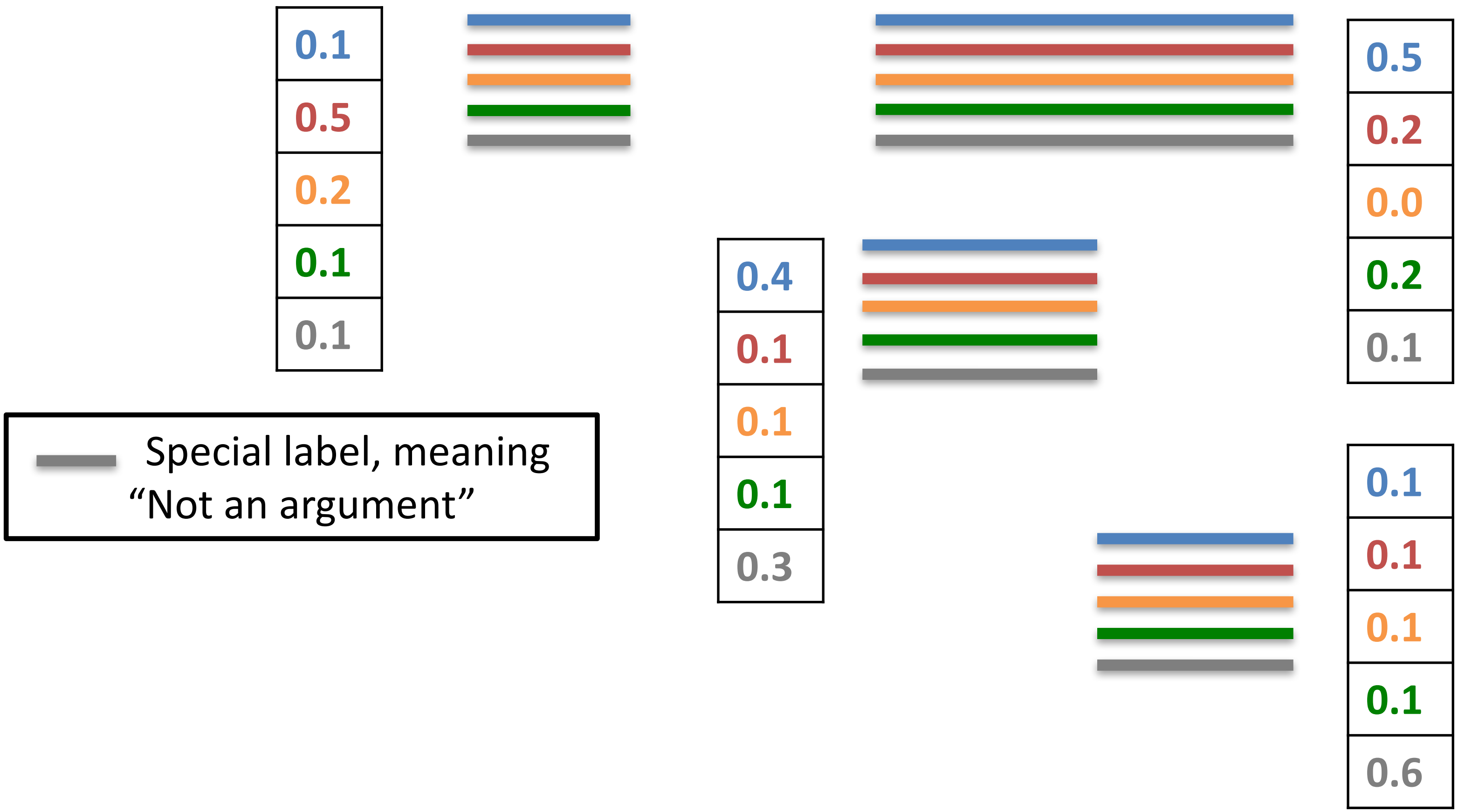
1. Identify candidate arguments for verb using parse tree [Xue & Palmer '04]
   – Filtered using a binary classifier
2. Classify argument candidates
   – Multi-class classifier (one of multiple labels per candidate)
3. Inference
   – Using probability estimates from argument classifier
   – Must respect structural and linguistic constraints
     • Eg: No overlapping arguments

The bus was **heading** for Nairobi in Kenya.



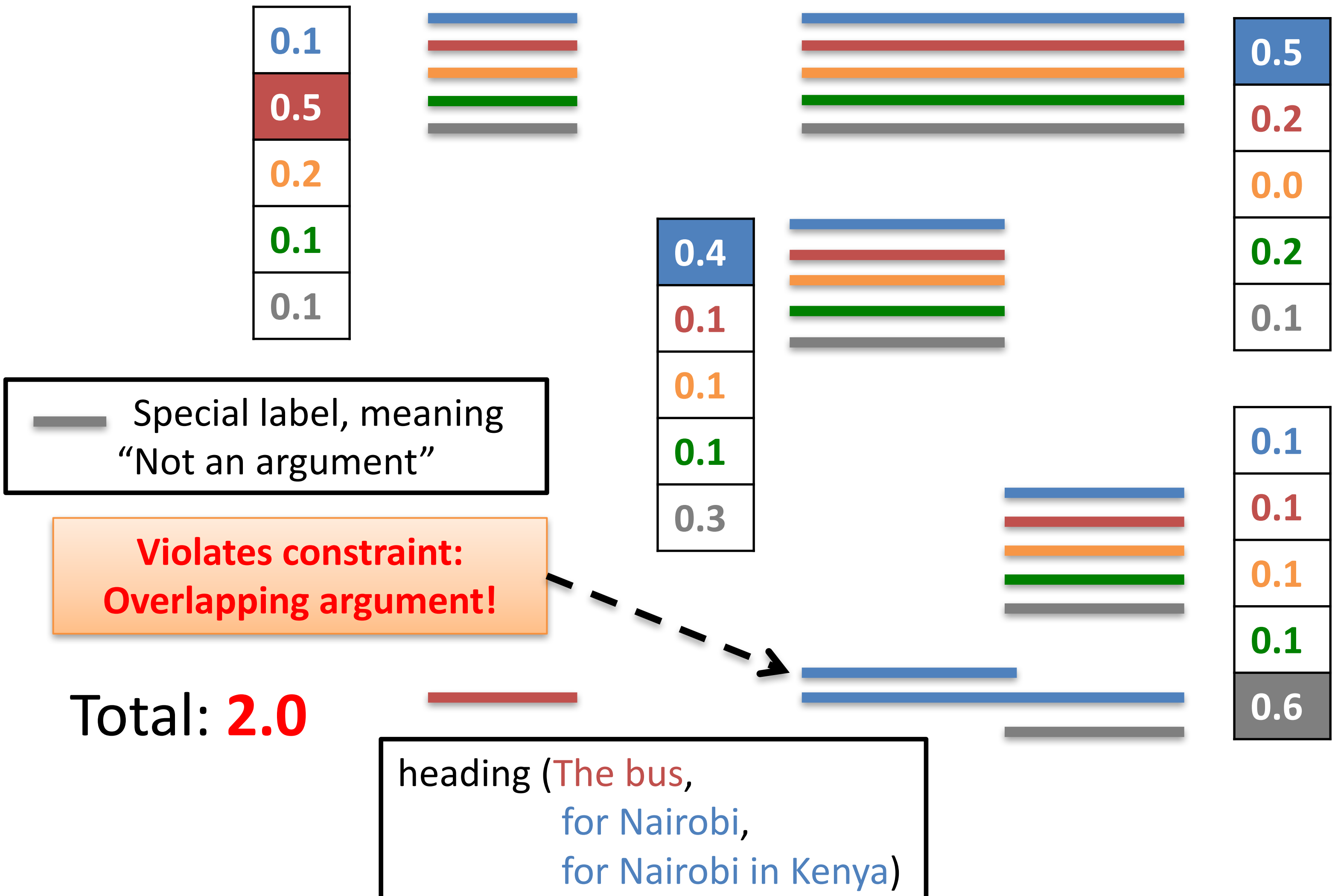
Inference: verb arguments

The bus was **heading** for Nairobi in Kenya.



