Learning and Inference

for

Natural Language Understanding

Dan Roth

Department of Computer Science

University of Illinois at Urbana-Champaign

September 2015

The 18th International Conference of Text, Speech and Dialogue (TSD 2015)

Plzen, Czech Republic

Cognitive Computation Group
University of Illinois at Urbana-Champaign

Learning and Inference
for
Natural Language Understanding

Dan Roth

Department of Computer Science

University of Illinois at Urbana-Champaign

With thanks to:

Collaborators: **Kai-Wei Chang, Xiao Chen, Dan Goldwasser, Daniel Khashabi, Gourab Kundu, Haoruo Peng, Lev Ratinov, Vivek Srikumar; others**

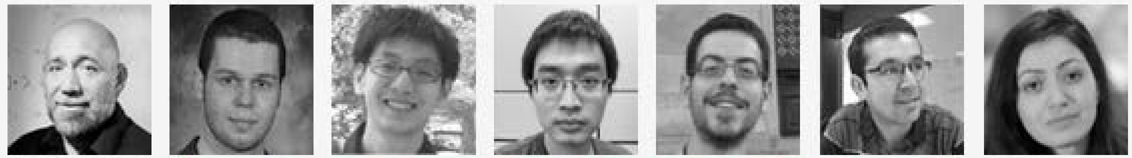
Funding: NSF; DHS; NIH; DARPA; IARPA, ARL, ONR, Google, AI2
DASH Optimization (Xpress-MP); Gurobi.

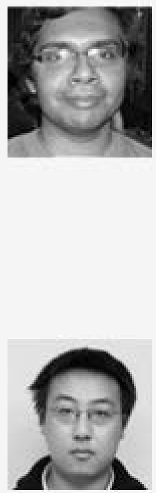
Cognitive Computation Group

University of Illinois at Urbana-Champaign

News | Research | People | Software | Demos | Publications | Resources | Schedule

People





CCG GOALS

Our research focuses on the computational foundations of intelligent behavior. We develop theories and systems pertaining to intelligent behavior using a unified methodology -- at the heart of which is the idea that learning has a central role in intelligence. Specifically, we focus on developing theories and ~~systems~~ for Natural Language Understanding and Information Access, as well as on Machine Learning and Inference approaches that facilitate it.
Our work spans several aspects of these problems -- from theoretical questions in machine learning, knowledge representation and reasoning to experimental paradigms and large scale system development -- and draws on methods from theoretical computer science, probability and statistics, artificial intelligence, linguistics and experimental computer science.

Past Visitors

Wei Lu - Singapore University of Technology and Design
Yao-zhong Zhang - University of Tokyo

Alumni - Ph.D. (First Position)

Vivek Srikumar - Stanford University
Rajhans Samdani - Google Research
Alla Rozovskaya - Columbia University

Comprehension

(ENGLAND, June, 1989) - Christopher Robin is alive and well. He lives in England. He is the same person that you read about in the book, Winnie the Pooh. As a boy, Chris lived in a pretty home called Cotchfield Farm. When Chris was three years old, his father wrote a poem about him. The poem was printed in a magazine for others to read. Mr. Robin then wrote a book. He made up a fairy tale land where Chris lived. His friends were animals. There was a bear called Winnie the Pooh. There was also an owl and a young pig, called a piglet. All the animals were stuffed toys that Chris owned. Mr. Robin made them come to life with his words. The places in the story were all near Cotchfield Farm. Winnie the Pooh was written in 1925. Children still love to read about Christopher Robin and his animal friends. Most people don't know he is a real person who is grown now. He has written two books of his own. They tell what it is like to be famous.

Comprehension

(ENGLAND, June, 1989) - Christopher Robin is alive and well. He lives in England. He is the same person that you read about in the book, Winnie the Pooh. As a boy, Chris lived in a pretty home called Cotchfield Farm. When Chris was three years old, his father wrote a poem about him. The poem was printed in a magazine for others to read. Mr. Robin then wrote a book. He made up a fairy tale land where Chris lived. His friends were animals. There was a bear called Winnie the Pooh. There was also an owl and a young pig, called a piglet. All the animals were stuffed toys that Chris owned. Mr. Robin made them come to life with his words. The places in the story were all near Cotchfield Farm. Winnie the Pooh was written in 1925. Children still love to read about Christopher Robin and his animal friends. Most people don't know he is a real person who is grown now. He has written two books of his own. They tell what it is like to be famous.

1. Christopher Robin was born in England.
2. Winnie the Pooh is a title of a book.
3. Christopher Robin's dad was a magician.
4. Christopher Robin must be at least 65 now.

Comprehension

(ENGLAND, June, 1989) - Christopher Robin is alive and well. He lives in England. He is the same person that you read about in the book, Winnie the Pooh. As a boy, Chris lived in a pretty home called Cotchfield Farm. When Chris was three years old, his father wrote a poem about him. The poem was printed in a magazine for others to read. Mr. Robin then wrote a book. He made up a fairy tale land where Chris lived. His friends were animals. There was a bear called Winnie the Pooh. There was also an owl and a young pig, called a piglet. All the animals were stuffed toys that Chris owned. Mr. Robin made them come to life with his words. The places in the story were all near Cotchfield Farm. Winnie the Pooh was written in 1925. Children still love to read about Christopher Robin and his animal friends. Most people don't know he is a real person who is grown now. He has written two books of his own. They tell what it is like to be famous.

1. Christopher Robin was born in England.
2. Winnie the Pooh is a title of a book.
3. Christopher Robin's dad was a magician.
4. Christopher Robin must be at least 65 now.

Comprehension

(ENGLAND, June, 1989) - Christopher Robin is alive and well. He lives in England. He is the same person that you read about in the book, Winnie the Pooh. As a boy, Chris lived in a pretty home called Cotchfield Farm. When Chris was three years old, his father wrote a poem about him. The poem was printed in a magazine for others to read. Mr. Robin then wrote a book. He made up a fairy tale land where Chris lived. His friends were animals. There was a bear called Winnie the Pooh. There was also an owl and a young pig, called a piglet. All the animals were stuffed toys that Chris owned. Mr. Robin made them come to life with his words. The places in the story were all near Cotchfield Farm. Winnie the Pooh was written in 1925. Children still love to read about Christopher Robin and his animal friends. Most people don't know he is a real person who is grown now. He has written two books of his own. They tell what it is like to be famous.

1. Christopher Robin was born in England.
2. Winnie the Pooh is a title of a book.
3. Christopher Robin's dad was a magician.
4. Christopher Robin must be at least 65 now.

Comprehension

(ENGLAND, June, 1989) - Christopher Robin is alive and well. He lives in England. He is the same person that you read about in the book, Winnie the Pooh. As a boy, Chris lived in a pretty home called Cotchfield Farm. When Chris was three years old, his father wrote a poem about him. The poem was printed in a magazine for others to read. Mr. Robin then wrote a book. He made up a fairy tale land where Chris lived. His friends were animals. There was a bear called Winnie the Pooh. There was also an owl and a young pig, called a piglet. All the animals were stuffed toys that Chris owned. Mr. Robin made them come to life with his words. The places in the story were all near Cotchfield Farm. Winnie the Pooh was written in 1925. Children still love to read about Christopher Robin and his animal friends. Most people don't know he is a real person who is grown now. He has written two books of his own. They tell what it is like to be famous.

1. Christopher Robin was born in England.
2. Winnie the Pooh is a title of a book.
3. Christopher Robin's dad was a magician.
4. Christopher Robin must be at least 65 now.

Comprehension

(ENGLAND, June, 1989) - Christopher Robin is alive and well. He lives in England. He is the same person that you read about in the book, Winnie the Pooh. As a boy, Chris lived in a pretty home called Cotchfield Farm. When Chris was three years old, his father wrote a poem about him. The poem was printed in a magazine for others to read. Mr. Robin then wrote a book. He made up a fairy tale land where Chris lived. His friends were animals. There was a bear called Winnie the Pooh. There was also an owl and a young pig, called a piglet. All the animals were stuffed toys that Chris owned. Mr. Robin made them come to life with his words. The places in the story were all near Cotchfield Farm. Winnie the Pooh was written in 1925. Children still love to read about Christopher Robin and his animal friends. Most people don't know he is a real person who is grown now. He has written two books of his own. They tell what it is like to be famous.

1. Christopher Robin was born in England.
2. Winnie the Pooh is a title of a book.
3. Christopher Robin's dad was a magician.
4. Christopher Robin must be at least 65 now.