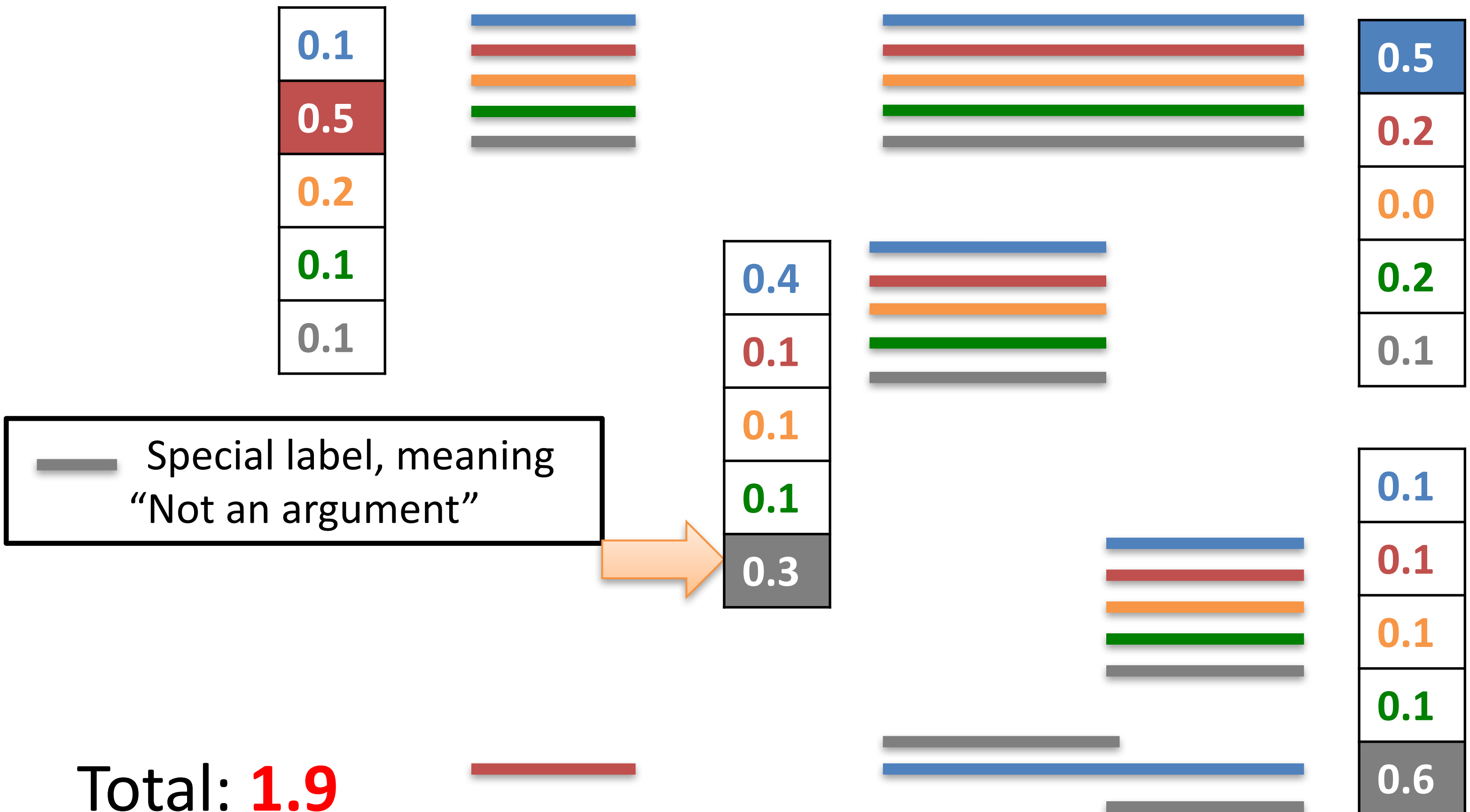
Inference: verb arguments

The bus was **heading** for Nairobi in Kenya.



Inference: verb arguments

The bus was **heading** for Nairobi in Kenya.



heading (The bus,
for Nairobi in Kenya)

Supervised Learning

Structured SVM: [Tsochantaridis, et al 2004]

- Given a collection of labeled examples $\{(\mathbf{x}_i, \mathbf{y}_i)\}$ find a scoring function that minimizes loss

Real valued weight vector that assigns scores to feature representations of structures

Score of annotated structure $\geq$ Score of any other structure $+$ Penalty for predicting other structure (margin)

Feature vector: (Sparse) vector of attributes that capture dependencies between input and output, collecting features from each part

- hms

Inference

Score for decision y^i
Denoted by c^i

- Find a score maximizing ***feasible*** structure

$$arg\max_{\mathbf{y}} \mathbf{w}^T \Phi(\mathbf{x}, \mathbf{y}) = arg\max_{\mathbf{y}} \sum_i y^i w^T \phi_i(\mathbf{x}, y^i)$$

- For some structures, inference is computationally easy
  - Eg: Using the Viterbi algorithm

- In general, NP-hard,
  - In NLP, can be solved using a 0-1 **Integer linear program solver**
  - Constraints specified by linear inequalities over the variables

0-1 integer linear program for SRL

- Introduce binary variable y^i for inference decisions
  - For this specific case, call this *decision* $y^{a,t}$ and the corresponding *score* $c^{a,t}$

- Constraints (simple version)
  1. Exactly one label per argument
  2. No overlapping arguments
  3. Unique core arguments

$$\max_{\mathbf{y}} \sum_i y^i c^i = \sum_{a,t} y^{a,t} c^{a,t}$$

$$\forall a, \quad \sum_t y^{a,t} = 1$$

Other constraints can be similarly written as linear (in)equalities

[Roth and Yih, 05], [Clarke and Lapata, 08], many others

Brief interlude: Solving ILPs

- Thinking about ILPs makes it easy to write down inference problems
  - Not any easier to solve!

- Some ideas to solve
  - If possible, use dynamic programming
  - Commercial solvers
  - Cutting plane technique [Riedel '08]
  - Relaxation techniques, like Lagrangian relaxation [Rush&Collins '11], others
  - *Amortized inference* [Srikumar, Kundu & Roth, 12]
  - Of course, approximate algorithms

Verb SRL is not sufficient

John, a fast-rising politician, slept on the train to Chicago.

Relation: **sleep**
- **Sleeper**: John, a fast-rising politician
- **Location**: on the train to Chicago.

What was John's destination?

train to Chicago

↳ `Destination`(train, Chicago)

EMNLP '11, TACL '13

Who was John?

John, a fast-rising politician

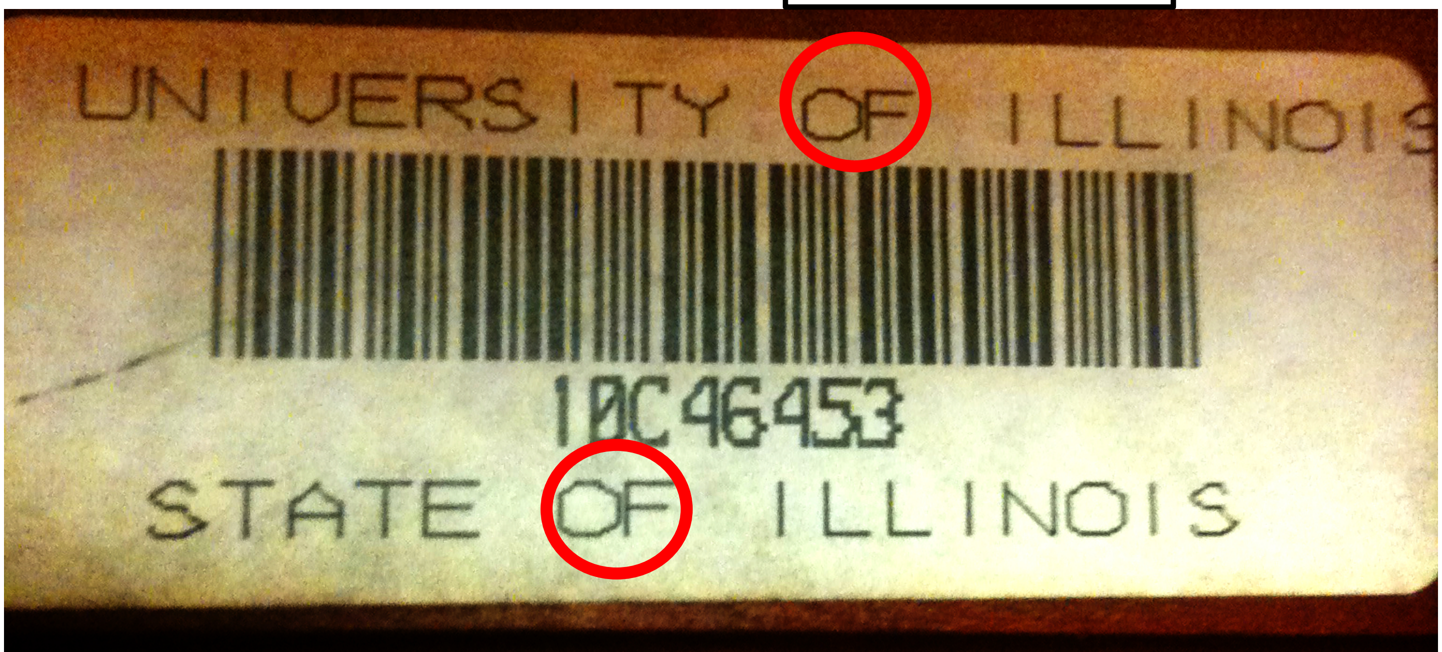
↳ `IS-A`(John, a fast-rising politician)