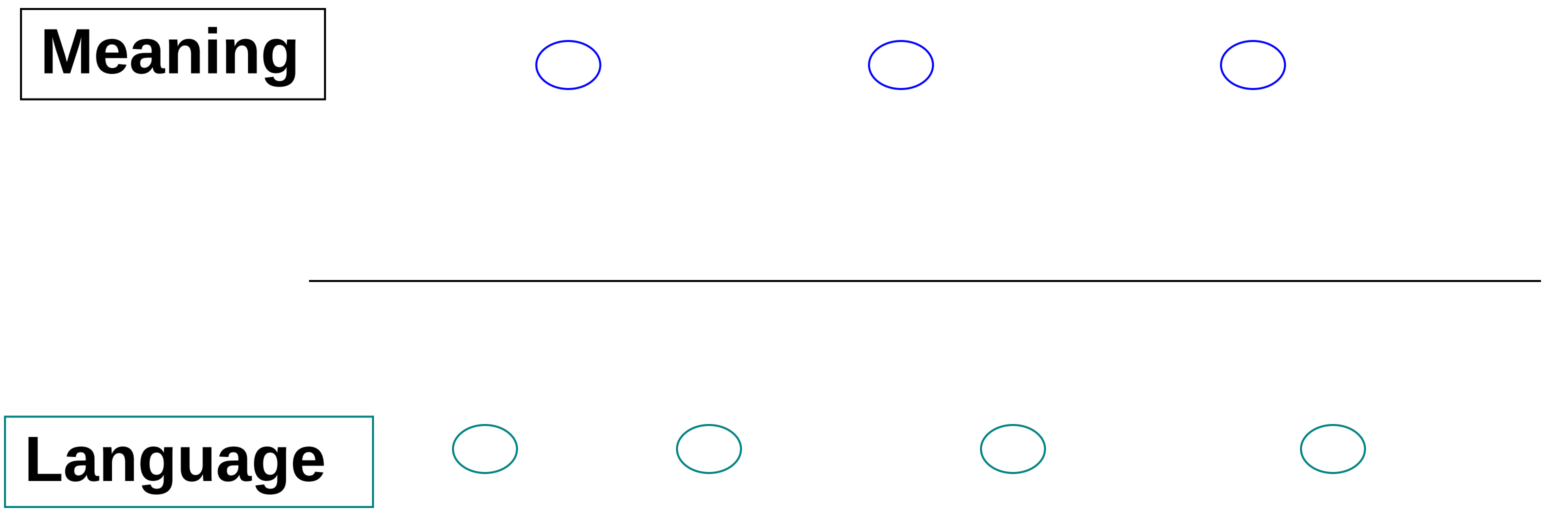
Comprehension

(ENGLAND, June, 1989) - Christopher Robin is alive and well. He lives in England. He is the same person that you read about in the book, Winnie the Pooh. As a boy, Chris lived in a pretty home called Cotchfield Farm. When Chris was three years old, his father wrote a poem about him. The poem was printed in a magazine for others to read. Mr. Robin then wrote a book. He made up a fairy tale land where Chris lived. His friends were animals. There was a bear called Winnie the Pooh. There was also an owl and a young pig, called a piglet. All the animals were stuffed toys that Chris owned. Mr. Robin made them come to life with his words. The places in the story were all near Cotchfield Farm. Winnie the Pooh was written in 1925. Children still love to read about Christopher Robin and his animal friends. Most people don't know he is a real person who is grown now. He has written two books of his own. They tell what it is like to be famous.

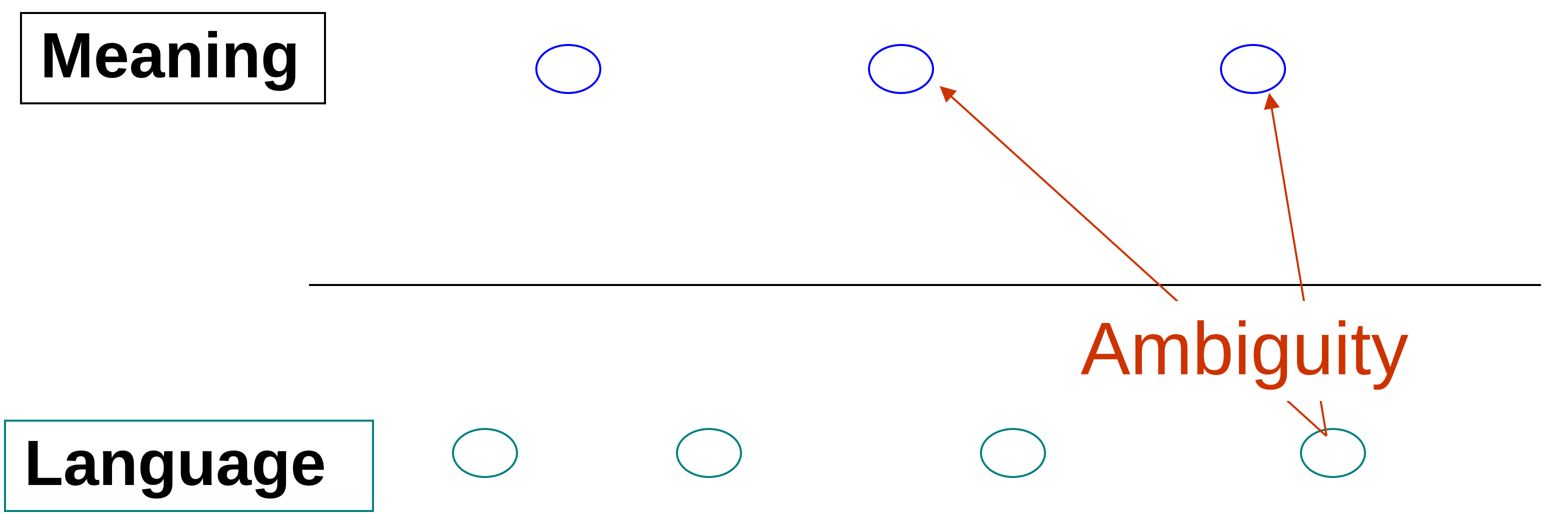
1. Christopher Robin was born in England.
2. Winnie the Pooh is a title of a book.
3. Christopher Robin's dad was a magician.
4. Christopher Robin must be at least 65 now.

This is an Inference Problem

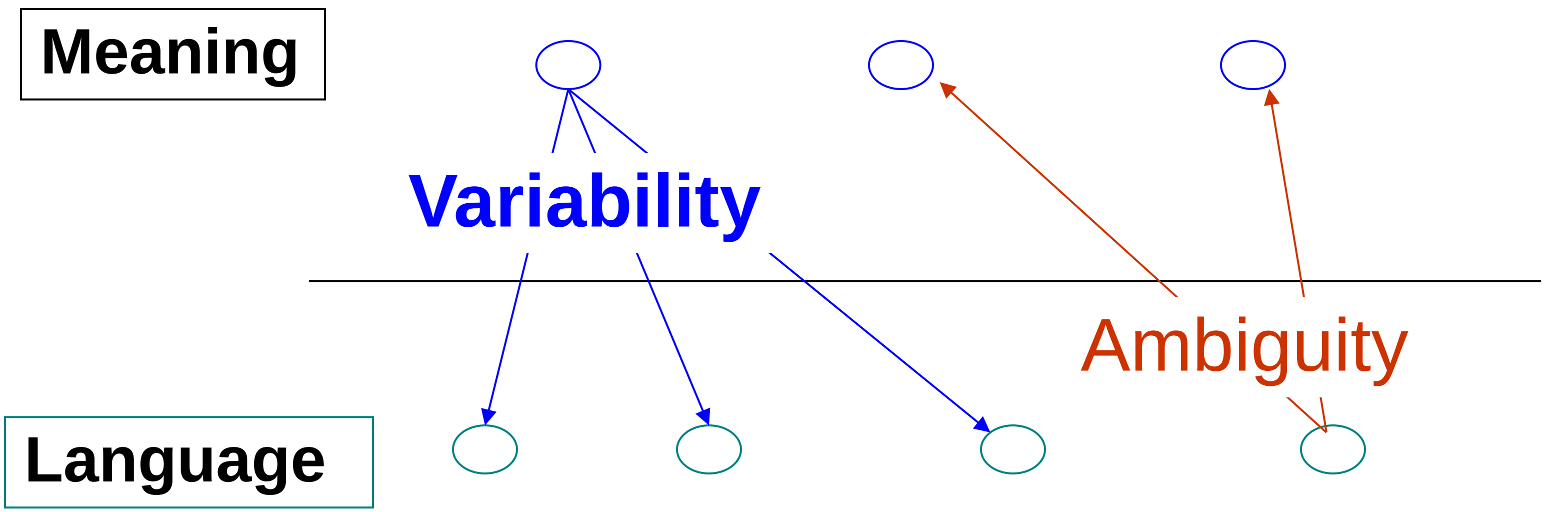
Why is it Difficult?



Why is it Difficult?



Why is it Difficult?



Ambiguity

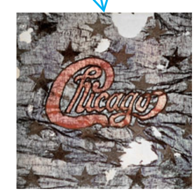
It's a version of ***Chicago*** – the standard classic ***Macintosh*** menu font, with that distinctive thick diagonal in the "N".	***Chicago*** was used by default for ***Mac*** menus through ***MacOS 7.6***, and ***OS 8*** was released mid-1997..	***Chicago VIII*** was one of the early 70s-era ***Chicago*** albums to catch my ear, along with ***Chicago II***.

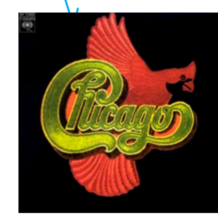
Ambiguity

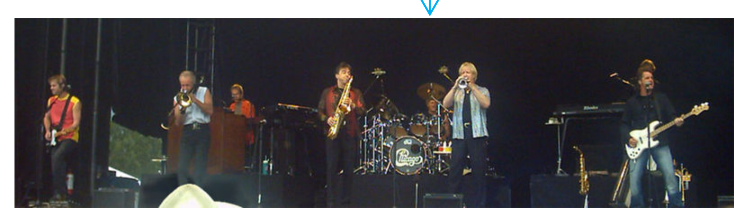
It's a version of ***Chicago*** – the standard classic ***Macintosh*** menu font, with that distinctive thick diagonal in the "N".

Chicago was used by default for ***Mac*** menus through ***MacOS 7.6***, and ***OS 8*** was released mid-1997..

Chicago VIII was one of the early 70s-era ***Chicago*** albums to catch my ear, along with ***Chicago II***.

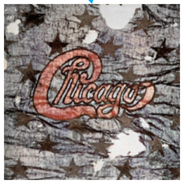






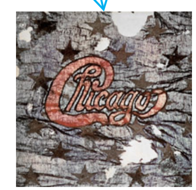
Ambiguity

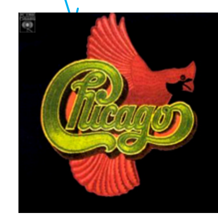
It's a version of ***Chicago*** – the standard classic ***Macintosh*** menu font, with that distinctive thick diagonal in the "N".	***Chicago*** was used by default for ***Mac*** menus through ***MacOS 7.6***, and ***OS 8*** was released mid-1997..	***Chicago VIII*** was one of the early 70s-era ***Chicago*** albums to catch my ear, along with ***Chicago II***.

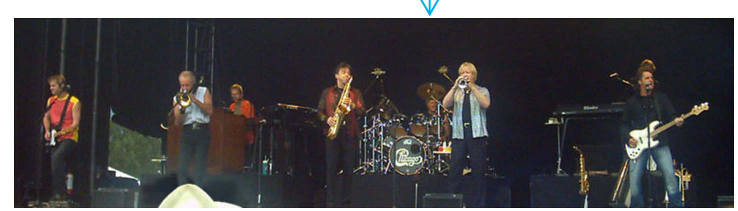


Ambiguity

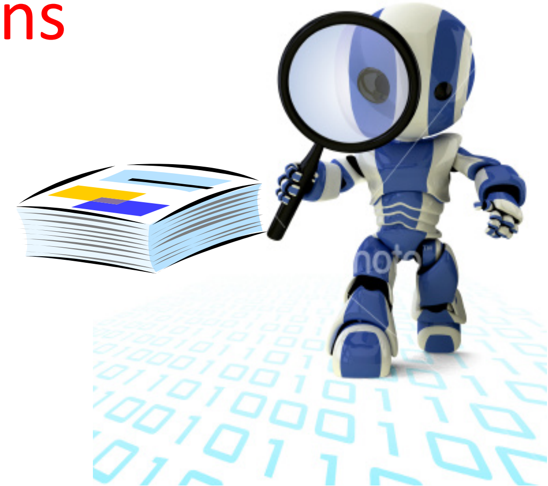
It's a version of ***Chicago*** – the standard classic ***Macintosh*** menu font, with that distinctive thick diagonal in the "N".	***Chicago*** was used by default for ***Mac*** menus through ***MacOS 7.6***, and ***OS 8*** was released mid-1997..	***Chicago VIII*** was one of the early 70s-era ***Chicago*** albums to catch my ear, along with ***Chicago II***.





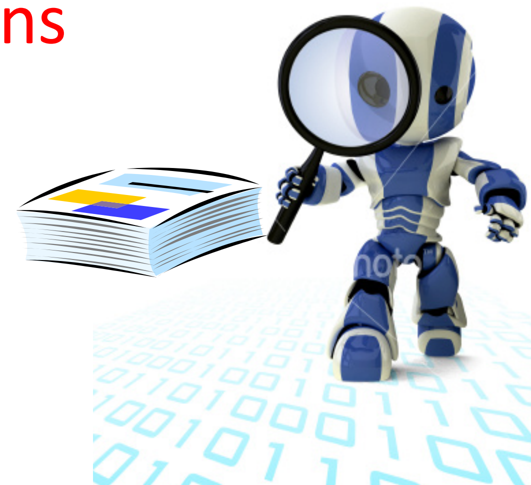


Variability in Natural Language Expressions

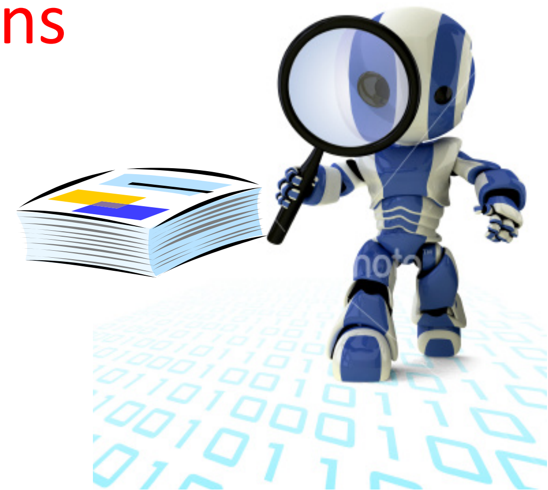


Variability in Natural Language Expressions

Determine if Jim Carpenter works for the government



Variability in Natural Language Expressions



Determine if Jim Carpenter works for the government

Jim Carpenter works for the U.S. Government.

Variability in Natural Language Expressions

Determine if Jim Carpenter works for the government

Jim Carpenter works for the U.S. Government.

The American government employed Jim Carpenter.

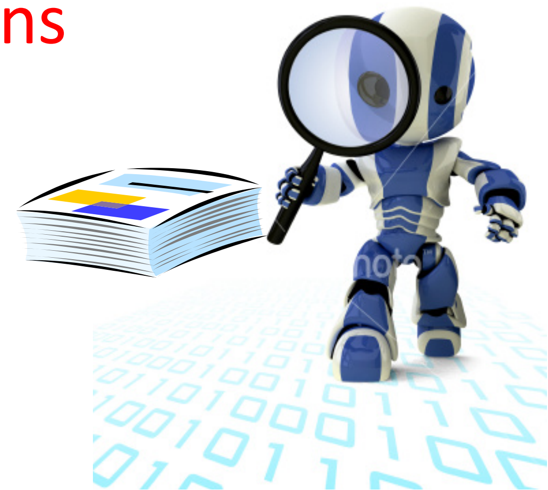
Jim Carpenter was fired by the US Government.

Jim Carpenter worked in a number of important positions.
…. As a press liaison for the IRS, he made contacts in the white house.

Russian interior minister Yevgeny Topolov met yesterday with his US counterpart, Jim Carpenter.

Former US Secretary of Defense Jim Carpenter spoke today…

Variability in Natural Language Expressions



- Determine if Jim Carpenter works for the government

Jim Carpenter works for the U.S. Government.

The American government employed Jim Carpenter.

Jim Carpenter was fired by the US Government.

Jim Carpenter worked in a number of important positions. …. As a press liaison for the IRS, he made contacts in the white house.

- Russian interior minister Yevgeny Topolov met yesterday with his US counterpart, Jim Carpenter.

Former US Secretary of Defense Jim Carpenter spoke today…

Cognitive Computation Group
University of Illinois at Urbana-Champaign

Variability in Natural Language Expressions

Determine if Jim Carpenter works for the government

Jim Carpenter works for the U.S. Government.

The American government employed Jim Carpenter.

Jim Carpenter was fired by the US Government.

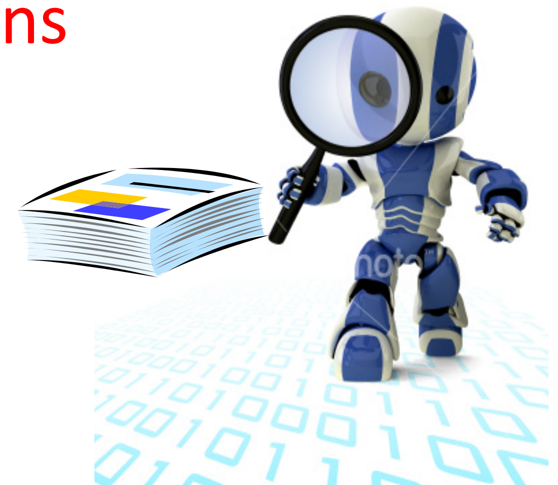
Jim Carpenter worked in a number of important positions.
…. As a press liaison for the IRS, he made contacts in the white house.

Russian interior minister Yevgeny Topolov met yesterday with his US counterpart, Jim Carpenter.

Former US Secretary of Defense Jim Carpenter spoke today…

Standard techniques cannot deal with the variability of expressing meaning nor with the ambiguity of interpretation

Variability in Natural Language Expressions



Determine if Jim Carpenter works for the government

Jim Carpenter works for the U.S. Government.

The American government employed Jim Carpenter.

Jim Carpenter was fired by the US Government.

Jim Carpenter worked in a number of important positions.
…. As a press liaison for the IRS, he made contacts in the white house.

Russian interior minister Yevgeny Topolov met yesterday with his US counterpart, Jim Carpenter.

Former US Secretary of Defense Jim Carpenter spoke today…

Standard techniques cannot deal with the variability of expressing meaning nor with the ambiguity of interpretation

Needs:

- Understanding Relations, Entities and Semantic Classes
- Acquiring knowledge from external resources; representing knowledge
- Identifying, disambiguating & tracking entities, events, etc.
- Time, quantities, processes…

What we do

- Understanding natural language
  - Lexical inference
  - Semantic Parsing
  - Discourse phenomena
  - .....
- Understanding a lot of domains
  - Events
  - Temporal Reasoning
  - Spatial Reasoning
  - Quantitative Reasoning
  - ....

What we do

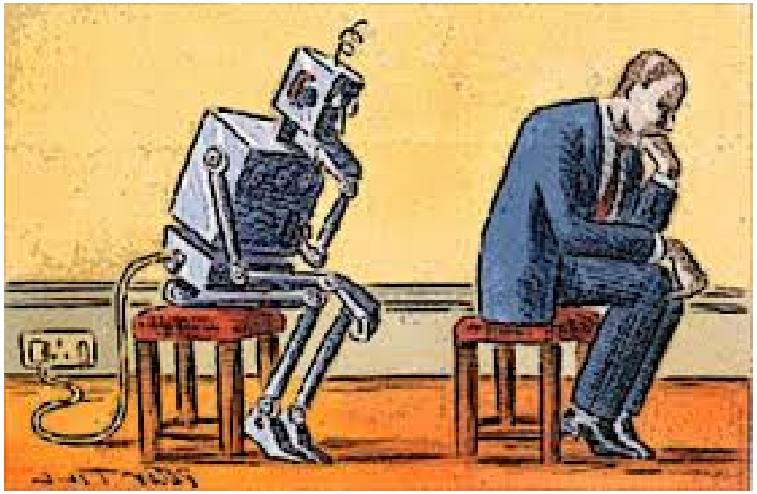
- Understanding natural language
  - Lexical inference
  - Semantic Parsing
  - Discourse phenomena
  - …..
- Understanding a lot of domains
  - Events
  - Temporal Reasoning
  - Spatial Reasoning
  - Quantitative Reasoning
  - …….
- Knowledge
  - Acquisition
  - Representation
  - ….

What we do

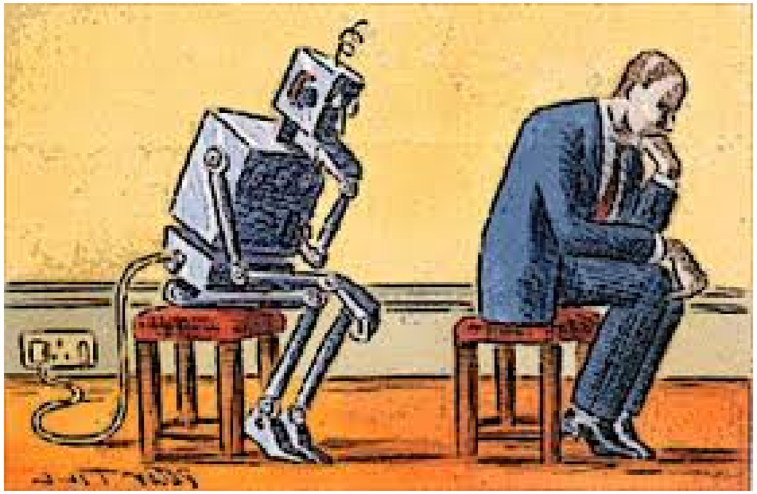
- Understanding natural language
  - Lexical inference
  - Semantic Parsing
  - Discourse phenomena
  - .....
- Understanding a lot of domains
  - Events
  - Temporal Reasoning
  - Spatial Reasoning
  - Quantitative Reasoning
  - ....
- Knowledge
  - Acquisition
  - Representation
  - ....
- How to use knowledge to support textual inference?