ACL '08

Ontology of Preposition Relations

Examples of preposition relations

Possessor



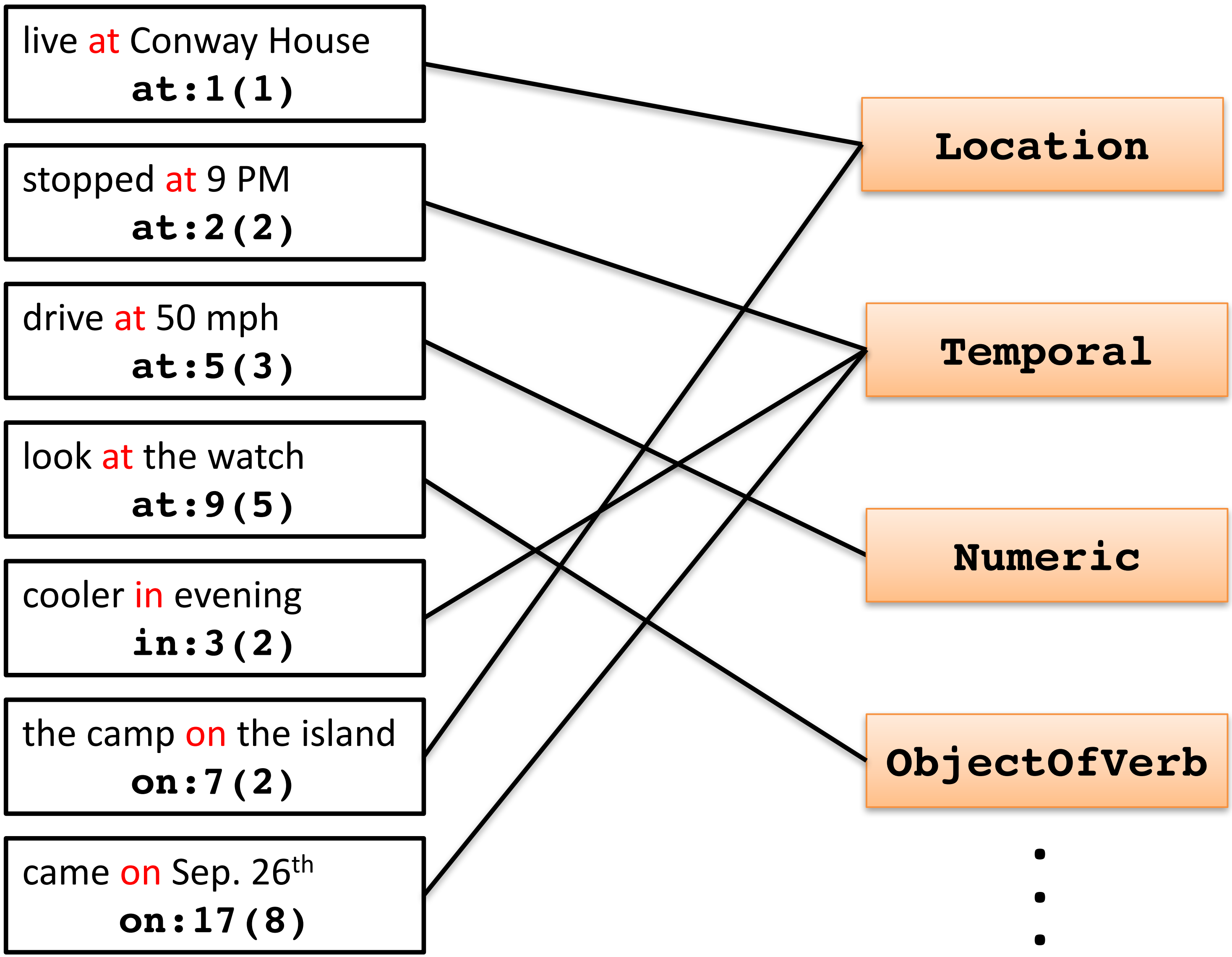
Species

Preposition Sense Disambiguation

Eg. *State* ***of*** *Illinois* vs. *University* ***of*** *Illinois*

- The Preposition Project [Litkowski and Hargraves, 2005]
  - Word sense for 34 prepositions
  - Based on preposition definitions in Oxford Dictionary of English

Mapping from senses to relations



An inventory of preposition relations

- Labels that act as the predicate
  - Semantically related senses of prepositions merged
  - ~250 senses ➔ 32 relation labels

- Word sense disambiguation data, re-labeled
  - SemEval 2007 shared task gives relation labeled data
    - ~16K training and ~8K test instances
    - 34 prepositions

List of preposition relations

Relation	Train	Test
Activity	63	39
Agent	367	159
Attribute	510	266
Beneficiary	205	105
Cause	591	289
Co-Participiants	112	58
Destination	1054	526
Direction	909	441
EndState	188	104
Experiencer	116	54
Instrument	565	290
Location	3096	1531
Manner	457	245
MediumOfCommunication	57	30
Numeric	113	48
ObjectOfVerb	1801	882

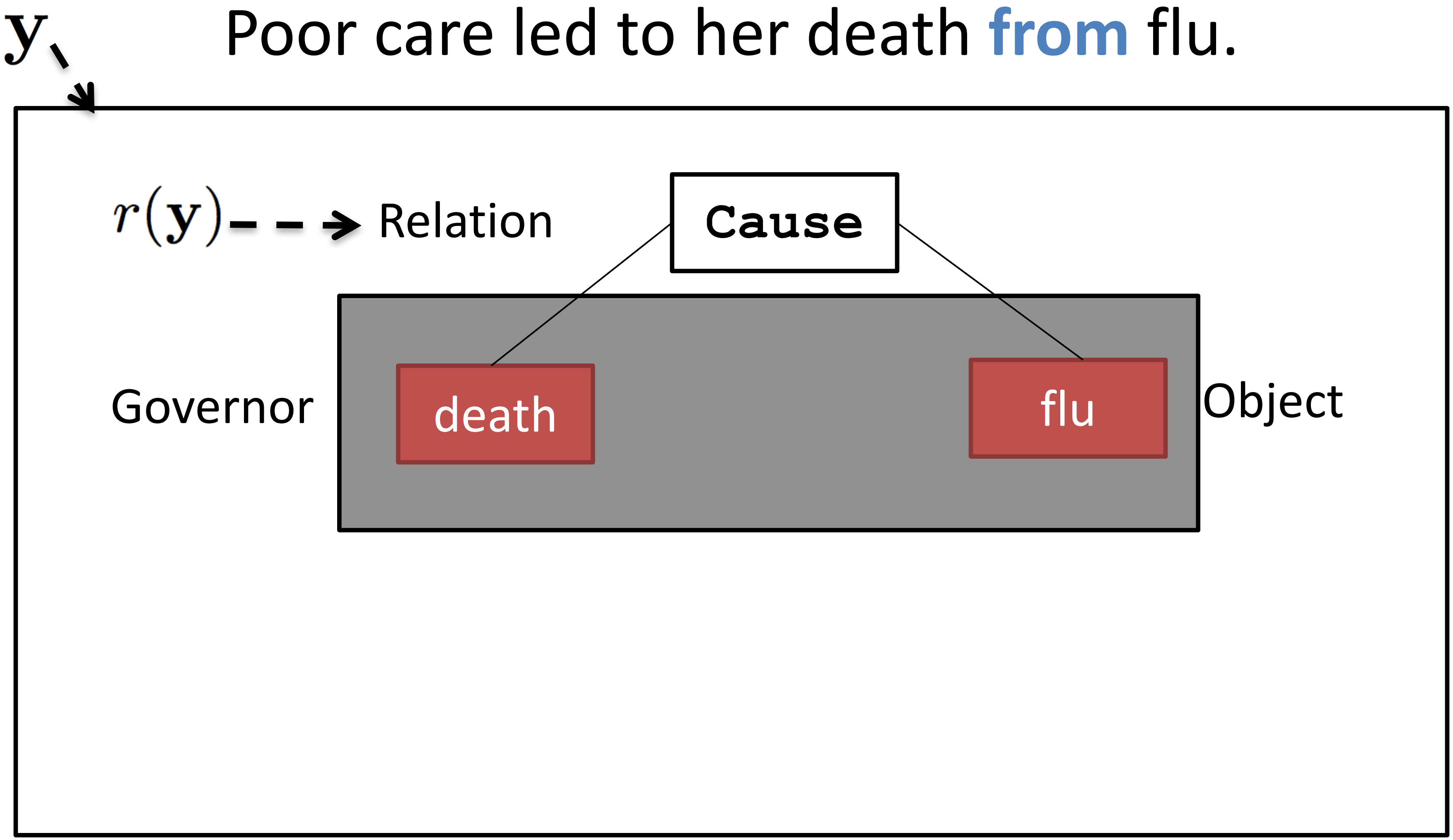
Relation	Train	Test
Opponent/Contrast	233	131
Other	72	42
PartWhole	958	471
Participant/Accompanier	292	142
PhysicalSupport	399	202
Possessor	508	269
ProfessionalAspect	45	22
Purpose	261	113
Recipient	378	190
Separation	345	172
Source	740	357
Species	394	198
StartState	69	37
Temporal	331	157
Topic	886	462
Via	61	26

“zoo in NYC” ⟶ `Location`(zoo, NYC)