What is Needed?

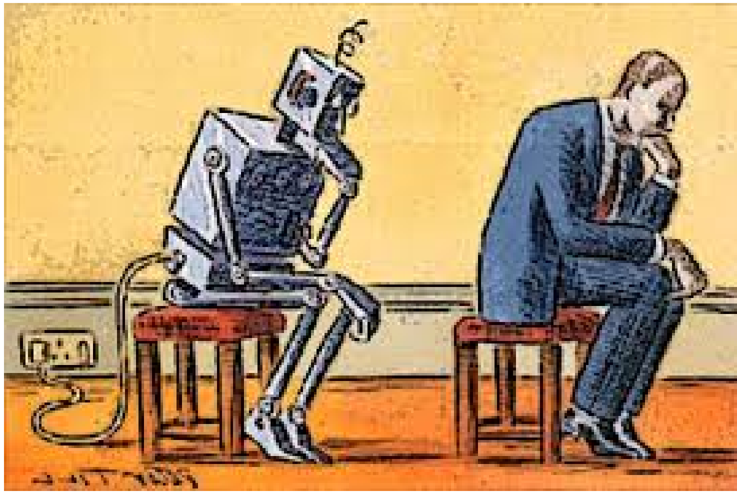
What is Needed?



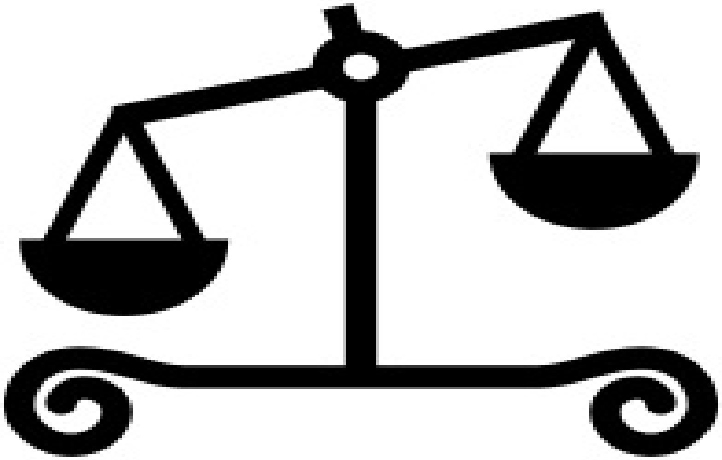
What is Needed?

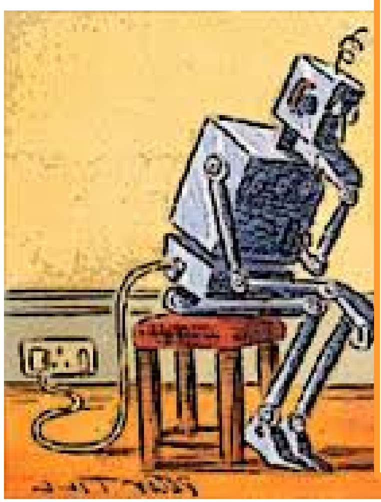


What is Needed?



What is Needed?





- A computational Framework
- Examples:
  - Modeling
  - Learning
  - Inference

Comprehension

(ENGLAND, June, 1989) - Christopher Robin is alive and well. He lives in England. He is the same person that you read about in the book, Winnie the Pooh. As a boy, Chris lived in a pretty home called Cotchfield Farm. When Chris was three years old, his father wrote a poem about him. The poem was printed in a magazine for others to read. Mr. Robin then wrote a book. He made up a fairy tale land where Chris lived. His friends were animals. There was a bear called Winnie the Pooh. There was also an owl and a young pig, called a piglet. All the animals were stuffed toys that Chris owned. Mr. Robin made them come to life with his words. The places in the story were all near Cotchfield Farm. Winnie the Pooh was written in 1925. Children still love to read about Christopher Robin and his animal friends. Most people don't know he is a real person who is grown now. He has written two books of his own. They tell what it is like to be famous.

1. Christopher Robin was born in England.
2. Winnie the Pooh is a title of a book.
3. Christopher Robin's dad was a magician.
4. Christopher Robin must be at least 65 now.

This is an Inference Problem

Comprehension

(ENGLAND, June, 1989) - Christopher Robin is alive and well. He lives in England. He is the same person that you read about in the book, Winnie the Pooh. As a boy, Chris lived in a pretty home called Cotchfield Farm. When Chris was three years old, his father wrote a poem about him. The poem was printed in a magazine for others to read. Mr. Robin then wrote a book. He made up a fairy tale land where Chris lived. His friends were animals. There was a bear called Winnie the Pooh. There was also an owl and a young pig, called a piglet. All the animals were stuffed toys that Chris owned. Mr. Robin made them come to life with his words. The places in the story were all near Cotchfield Farm. Winnie the Pooh was written in 1925. Children still love to read about Christopher Robin and his animal friends. Most people don't know he is a real person who is grown now. He has written two books of his own. They tell what it is like to be famous

1. Christopher Robin was born in England.
2. Winnie the Pooh is a title of a book.
3. Christopher Robin's dad was a magician.
4. Christopher Robin must be at least 65 now.

This is an Inference Problem

Comprehension

(ENGLAND, June, 1989) - Christopher Robin is alive and well. He lives in England. He is the same person that you read about in the book, Winnie the Pooh. As a boy, Chris lived in a pretty home called Cotchfield Farm. When Chris was three years old, his father wrote a poem about him. The poem was printed in a magazine for others to read. Mr. Robin then wrote a book. He made up a fairy tale land where Chris lived. His friends were animals. There was a bear called Winnie the Pooh. There was also an owl and a young pig, called a piglet. All the animals were stuffed toys that Chris owned. Mr. Robin made them come to life with his words. The places in the story were all near Cotchfield Farm. Winnie the Pooh was written in 1925. Children still love to read about Christopher Robin and his animal friends. Most people don't know he is a real person who is grown now. He has written two books of his own. They tell what it is like to be famous.

1. Christopher Robin was born in England.
2. Winnie the Pooh is a title of a book.
3. Christopher Robin's dad was a magician.
4. Christopher Robin must be at least 65 now.

This is an Inference Problem

Comprehension

(ENGLAND, June, 1989) - Christopher Robin is alive and well. He lives in England. He is the same person that you read about in the book, Winnie the Pooh. As a boy, Chris lived in a pretty home called Cotchfield Farm. When Chris was three years old, his father wrote a poem about him. The poem was printed in a magazine for others to read. Mr. Robin then wrote a book. He made up a fairy tale land where Chris lived. His friends were animals. There was a bear called Winnie the Pooh. There was also an owl and a young pig, called a piglet. All the animals were stuffed toys that Chris owned. Mr. Robin made them come to life with his words. The places in the story were all near Cotchfield Farm. Winnie the Pooh was written in 1925. Children still love to read about Christopher Robin and his animal friends. Most people don't know he is a real person who is grown now. He has written two books of his own. They tell what it is like to be famous.

1. Christopher Robin was born in England.
2. Winnie the Pooh is a title of a book.
3. Christopher Robin's dad was a magician.
4. Christopher Robin must be at least 65 now.

This is an Inference Problem

Comprehension

(ENGLAND, June, 1989) - Christopher Robin is alive and well. He lives in England. He is the same person that you read about in the book, Winnie the Pooh. As a boy, Chris lived in a pretty home called Cotchfield Farm. When Chris was three years old, his father wrote a poem about him. The poem was printed in a magazine for others to read. Mr. Robin then wrote a book. He made up a fairy tale land where Chris lived. His friends were animals. There was a bear called Winnie the Pooh. There was also an owl and a young pig, called a piglet. All the animals were stuffed toys that Chris owned. Mr. Robin made them come to life with his words. The places in the story were all near Cotchfield Farm. Winnie the Pooh was written in 1925. Children still love to read about Christopher Robin and his animal friends. Most people don't know he is a real person who is grown now. He has written two books of his own. They tell what it is like to be famous.

1. Christopher Robin was born in England.
2. Winnie the Pooh is a title of a book.
3. Christopher Robin's dad was a magician.
4. Christopher Robin must be at least 65 now.

This is an Inference Problem

Comprehension

(ENGLAND, June, 1989) - Christopher Robin is alive and well. He lives in England. He is the same person that you read about in the book, Winnie the Pooh. As a boy, Chris lived in a pretty home called Cotchfield Farm. When Chris was three years old, his father wrote a poem about him. The poem was printed in a magazine for others to read. Mr. Robin then wrote a book. He made up a fairy tale land where Chris lived. His friends were animals. There was a bear called Winnie the Pooh. There was also an owl and a young pig, called a piglet. All the animals were stuffed toys that Chris owned. Mr. Robin made them come to life with his words. The places in the story were all near Cotchfield Farm. Winnie the Pooh was written in 1925. Children still love to read about Christopher Robin and his animal friends. Most people don't know he is a real person who is grown now. He has written two books of his own. They tell what it is like to be famous.

1. Christopher Robin was born in England.
2. Winnie the Pooh is a title of a book.
3. Christopher Robin's dad was a magician.
4. Christopher Robin must be at least 65 now.

This is an Inference Problem

Comprehension

(ENGLAND, June, 1989) - Christopher Robin is alive and well. He lives in England. He is the same person that you read about in the book, Winnie the Pooh. As a boy, Chris lived in a pretty home called Cotchfield Farm. When Chris was three years old, his father wrote a poem about him. The poem was printed in a magazine for others to read. Mr. Robin then wrote a book. He made up a fairy tale land where Chris lived. His friends were animals. There was a bear called Winnie the Pooh. There was also an owl and a young pig, called a piglet. All the animals were stuffed toys that Chris owned. Mr. Robin made them come to life with his words. The places in the story were all near Cotchfield Farm. Winnie the Pooh was written in 1925. Children still love to read about Christopher Robin and his animal friends. Most people don't know he is a real person who is grown now. He has written two books of his own. They tell what it is like to be famous.

1. Christopher Robin was born in England.
2. Winnie the Pooh is a title of a book.
3. Christopher Robin's dad was a magician.
4. Christopher Robin must be at least 65 now.

This is an Inference Problem

Natural Language Understanding

- Natural language understanding decisions are global decisions that require
  - Making (local) predictions driven by different models trained in different ways, at different times/conditions/scenarios
  - The ability to put these predictions together coherently
  - Knowledge, that guides the decisions so they satisfy our expectations

Natural Language Understanding

Expectation is a knowledge intensive component

- Natural language understanding decisions are global decisions that require
  - Making (local) predictions driven by different models trained in different ways, at different times/conditions/scenarios
  - The ability to put these predictions together coherently
  - Knowledge, that guides the decisions so they satisfy our expectations

Natural Language Understanding

Expectation is a knowledge intensive component

- Natural language understanding decisions are global decisions that require
  - Making (local) predictions driven by different models trained in different ways, at different times/conditions/scenarios
  - The ability to put these predictions together coherently
  - Knowledge, that guides the decisions so they satisfy our expectations

Natural Language Understanding

Expectation is a knowledge intensive component

- Natural language understanding decisions are global decisions that require
  - Making (local) predictions driven by different models trained in different ways, at different times/conditions/scenarios
  - The ability to put these predictions together coherently
  - Knowledge, that guides the decisions so they satisfy our expectations

Natural Language Interpretation is a Common Sense driven Inference Process that is best thought of as a knowledge constrained optimization problem, done on top of multiple statistically learned models.

Natural Language Understanding

Expectation is a knowledge intensive component

- Natural language understanding decisions are global decisions that require
  - Making (local) predictions driven by different models trained in different ways, at different times/conditions/scenarios
  - The ability to put these predictions together coherently
  - Knowledge, that guides the decisions so they satisfy our expectations

Natural Language Interpretation is a Common Sense driven Inference Process that is best thought of as a knowledge constrained optimization problem, done on top of multiple statistically learned models.

Many forms of Inference; a lot boil down to determining best assignment

Natural Language Understanding

Expectation is a knowledge intensive component

- Natural language understanding decisions are global decisions that require
  - Making (local) predictions driven by different models trained in different ways, at different times/conditions/scenarios
  - The ability to put these predictions together coherently
  - Knowledge, that guides the decisions so they satisfy our expectations

Natural Language Interpretation is a Common Sense driven Inference Process that is best thought of as a knowledge constrained optimization problem, done on top of multiple statistically learned models.

- **How can we get there?**

Many forms of Inference; a lot boil down to determining best assignment

Joint Inference with General Constraint Structure [Roth&Yih'04,07,….]

Recognizing Entities and Relations

$\boxed{\text{Dole}}$'s wife, $\boxed{\text{Elizabeth}}$, is a native of $\boxed{\text{N.C.}}$

$E1 \xrightarrow{R_{12}} E2 \xrightarrow{R_{23}} E3$

Joint Inference with General Constraint Structure [Roth&Yih'04,07,….]

Recognizing Entities and Relations

other	0.05
per	0.85
loc	0.10

other	0.10
per	0.60
loc	0.30

other	0.05
per	0.50
loc	0.45

Dole 's wife, Elizabeth , is a native of N.C.

E_1 E_2 E_3

R_{12} R_{23}

irrelevant	0.05
spouse_of	0.45
born_in	0.50

irrelevant	0.10
spouse_of	0.05
born_in	0.85

Joint Inference with General Constraint Structure [Roth&Yih'04,07,….]

Recognizing Entities and Relations

other	0.05
per	**0.85**
loc	0.10

other	0.10
per	**0.60**
loc	0.30

other	0.05
per	**0.50**
loc	0.45

Dole 's wife, Elizabeth , is a native of N.C.

$$E_1 \xrightarrow{R_{12}} E_2 \xrightarrow{R_{23}} E_3$$

irrelevant	0.05
spouse_of	0.45
born_in	**0.50**

irrelevant	0.10
spouse_of	0.05
born_in	**0.85**

Joint Inference with General Constraint Structure [Roth&Yih'04,07,….]

Recognizing Entities and Relations

other	0.05
per	**0.85**
loc	0.10

other	0.10
per	**0.60**
loc	0.30

other	0.05
per	**0.50**
loc	0.45

Dole 's wife, Elizabeth , is a native of N.C.

$E_1 \xrightarrow{R_{12}} E_2 \xrightarrow{R_{23}} E_3$

irrelevant	0.05
spouse_of	0.45
born_in	**0.50**

irrelevant	0.10
spouse_of	0.05
born_in	**0.85**

Joint Inference with General Constraint Structure [Roth&Yih'04,07,….]

Recognizing Entities and Relations

other	0.05
per	**0.85**
loc	0.10

other	0.10
per	**0.60**
loc	0.30

other	0.05
per	0.50
loc	**0.45**

Dole 's wife, Elizabeth , is a native of N.C.

E_1 $\xrightarrow{R_{12}}$ E_2 $\xrightarrow{R_{23}}$ E_3

irrelevant	0.05
spouse_of	0.45
born_in	**0.50**

irrelevant	0.10
spouse_of	0.05
born_in	**0.85**

Joint Inference with General Constraint Structure [Roth&Yih'04,07,….]

Recognizing Entities and Relations

other	0.05
per	**0.85**
loc	0.10

other	0.10
per	**0.60**
loc	0.30

other	0.05
per	0.50
loc	**0.45**

Dole 's wife, Elizabeth , is a native of N.C.

E1 E2 E3

R_{12} R_{23}

irrelevant	0.05
spouse_of	0.45
born_in	**0.50**

irrelevant	0.10
spouse_of	0.05
born_in	**0.85**

Joint Inference with General Constraint Structure [Roth&Yih'04,07,….]

Recognizing Entities and Relations

other	0.05
per	**0.85**
loc	0.10

other	0.10
per	**0.60**
loc	0.30

other	0.05
per	0.50
loc	**0.45**

Dole 's wife, Elizabeth , is a native of N.C.

E1 E2 E3

R_{12} R_{23}

irrelevant	0.05
spouse_of	**0.45**
born_in	0.50

irrelevant	0.10
spouse_of	0.05
born_in	**0.85**

Joint Inference with General Constraint Structure [R

Recognizing Entities and Relations

Joint inference gives good improvement

other	0.05
per	**0.85**
loc	0.10

other	0.10
per	**0.60**
loc	0.30

othe	0.05
p	0.50
loc	**0.45**

Dole 's wife, Elizabeth , is a native of N.C.

E_1 E_2 E_3

R_{12} R_{23}

irrelevant	0.05
spouse_of	**0.45**
born_in	0.50

irrelevant	0.10
spouse_of	0.05
born_in	**0.85**

Joint Inference with General Constraint Structure [R

Recognizing Entities and Relations

Joint inference gives good improvement

other	0.05
per	**0.85**
loc	0.10

other	0.10
per	**0.60**
loc	0.30

othe	0.05
p	0.50
loc	**0.45**

Dole 's wife, Elizabeth , is a na

E1 E2

R_{12} R_{23}

Key Questions:

How to guide the global inference?

How to learn the model(s)?

irrelevant	0.05
spouse_of	**0.45**
born_in	0.50

irrelevant	0.10
spouse_of	0.05
born_in	**0.85**

Joint Inference with General Constraint Structure [R

Recognizing Entities and Relations

Joint inference gives good improvement

other	0.05
per	**0.85**
loc	0.10

other	0.10
per	**0.60**
loc	0.30

othe	0.05
[illegible]	0.50
loc	**0.45**

Dole 's wife, Elizabeth , is a na

E1 $\xrightarrow{R_{12}}$ E2 — R_{23}

Key Questions:

How to guide the global inference?

How to learn the model(s)?

irrelevant	0.05
spouse_of	**0.45**
born_in	0.50

irrelevant	0.10
spouse_of	0.05
born_in	**0.85**

Models could be learned separately/jointly; constraints may come up only at decision time.

Joint Inference with General Constraint Structure [R

Recognizing Entities and Relations

Joint inference gives good improvement

other	0.05
per	**0.85**
	0.10

other	0.10
per	**0.60**
loc	0.30

othe	0.05
p	0.50
loc	**0.45**

Key Questions:

global inference?

An Objective function that incorporates learned models with knowledge (output constraints)

A Constrained Conditional Model

irrelevant	0.05
spouse_of	**0.45**
born_in	0.50

spouse_of	0.05
born_in	**0.85**

Models could be learned separately/jointly; constraints may come up only at decision time.

Constrained Conditional Models

$$\mathbf{y} = \mathbf{argmax}_{\mathbf{y} \in \mathcal{Y}}\ \mathbf{w}^{\mathbf{T}}\phi(\mathbf{x}, \mathbf{y})$$

Constrained Conditional Models

$$\mathbf{y} = \mathbf{argmax}_{\mathbf{y} \in \mathcal{Y}} \; \mathbf{w}^{\mathbf{T}}\phi(\mathbf{x}, \mathbf{y})$$

Constrained Conditional Models

$$\mathbf{y} = \mathbf{argmax}_{\mathbf{y} \in \mathcal{Y}} \ \mathbf{w}^{\mathbf{T}}\phi(\mathbf{x}, \mathbf{y})$$

Features, classifiers; log-linear models (HMM, CRF) or a combination

Constrained Conditional Models

$$\mathbf{y} = \mathbf{argmax}_{\mathbf{y} \in \mathcal{Y}}\ \mathbf{w}^{\mathbf{T}}\phi(\mathbf{x}, \mathbf{y})$$

Weight Vector for "local" models

Features, classifiers; log-linear models (HMM, CRF) or a combination

Constrained Conditional Models

$$\mathbf{y} = \mathrm{argmax}_{\mathbf{y} \in \mathcal{Y}}\ \mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}, \mathbf{y}) + \mathbf{u}^{\mathrm{T}}\mathcal{C}(\mathbf{x}, \mathbf{y})$$

Knowledge component: (Soft) constraints

Weight Vector for "local" models

Features, classifiers; log-linear models (HMM, CRF) or a combination

Constrained Conditional Models

Penalty for violating the constraint.

$$\mathbf{y} = \mathbf{argmax}_{\mathbf{y} \in \mathcal{Y}} \ \mathbf{w}^T \phi(\mathbf{x}, \mathbf{y}) + \mathbf{u}^T \mathcal{C}(\mathbf{x}, \mathbf{y})$$

Knowledge component: (Soft) constraints

Weight Vector for “local” models

Features, classifiers; log-linear models (HMM, CRF) or a combination

How far y is from a “legal/expected” assignment

Constrained Conditional Models

Penalty for violating the constraint.

$$\mathbf{y = argmax_{y \in \mathcal{Y}}\ w^T \phi(x, y) + u^T \mathcal{C}(x, y)}$$

Knowledge component: (Soft) constraints

Weight Vector for "local" models

Features, classifiers; log-linear models (HMM, CRF) or a combination

How far y is from a "legal/expected" assignment

- Training: learning the objective function (**w, u**)
  - Decouple? Decompose? Force **u** to model hard constraints?