Two Models for Predicting Preposition Relations

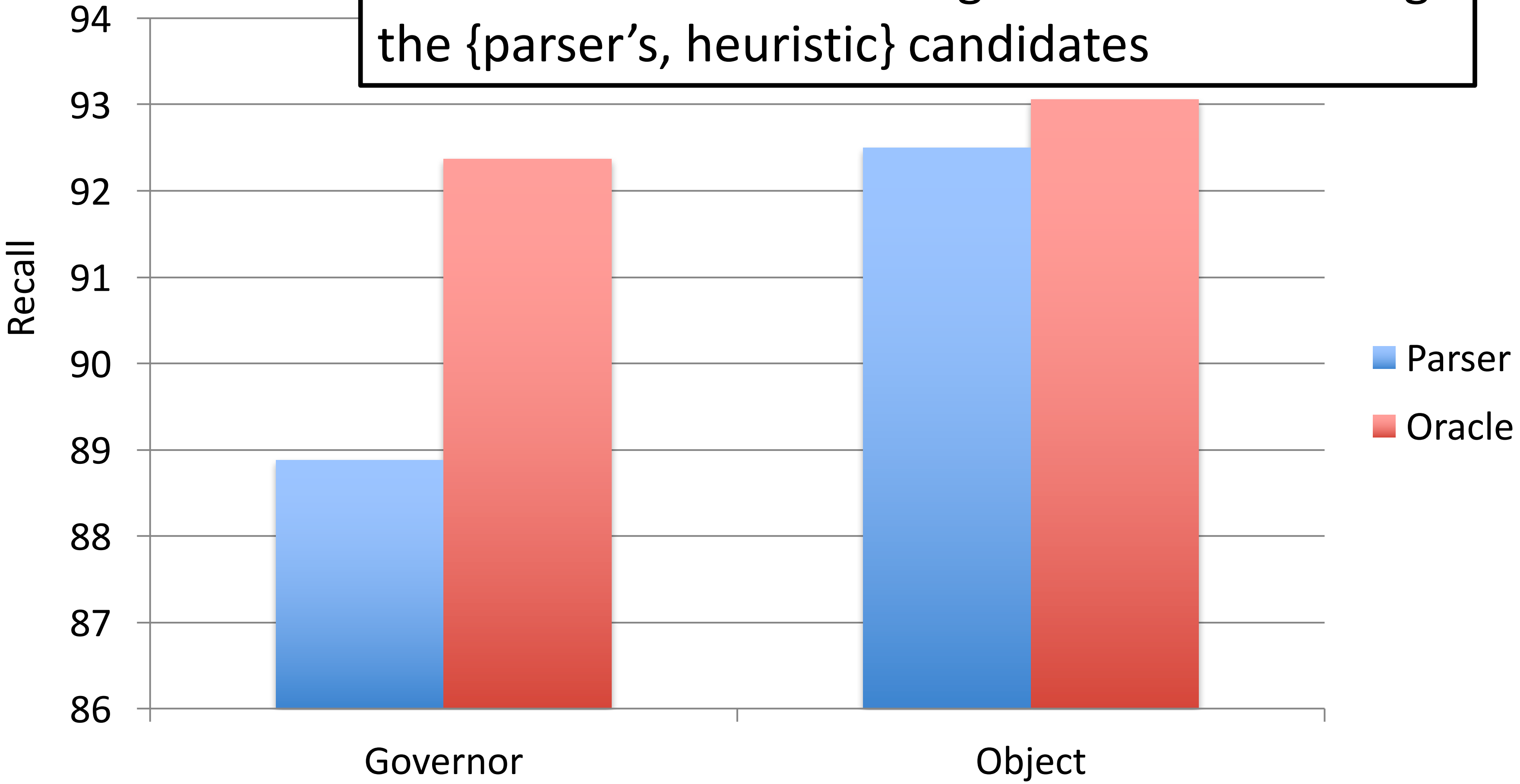
Structure of prepositions

Poor care led to her death **from** flu.



Parser not sufficient for arguments

Oracle: Selects the true argument if it is among the {parser's, heuristic} candidates



Easy-first dependency parser [Goldberg & Elhadad, 2010]

Relation depends on argument types

Poor care led to her death **from** flu.

`Cause`(death, flu)

Poor care led to her death **from** pneumonia.

How do we generalize the classifier to unseen arguments in the same “type”?

Why are types important?

- Goes beyond words
  - Abstract *flu and pneumonia* into the same group

- Some semantic relations hold only for certain types of entities

- Two notions of type
    - WordNet hypernyms
    - Distributional word clusters
  - Allow for multiple meanings and concept hierarchies

WordNet IS-A hierarchy

```
pneumonia
   => respiratory disease
      => disease
         => illness
            => ill health
               => pathological state
                  => physical condition
                     => condition
                        => state
                           => attribute
                              => abstraction
                                 => entity
```

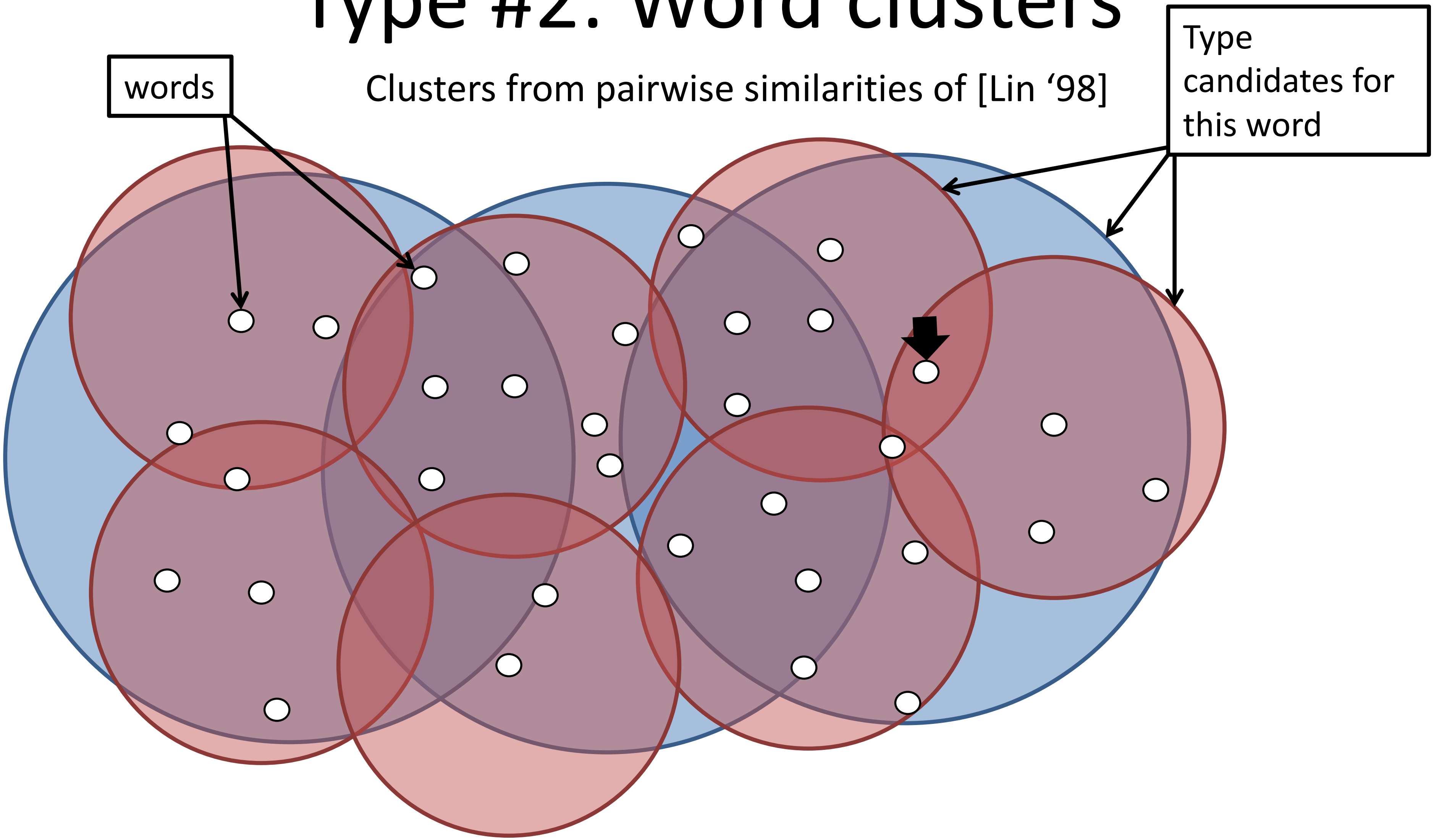
Picking the right level in this hierarchy can generalize pneumonia and flu

More general, but less discrimniative

Picking incorrectly will over-generalize

Type #2: Word clusters

Clusters from pairwise similarities of [Lin ‘98]