Constrained Conditional Models

Penalty for violating the constraint.

$$y = \text{argmax}_{y \in \mathcal{Y}} \; w^T\phi(x, y) + u^T\mathcal{C}(x, y)$$

Knowledge component: (Soft) constraints

Weight Vector for "local" models

Features, classifiers; log-linear models (HMM, CRF) or a combination

How far y is from a "legal/expected" assignment

- Training: learning the objective function (**w, u**)
  - Decouple? Decompose? Force **u** to model hard constraints?
- A way to push the learned model to **satisfy our output expectations** (or expectations from a latent representation)
  - [CoDL, Chang et. al (07, 12); Posterior Regularization, Ganchev et. al (10); Unified EM (Samdani et. al (12)]

Constrained Conditional Models

Penalty for violating the constraint.

$y = \text{argmax}_y \sum \mathbf{1}_{\phi(x,y)} w_{x,y}$ subject to Constraints C(x,y)

Knowledge component: (Soft) constraints

Weight Vector for "local" models

Features, classifiers; log-linear models (HMM, CRF) or a combination

How far y is from a "legal/expected" assignment

- Training: learning the objective function (**w, u**)
  - Decouple? Decompose? Force **u** to model hard constraints?
- A way to push the learned model to **satisfy our output expectations** (or expectations from a latent representation)
  - [CoDL, Chang et. al (07, 12); Posterior Regularization, Ganchev et. al (10); Unified EM (Samdani et. al (12)]

Constrained Conditional Models

Any MAP problem w.r.t. any probabilistic model, can be formulated as an ILP [Roth+ 04, Taskar 04]

$$y = \text{argmax}_y \sum \mathbf{1}_{\phi(x,y)}\, w_{x,y} \quad \text{subject to Constraints } C(x,y)$$

Knowledge component: (Soft) constraints

Weight Vector for "local" models

Features, classifiers; log-linear models (HMM, CRF) or a combination

How far y is from a "legal/expected" assignment

- Training: learning the objective function (**w, u**)
  - Decouple? Decompose? Force **u** to model hard constraints?
- A way to push the learned model to **satisfy our output expectations** (or expectations from a latent representation)
  - [CoDL, Chang et. al (07, 12); Posterior Regularization, Ganchev et. al (10); Unified EM (Samdani et. al (12)]

Constrained Conditional Models

Any MAP problem w.r.t. any probabilistic model, can be formulated as an ILP [Roth+ 04, Taskar 04]

Variables are models

$y = \text{argmax}_y \sum \mathbf{1}_{\phi(x,y)}\, w_{x,y}$ subject to Constraints C(x,y)

Knowledge component: (Soft) constraints

Weight Vector for “local” models

Features, classifiers; log-linear models (HMM, CRF) or a combination

How far y is from a “legal/expected” assignment

- Training: learning the objective function (**w, u**)
  - Decouple? Decompose? Force **u** to model hard constraints?
- A way to push the learned model to **satisfy our output expectations** (or expectations from a latent representation)
  - [CoDL, Chang et. al (07, 12); Posterior Regularization, Ganchev et. al (10); Unified EM (Samdani et. al (12)]

Constrained Conditional Models

Any MAP problem w.r.t. any probabilistic model, can be formulated as an ILP [Roth+ 04, Taskar 04]

Variables are models

$y = \text{argmax}_y \sum \mathbf{1}_{\phi(x,y)}\, w_{x,y}$ subject to Constraints C(x,y)

Knowledge component: (Soft) constraints

Weight Vector for "local" models

Features, classifiers; log-linear models (HMM, CRF) or a combination

How far y is from a "legal/expected" assignment

- Training: learning the objective function (**w, u**)
  - Decouple? Decompose? Force **u** to model hard constraints?
- A way to push the learned model to **satisfy our output expectations** (or expectations from a latent representation)
  - [CoDL, Chang et. al (07, 12); Posterior Regularization, Ganchev et. al (10); Unified EM (Samdani et. al (12)]
- The benefits of thinking about it as an ILP are conceptual and computational.

Examples: CCM Formulations

$$\mathbf{y} = \mathrm{argmax}_{\mathbf{y} \in \mathcal{Y}}\ \mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}, \mathbf{y}) + \mathbf{u}^{\mathrm{T}}\mathcal{C}(\mathbf{x}, \mathbf{y})$$

Examples: CCM Formulations

$$\mathbf{y} = \mathbf{argmax}_{\mathbf{y} \in \mathcal{Y}}\ \mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}, \mathbf{y}) + \mathbf{u}^{\mathrm{T}}\mathcal{C}(\mathbf{x}, \mathbf{y})$$

While ϕ**(x, y) and** $\mathcal{C}$**(x, y)** could be the same; we want C(x, y) to express high level declarative knowledge over the statistical models.

Examples: CCM Formulations

$$\mathbf{y} = \mathbf{argmax}_{\mathbf{y} \in \mathcal{Y}}\ \mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}, \mathbf{y}) + \mathbf{u}^{\mathrm{T}}\mathcal{C}(\mathbf{x}, \mathbf{y})$$

While ϕ**(x, y) and** $\mathcal{C}$**(x, y)** could be the same; we want C(x, y) to express high level declarative knowledge over the statistical models.

Formulate NLP Problems as ILP problems (inference may be done otherwise)

1. Sequence tagging (HMM/CRF + Global constraints)
2. Sentence Compression (Language Model + Global Constraints)
3. SRL (Independent classifiers + Global Constraints)

Examples: CCM Formulations

$$\mathbf{y = argmax_{y \in \mathcal{Y}}\ w^T\phi(x, y) + u^T\mathcal{C}(x, y)}$$

While ϕ**(x, y) and** $\mathcal{C}$**(x, y)** could be the same; we want C(x, y) to express high level declarative knowledge over the statistical models.

Formulate NLP Problems as ILP problems (inference may be done otherwise)

1. Sequence tagging (HMM/CRF + Global constraints)
2. Sentence Compression (Language Model + Global Constraints)
3. SRL (Independent classifiers + Global Constraints)

Sequential Prediction

HMM/CRF based:

Argmax $\sum \lambda_{ij} x_{ij}$

Knowledge/Linguistics Constraints

Cannot have both A states and B states in an output sequence.

Examples: CCM Formulations

$$\mathbf{y} = \mathbf{argmax}_{\mathbf{y} \in \mathcal{Y}} \; \mathbf{w}^{\mathbf{T}}\phi(\mathbf{x}, \mathbf{y}) + \mathbf{u}^{\mathbf{T}}\mathcal{C}(\mathbf{x}, \mathbf{y})$$

While ϕ**(x, y) and** $\mathcal{C}$**(x, y)** could be the same; we want C(x, y) to express high level declarative knowledge over the statistical models.

Formulate NLP Problems as ILP problems (inference may be done otherwise)

1. Sequence tagging (HMM/CRF + Global constraints)
2. Sentence Compression (Language Model + Global Constraints)
3. SRL (Independent classifiers + Global Constraints)

Sentence Compression/Summarization:

Language Model based:

Argmax $\sum \lambda_{ijk}$ x_{ijk}

Knowledge/Linguistics Constraints

If a modifier chosen, include its head

If verb is chosen, include its arguments

Examples: CCM Formulations

$$\mathbf{y} = \mathrm{argmax}_{\mathbf{y} \in \mathcal{Y}}\ \mathbf{w}^T\phi(\mathbf{x}, \mathbf{y}) + \mathbf{u}^T\mathcal{C}(\mathbf{x}, \mathbf{y})$$

While ϕ**(x, y) and** $\mathcal{C}$**(x, y)** could be the same; we want C(x, y) to express high level declarative knowledge over the statistical models.

Formulate NLP Problems as ILP problems (inference may be done otherwise)

1. Sequence tagging (HMM/CRF + Global constraints)
2. Sentence Compression (Language Model + Global Constraints)
3. SRL (Independent classifiers + Global Constraints)

Sentence Compression/Summarization:

Language Model based:

Argmax $\sum \lambda_{ijk} x_{ijk}$

Knowledge/Linguistics Constraints

If a modifier chosen, include its head

If verb is chosen, include its arguments

Examples: CCM Formulations

$$\mathbf{y} = \mathbf{argmax}_{\mathbf{y} \in \mathcal{Y}}\ \mathbf{w}^T\phi(\mathbf{x}, \mathbf{y}) + \mathbf{u}^T\mathcal{C}(\mathbf{x}, \mathbf{y})$$

While ϕ**(x, y) and** $\mathcal{C}$**(x, y)** could be the same; we want C(x, y) to express high level declarative knowledge over the statistical models.

Formulate NLP Problems as ILP problems (inference may be done otherwise)

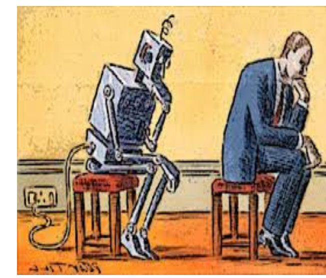
1. Sequence tagging (HMM/CRF + Global constraints)
2. Sentence Compression (Language Model + Global Constraints)
3. SRL (Independent classifiers + Global Constraints)

Constrained Conditional Models Allow:

- Decouple complexity of the learned model from that of the desired output
- Learn a simple model (multiple; pipelines); reason with a complex one.
- Accomplished by incorporating constraints to bias/re-rank global decisions to satisfy (minimally violate) expectations.

Outline

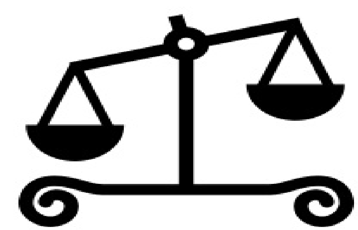
- Knowledge and Inference
  - Combining the soft with the logical/declarative nature of Natural Language
    - **Constrained Conditional Models: A formulation for global inference with knowledge modeled as expressive structural constraints**
    - **Some examples**



- Cycles of Knowledge
  - Grounding/Acquisition – knowledge – inference

- Learning with Indirect Supervision
  - Response Based Learning: learning from the world's feedback

- Scaling Up: Amortized Inference
  - Can the k-th inference problem be cheaper than the 1st?



Semantic Role Labeling (SRL)

I *left* my pearls to my daughter in my will .

[I]$_{A0}$ *left* [my pearls]$_{A1}$ [to my daughter]$_{A2}$ [in my will]$_{AM\text{-}LOC}$.

- ***A0*** Leaver
- ***A1*** Things left
- ***A2*** Benefactor
- ***AM-LOC*** Location

I *left* my pearls to my daughter in my will .

Semantic Role Labeling (SRL)

Archetypical Information Extraction Problem: E.g., Concept Identification and Typing, Event Identification, etc.

I *left* my pearls to my daughter in my will .

[I]$_{A0}$ *left* [my pearls]$_{A1}$ [to my daughter]$_{A2}$ [in my will]$_{AM\text{-}LOC}$.

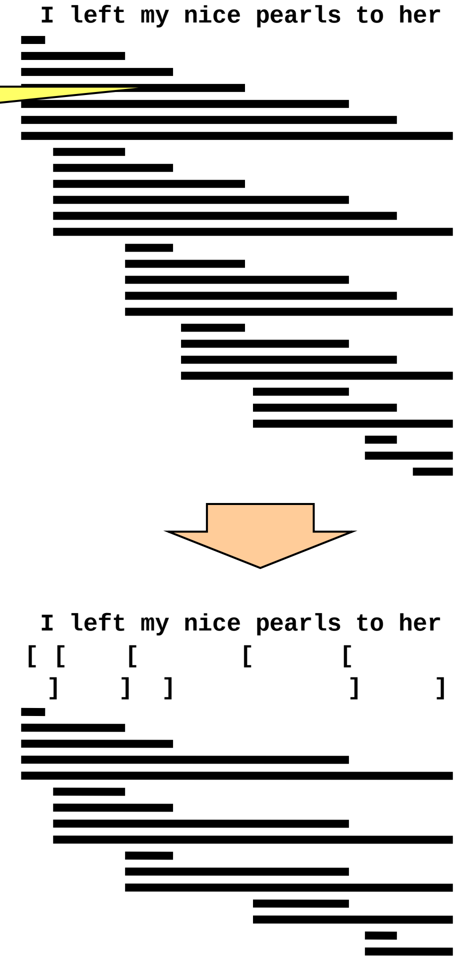
- ***A0*** Leaver
- ***A1*** Things left
- ***A2*** Benefactor
- ***AM-LOC*** Location

I *left* my pearls to my daughter in my will .

Algorithmic Approach

- Identify argument candidates
  - Pruning [Xue&Palmer, EMNLP'04]
  - Argument Identifier
    - **Binary classification**
- Classify argument candidates
  - Argument Classifier
    - **Multi-class classification**
- Inference
  - Use the estimated probability distribution given by the argument classifier
  - Use structural and linguistic constraints
  - Infer the optimal global output

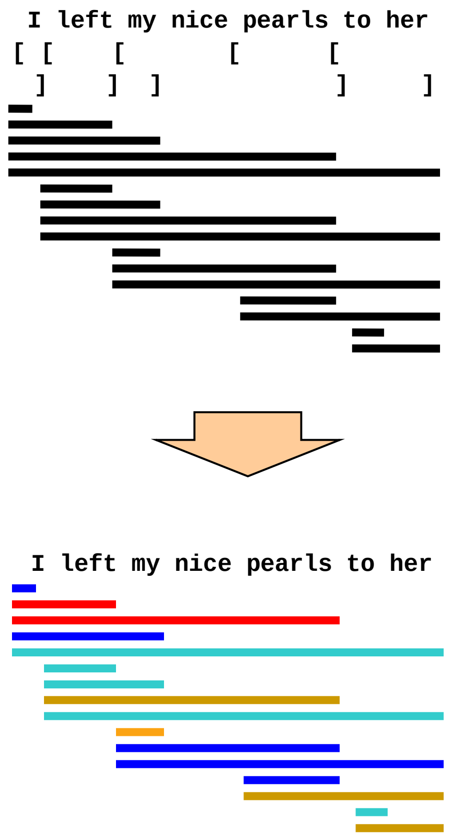
Algorithmic Approach



- **Identify** argument candidates
  - Pruning [Xue&Palmer, EMNLP'04]
  - Argument Identifier
    - **Binary classification**
- **Classify** argument candidates
  - Argument Classifier
    - **Multi-class classification**
- **Inference**
  - Use the estimated probability distribution given by the argument classifier
  - Use structural and linguistic constraints
  - Infer the optimal global output

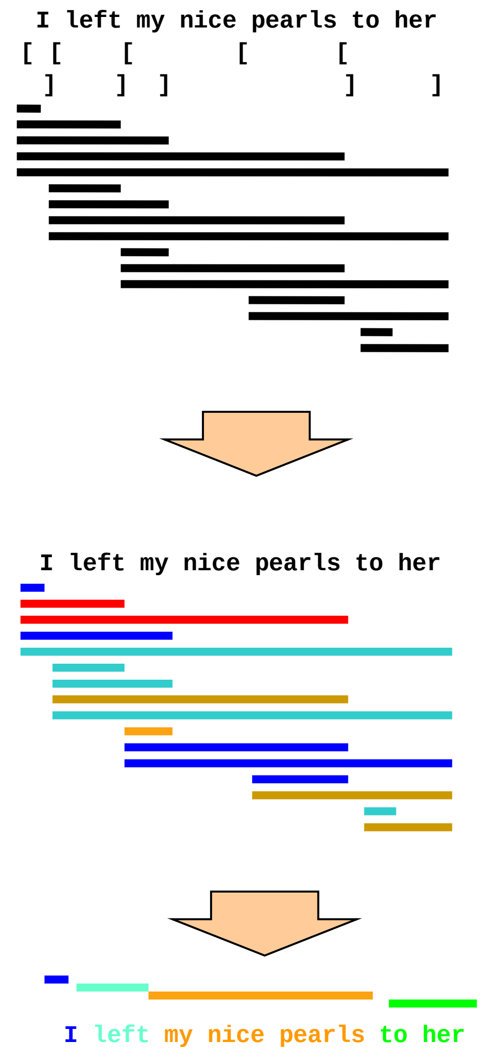
Algorithmic Approach

- Identify argument candidates
  - Pruning [Xue&Palmer, EMNLP'04]
  - Argument Identifier
    - **Binary classification**
- Classify argument candidates
  - Argument Classifier
    - **Multi-class classification**
- Inference
  - Use the estimated probability distribution given by the argument classifier
  - Use structural and linguistic constraints
  - Infer the optimal global output



Algorithmic Approach

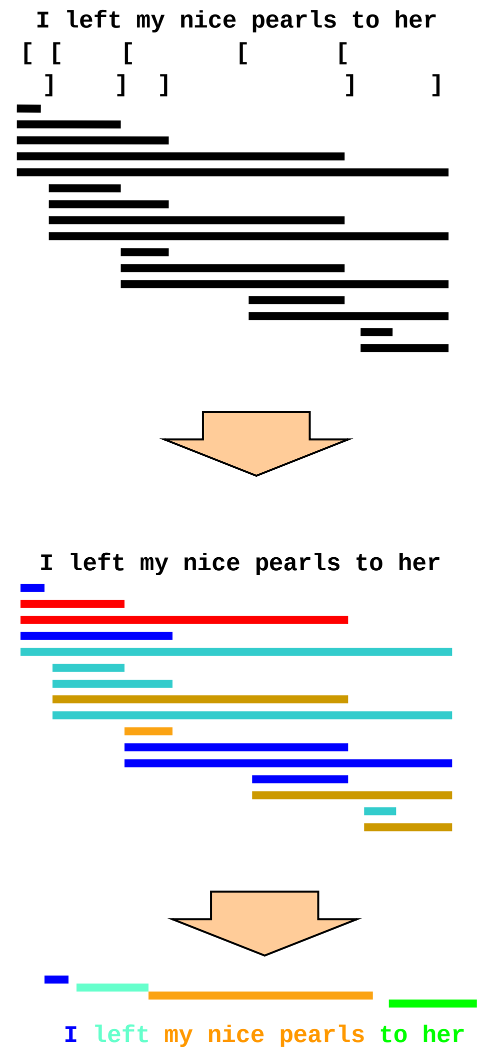
- Identify argument candidates
  - Pruning [Xue&Palmer, EMNLP'04]
  - Argument Identifier
    - **Binary classification**
- Classify argument candidates
  - Argument Classifier
    - **Multi-class classification**
- Inference
  - Use the estimated probability distribution given by the argument classifier
  - Use structural and linguistic constraints
  - Infer the optimal global output



Algorithmic Approach

- Identify argument candidates
  - Pruning [Xue&Palmer, EMNLP'04]
  - Argument Identifier
    - **Binary classification**
- Classify argument candidates
  - Argument Classifier
    - **Multi-class classification**
- Inference
  - Use the estimated probability distribution given by the argument classifier
  - Use st ic constraints
  - Infer t utput

One inference problem for each verb predicate.