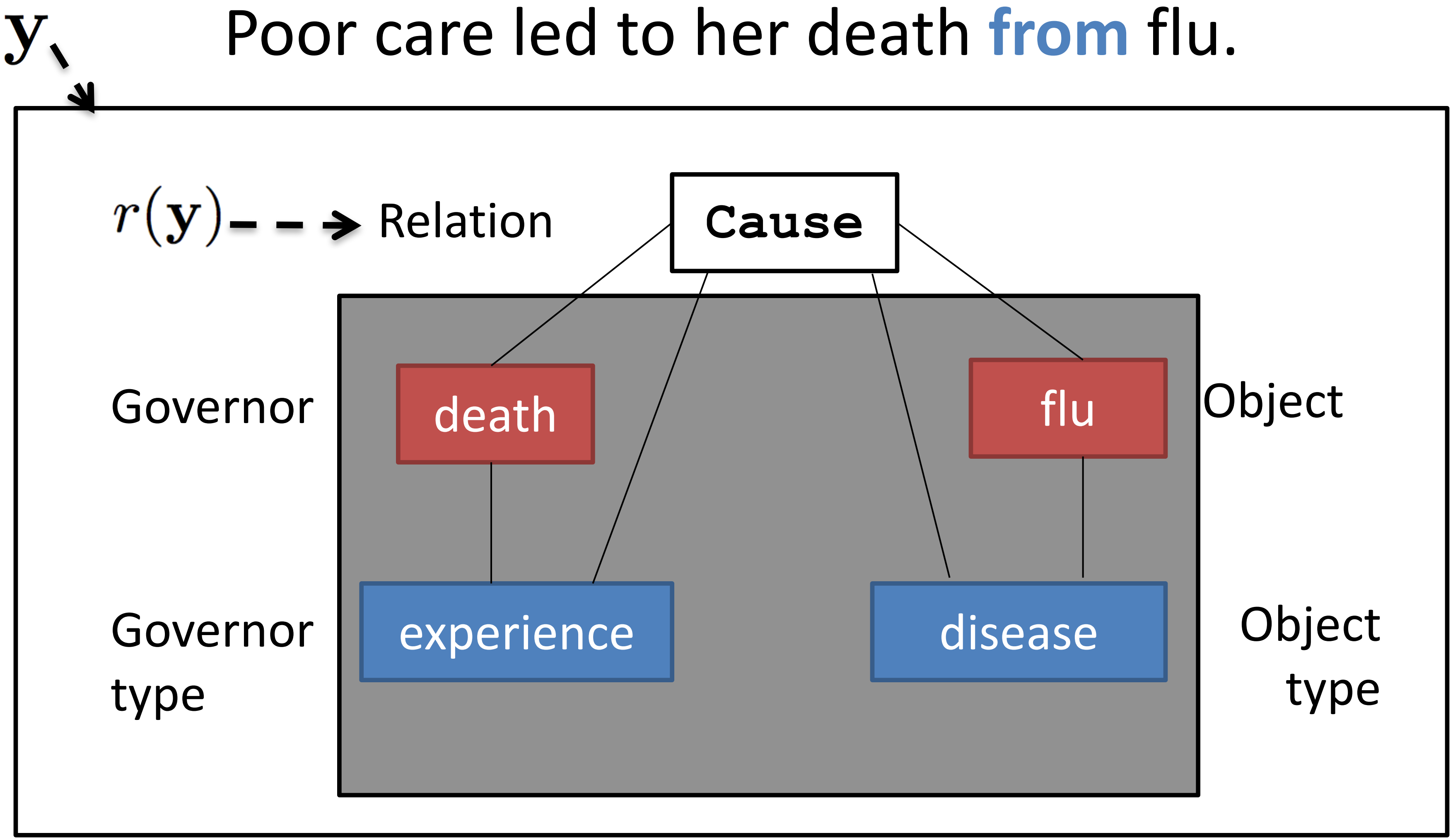
Cover all points with fixed radius clusters. Repeat with a different radii.

Example clusters

Jimmy Carter	metalwork	fox news channel	expert
Ronald Reagan	porcelain	NBC News	Wall Street analyst
richard nixon	handicraft	MSNBC	analyst
George Bush	jade	Fox News	economist
Lyndon Johnson	bronzeware	CNBC	telecommunications analyst
Richard M.	carving	CNNfn	strategist
Nixon	pottery	C-Span	media analyst
Gerald Ford	ceramic		
	earthenware		
	jewelry		
	stoneware		
	lacquerware		

Structure of prepositions

Poor care led to her death **from** flu.



Governor and object base features

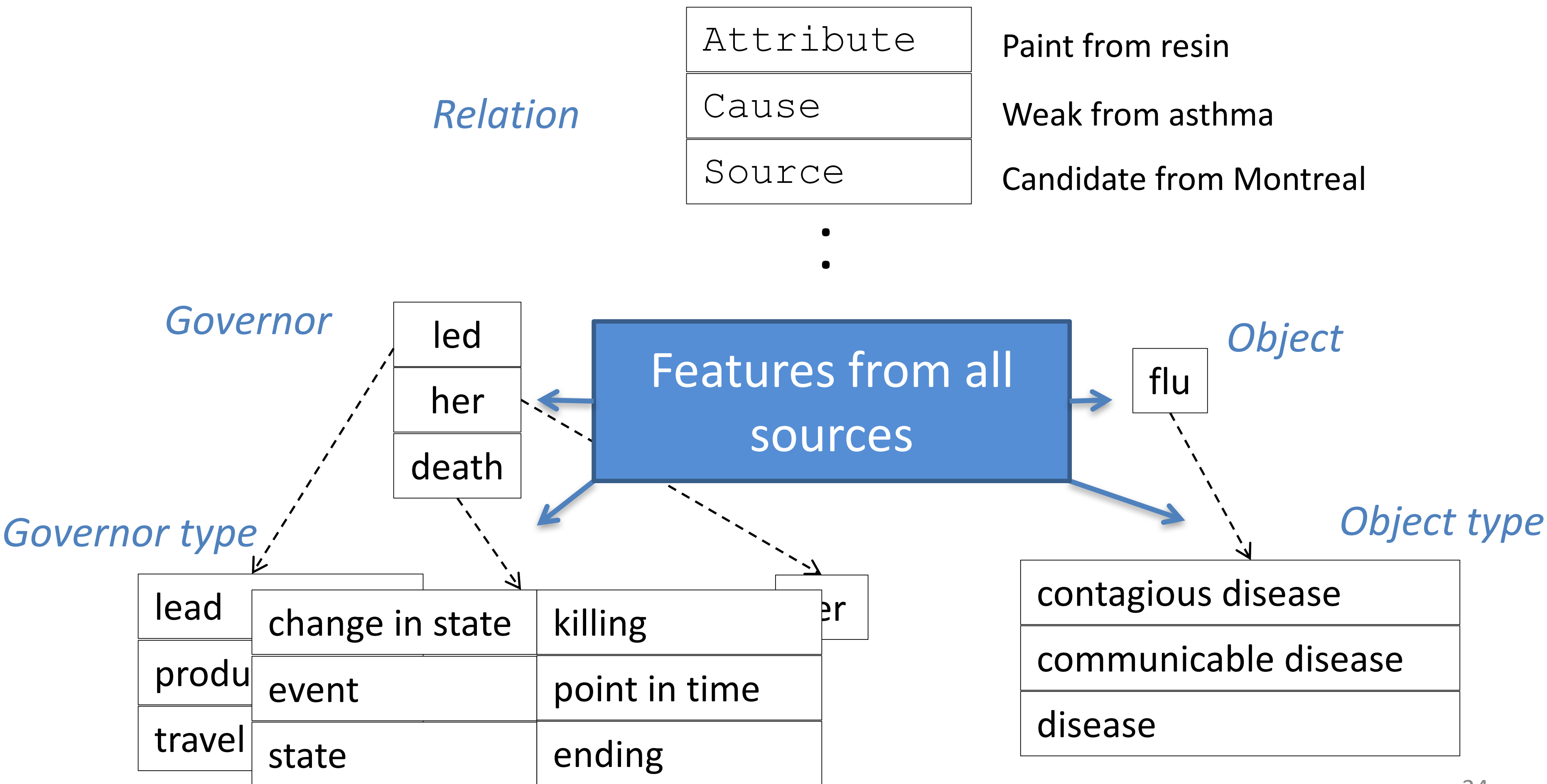
1. Word, POS, lemma and capitalization indicator
2. Conflated part-of-speech (Noun/Verb/Adjective/Adverb/Other)
3. Indicator for existence in WordNet,
4. WordNet synsets for the first and all senses,
5. WordNet lemma, lexicographer file names and part, member and substance holonyms,
6. Roget thesaurus divisions for the word,
7. The first and last two and three letters, and
8. Indicators for known affixes.

Two models

- Model 1
  - Predict only relation label: Multi-class
  - Use features from all possible governor and object candidates
    - Also types
- Model 2 uses features from the structure
  - Predict full structure: relation and arguments
    - Also types

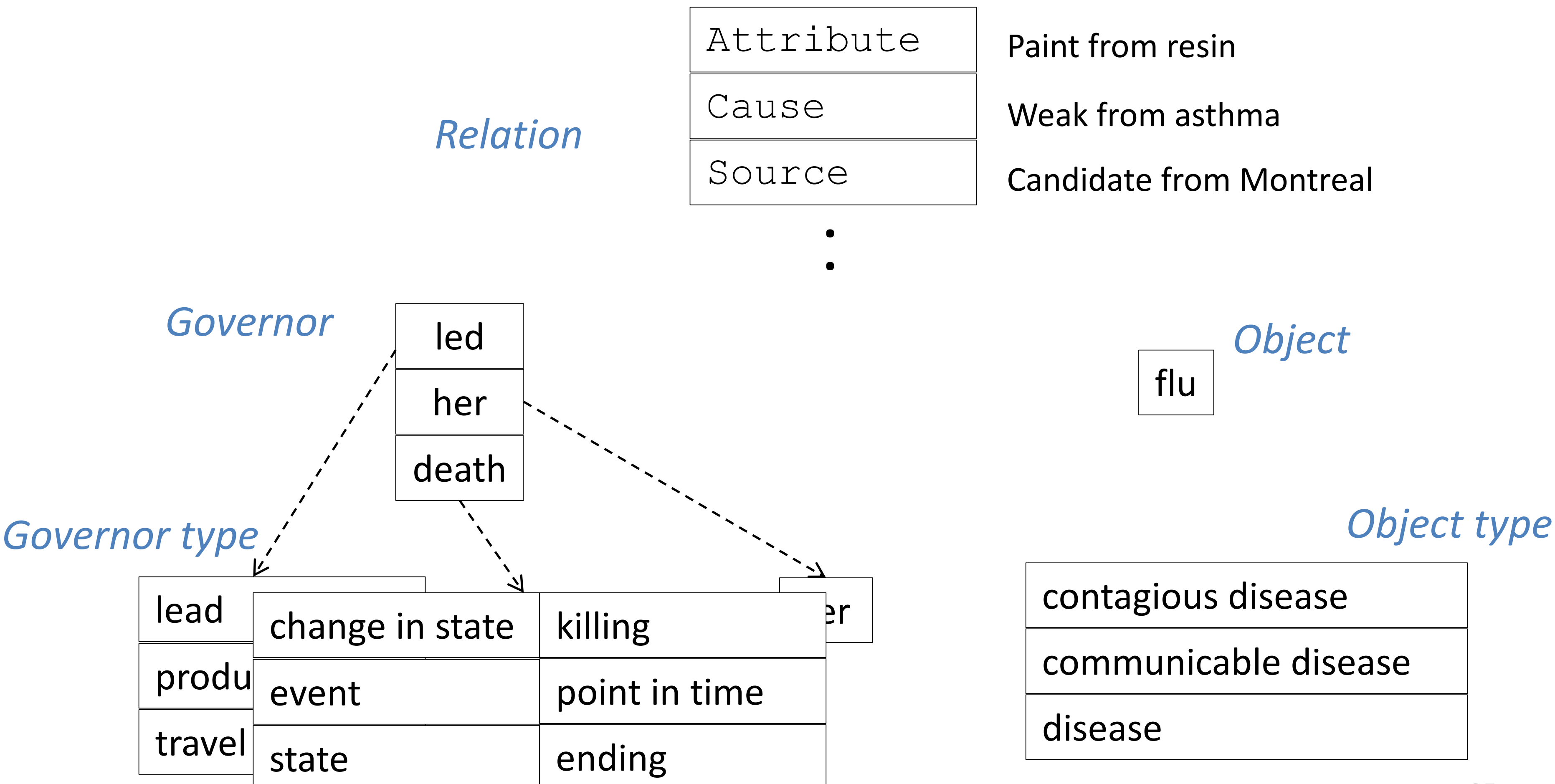
Model 1: Predict relation label

Poor care led to her death **from** flu.



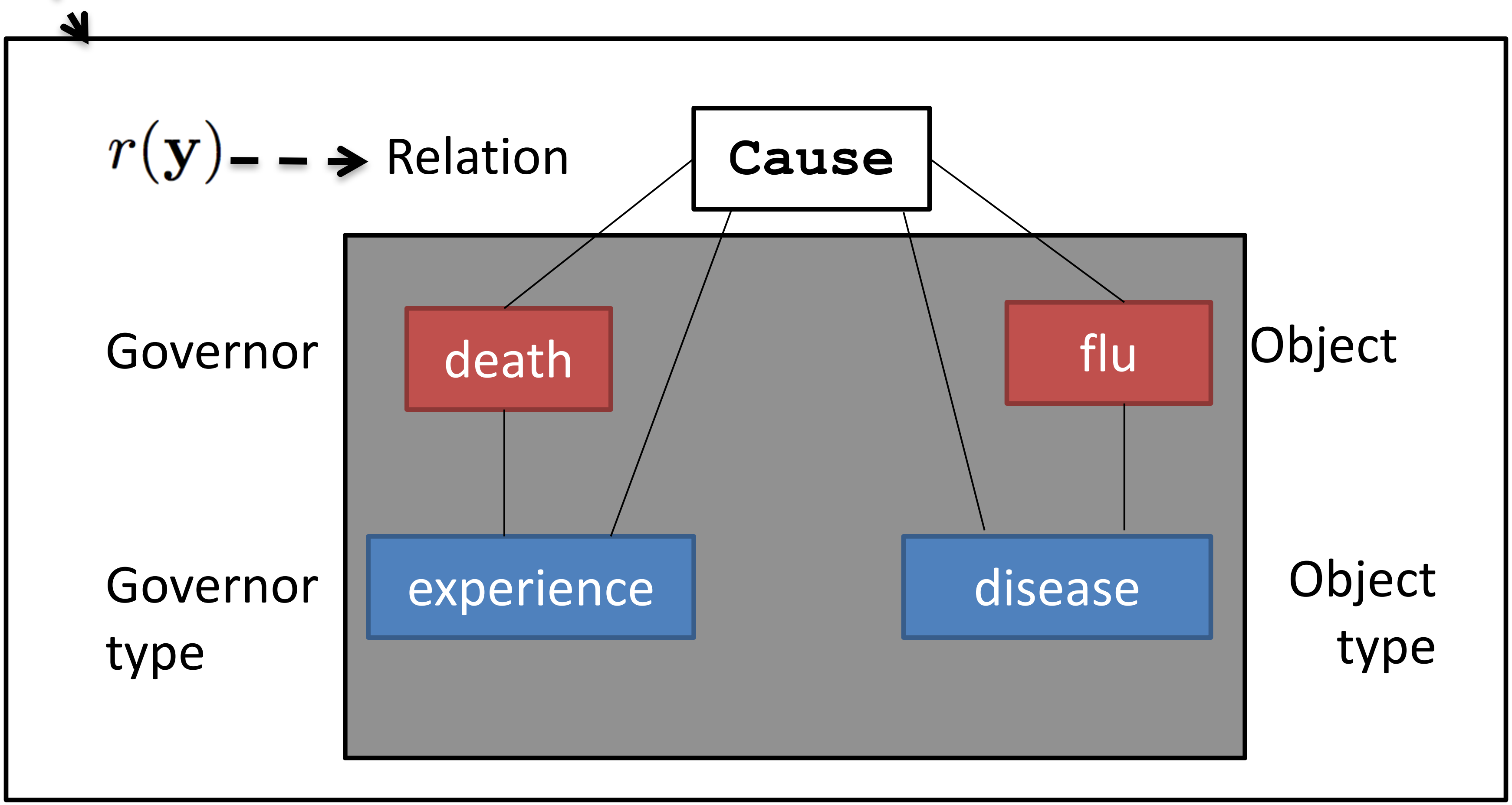
Model 2: Predict full structure

Poor care led to her death **from** flu.



Structure of prepositions

$\mathbf{y}$ Poor care led to her death **from** flu.



Learning Model 2: Latent inference

- ***Standard inference***: Find an assignment to the full structure

$$\max_{\mathbf{y}} \mathbf{w}^T \Phi(\mathbf{x}, \mathbf{y})$$

- ***Latent inference***: Given an example with annotated $r(\mathbf{y}^*)$

$$\begin{aligned} \max_{\mathbf{y}} \quad & \mathbf{w}^T \Phi(\mathbf{x}, \mathbf{y}) \\ \text{s.t.} \quad & r(\mathbf{y}^*) = r(\mathbf{y}) \end{aligned}$$

- *"Complete the structure given current model"*

Standard Supervised Learning

Structured SVM: [Tsochantaridis, et al 2004]

- Given a collection of labeled examples $\{(\mathbf{x}_i, \mathbf{y}_i)\}$ find a scoring function that minimizes loss

$$\min_{\mathbf{w},\xi} \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_i \xi_i$$

Score of annotated structure $\geq$ Score of any other structure $+$ Penalty for predicting other structure

- Several algorithms for learning.
  - We use the SVM solver from [Chang et al, 2010]

Learning Model 2

- Initialize weight vector using Model 1
- Repeat
  - Use latent inference with current weight to "complete" all missing pieces
  - Train with Structured SVM
    - During training, the learning algorithm is penalized more if it makes a mistake on $r(\mathbf{y})$

Generalization of Latent Structure SVM [Yu & Joachims '09]

Penalizing mistakes during learning

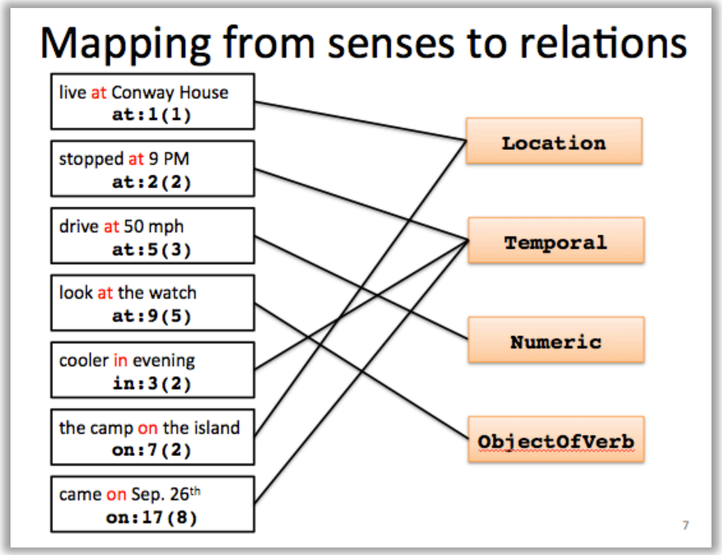
$$\min_{\mathbf{w},\xi} \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_i \xi_i$$

Score of annotated structure $\geq$ Score of any other structure $+\Delta(\mathbf{y}, \mathbf{y}_i) - \xi_i$

- Penalty for making a mistake must not be the same for the labeled and inferred parts
  - Reduced penalty for mistakes made for anything except $r(\mathbf{y})$
  - Modified loss-augmented inference

Coherence with word sense

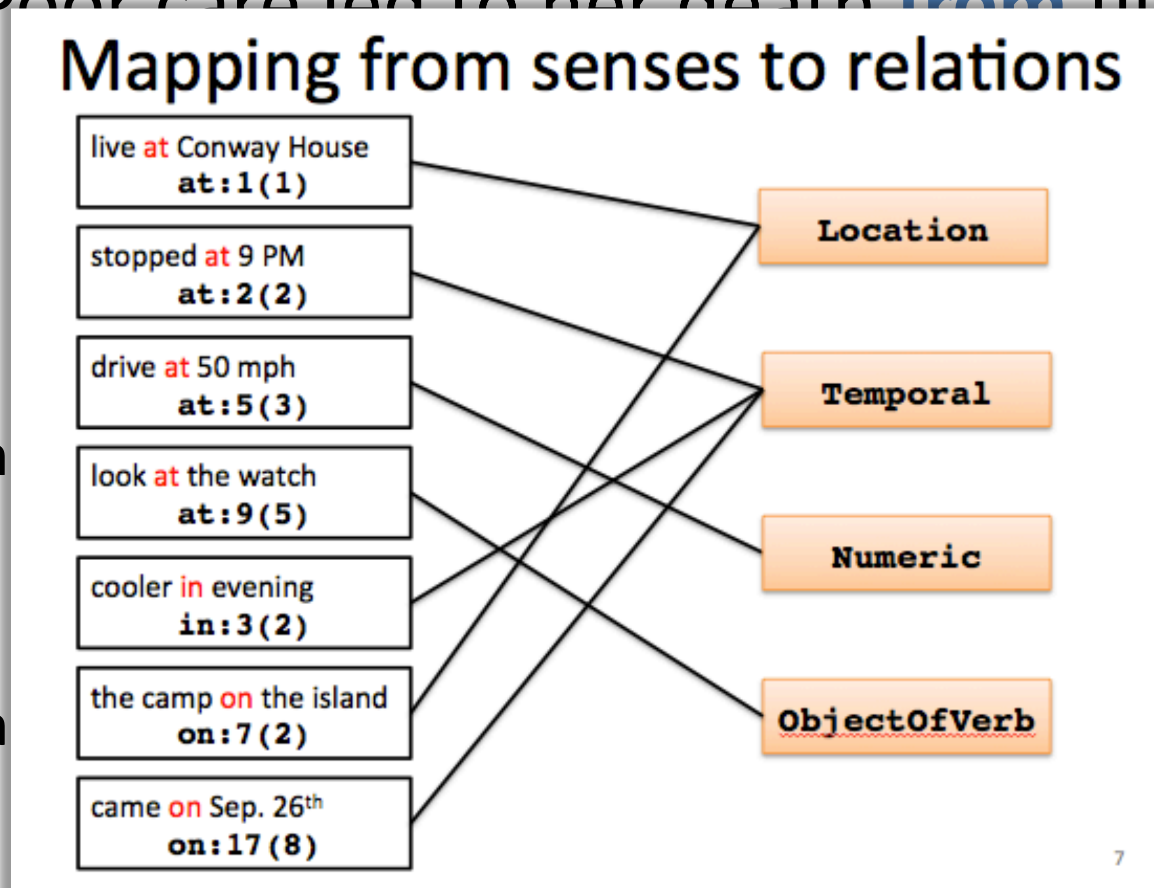
- Each sense corresponds to a relation



- Can explicitly enforce coherence
  – Model 1 or 2 for predicting relations
  – State-of-the-art preposition WSD [Hovy et al, 2010]
  – Coherence constraints from relation definition

Preposition Sense and Relations

Poor care led to her death **from** flu.



[Hovy et al, 2010]

n:12(9)

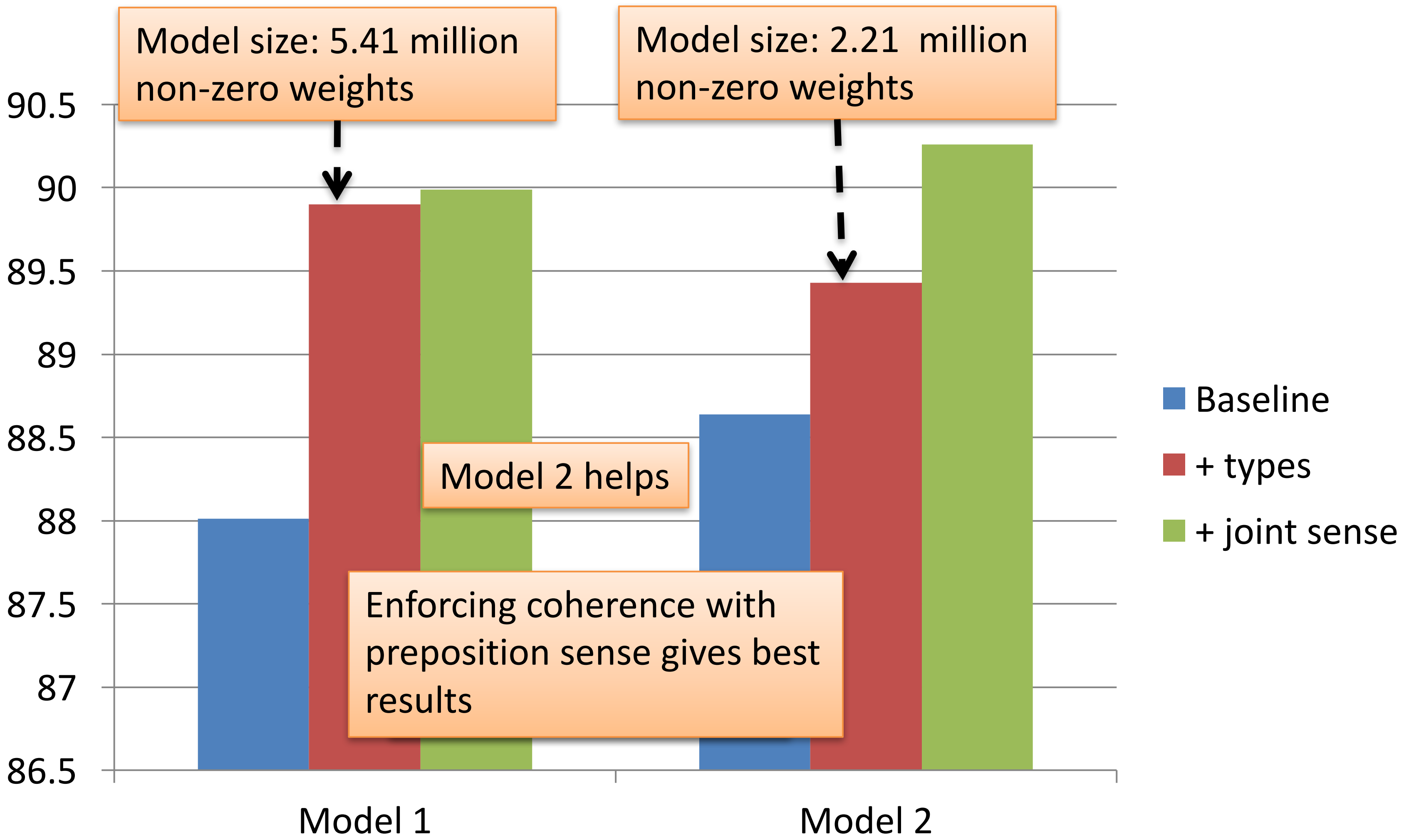
Govern

bject

Govern type

Object type

Accuracy of relation labeling



Analysis

- Sharing features across prepositions helps
  – Possible because the label space is the same
  – Not the case for WSD

- Parser or heuristics alone is not good enough
  – Combining them gives the best performance

Results summary

- Selecting a single assignment to arguments and types helps
- Enforcing coherence with preposition sense also helps
- Not fully believing the parser is useful
- Types are important

What do we have?

Input	Relation	Governor type	Object type
Died of pneumonia	`Cause`	Experience	Disease
Suffering from flu	`Cause`	Experience	Disease
Recovered from flu	`StartState`	Change	Disease

Governor, object and their types as a certificate for the choice of relation label

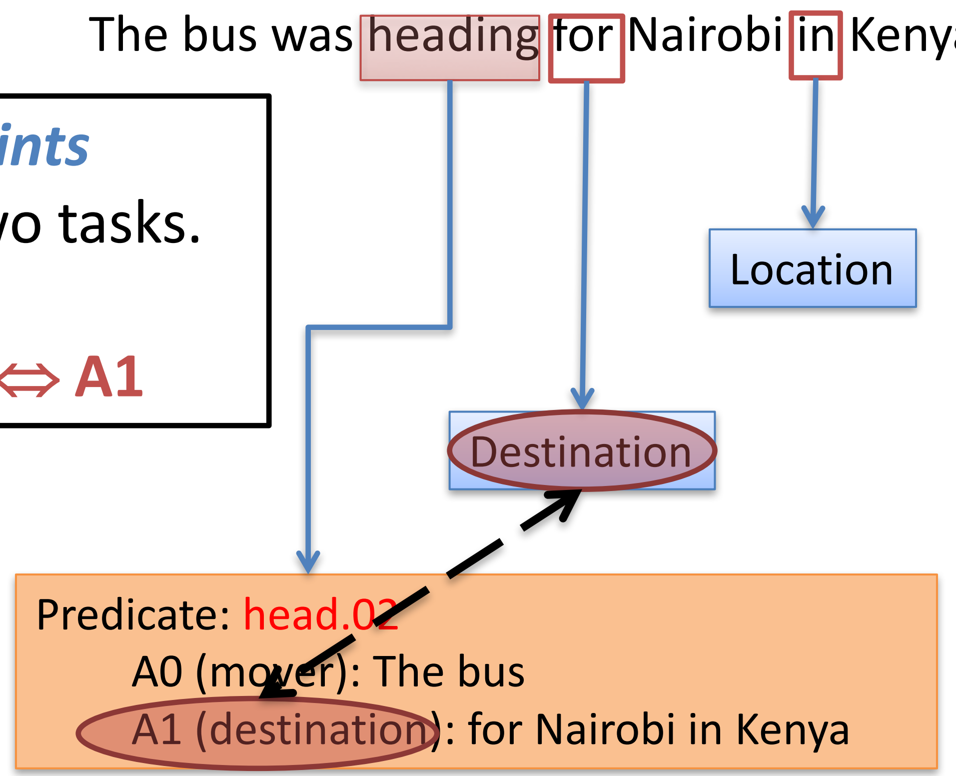
Verb and preposition interplay

Coherence of predictions

Predicate arguments from different triggers should be consistent

Joint constraints
linking the two tasks.

Destination ⇔ A1



Joint Constraints

- Hand written constraints
  1. Each verb attached preposition that is labeled as **Temporal** should correspond to the start of some **AM-TMP**
  2. For verb attached prepositions, some roles should never correspond to a **null** argument label
- Mined constraints
  - Set of allowed verb argument labels for each preposition relation and vice versa

- Constraints written as linear inequalities

Joint inference

Verb arguments

Preposition relations

$$\max_{\mathbf{y}} \sum_t \boxed{\lambda^t} \sum_a y^{a,t} c^{a,t} + \sum_r \boxed{\lambda^r} \sum_p y^{r,p} c^{r,p}$$

Argument candidates

Re-scaling parameters (one per label)

Preposition relation label

Constraints:

Each argument label

Preposition

Verb SRL constraints

Only one label per preposition

Joint constraints (as linear inequalities)

Desiderata for joint prediction

Intuition: *The correct interpretation of a sentence is the one that gives a consistent analysis across all the linguistic phenomena expressed in it*

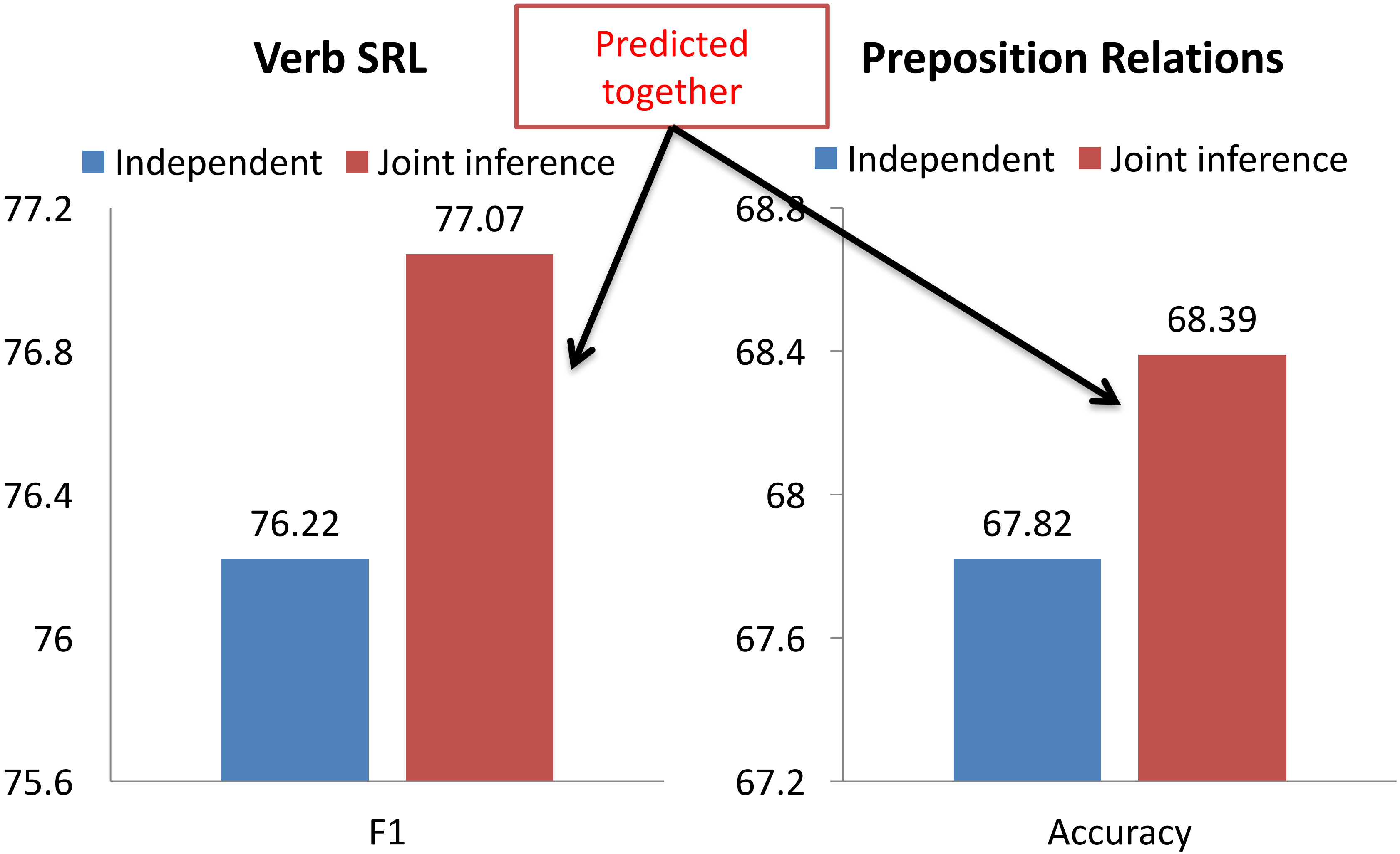
1. Should account for dependencies between linguistic phenomena

Joint constraints between tasks, easy with ILP inference

1. Should be able to use existing state of the art models minimal use of expensive jointly labeled data

Use small amount of joint data to re-scale scores to be in the same numeric range

Joint prediction helps [EMNLP, ‘11]



All results on Penn Treebank Section 23

Closing thoughts

- Semantic Role Labeling
    - Verbs traditionally studied
    - Same model also works for nominalizations

- Extending semantic role labeling
  - Prepositions express a diverse set of relations
    - An ontology of preposition relations
    - Can enrich existing PropBank/FrameNet representation
  - Also some work on comma relations

- *Data, word clusters, software available*
  - Check out the **NLP curator** that packages all of these together