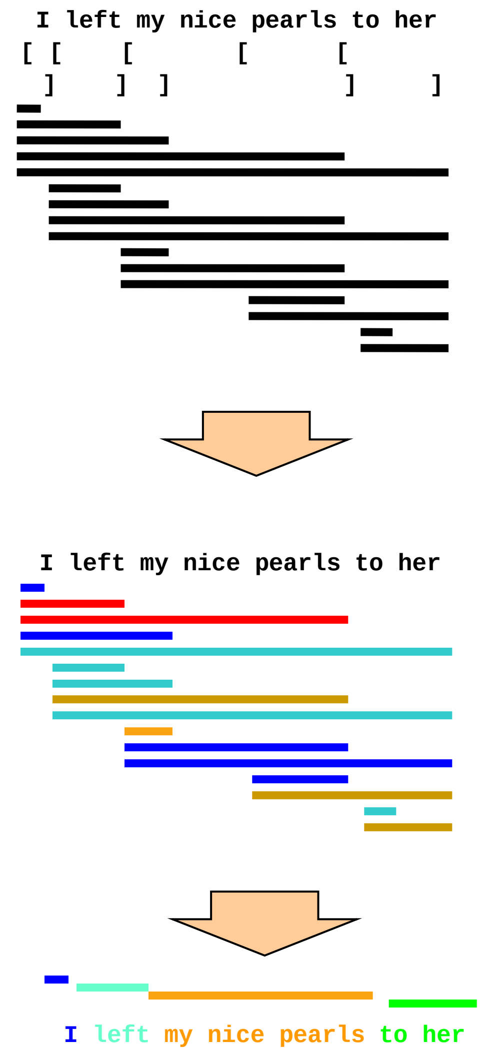
Algorithmic Approach

- Identify argument candidates
  - Pruning [Xue&Palmer, EMNLP'04]
  - Argument Identifier
    - **Binary classification**
- Classify argument candidates
  - Argument Classifier
    - **Multi-class classification**
- Inference

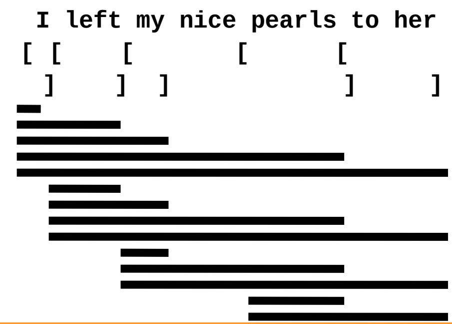
$$\operatorname{argmax} \sum_{a,t} y^{a,t} c^{a,t} = \sum_{a,t} 1_{a=t} c_{a=t}$$

Subject to:

- One label per argument: $\sum_t y^{a,t} = 1$
- No overlapping or embedding
- Relations between verbs and arguments,....



Algorithmic Approach



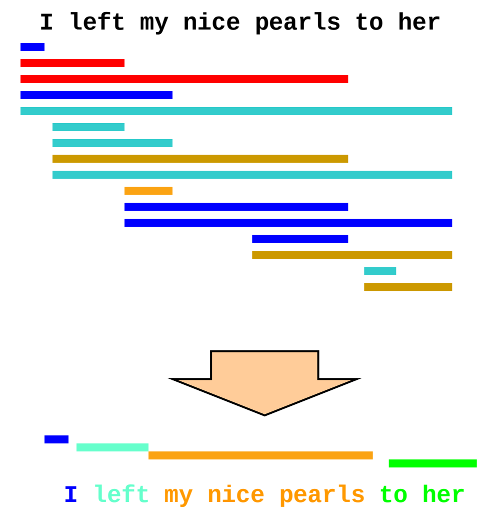
- Identify argument candidates
  - Pruning [Xue&Palmer, EMNLP'04]
  - Argument Identifier
    - **Binary classification**
- Classify argument candidates
  - Argument Classifier
    - **Multi-class classification**
- Inference

Variable $y^{a,t}$ indicates whether candidate argument a is assigned a label t.
$c^{a,t}$ is the corresponding model score

$$\mathrm{argmax} \sum_{a,t} y^{a,t} c^{a,t} = \sum_{a,t} 1_{a=t}\, c_{a=t}$$

Subject to:

- One label per argument: $\sum_t y^{a,t} = 1$
- No overlapping or embedding
- Relations between verbs and arguments,….



Algorithmic Approach

- Identify argument candid
  - Pruning [Xue&Palmer, EM
  - Argument Identifier
    - **Binary classification**
- Classify argument candid
  - Argument Classifier
    - **Multi-class classification**
- Inference

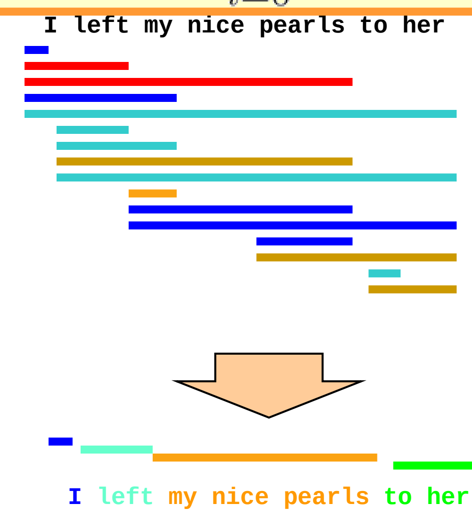
No duplicate argument classes $\forall i, \sum_{y \in \mathcal{Y}} 1_{\{y_i = y\}} = 1$

Unique labels

$$\forall y \in \mathcal{Y}, \sum_{i=0}^{n-1} 1_{\{y_i = y\}} \leq 1$$

$$\forall y \in \mathcal{Y}_R, \sum_{i=0}^{n-1} 1_{\{y_i = y = \text{“R-Ax”}\}} \leq \sum_{i=0}^{n-1} 1_{\{y_i = \text{“Ax”}\}}$$

$$\forall j, y \in \mathcal{Y}_C, 1_{\{y_j = y = \text{“C-Ax”}\}} \leq \sum_{i=0}^{j} 1_{\{y_i = \text{“Ax”}\}}$$



argmax $\sum_{a,t} y^{a,t} c^{a,t} = \sum_{a,t} 1_{a=t} c_{a=t}$

Subject to:

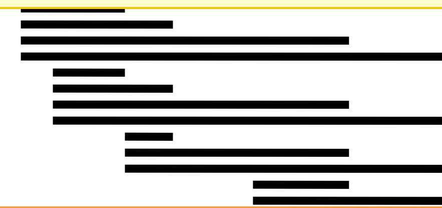
- One label per argument: $\sum_t y^{a,t} = 1$
- No overlapping or embedding
- Relations between verbs and arguments,….

n

Algorithmic Approach

Learning Based Java: allows a developer to encode constraints in First Order Logic; these are compiled into linear inequalities automatically.

- Identify argument candidates
  - Pruning [Xue&Palmer, EMNLP'04]
  - Argument Identifier
    - **Binary classification**
- Classify argument candidates
  - Argument Classifier
    - **Multi-class classification**
- Inference

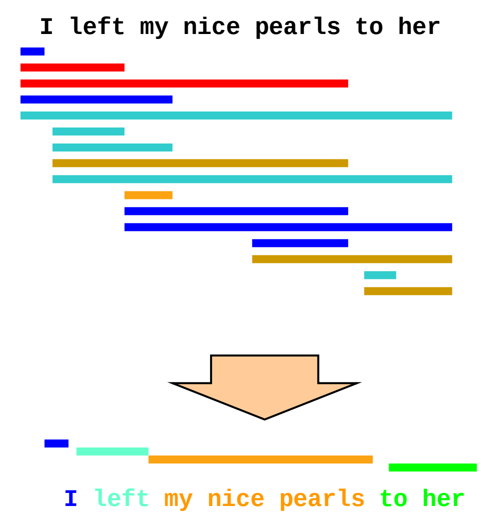


Variable $y^{a,t}$ indicates whether candidate argument a is assigned a label t.
$c^{a,t}$ is the corresponding model score

$$\text{argmax} \sum_{a,t} y^{a,t} c^{a,t} = \sum_{a,t} 1_{a=t} c_{a=t}$$

Subject to:

- One label per argument: $\sum_t y^{a,t} = 1$
- No overlapping or embedding
- Relations between verbs and arguments,….

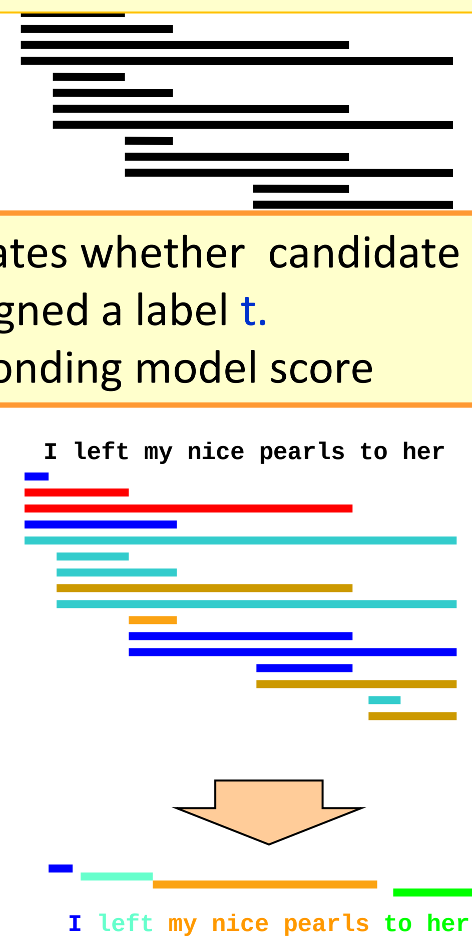


Algorithmic Approach

Learning Based Java: allows a developer to encode constraints in First Order Logic; these are compiled into linear inequalities automatically.

- Identify argument candidates
  - Pruning [Xue&Palmer, EMNLP'04]
  - Argument Identifier
    - **Binary classification**
- Classify argument candidates
  - Argument Classifier
    - **Multi-class classification**
- Inference

Variable $y^{a,t}$ indicates whether candidate argument a is assigned a label t.
$c^{a,t}$ is the corresponding model score



argmax $\sum_{a,t} y^{a,t} c^{a,t} = \sum_{a,t} 1_{a=t} c_{a=t}$

Subject to:

- One label per argument: $\sum_t y^{a,t} = 1$
- No overlapping or embedding
- Relations between verbs and arguments,….

Use the **pipeline architecture's simplicity** while **maintaining uncertainty**: keep probability distributions over decisions & use global inference at decision time.

Verb SRL is not Sufficient

- *John, a fast-rising politician, slept on the train to Chicago*.
- **Verb Predicate: sleep**

Verb SRL is not Sufficient

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

Verb SRL is not Sufficient

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

- **Who was John?**

Verb SRL is not Sufficient

- *John, a fast-rising politician, slept on the train to Chicago.*
- **Verb Predicate: sleep**
  - Sleeper: John, a fast-rising politician
  - Location: on the train to Chicago
- **Who was John?**
  - Relation: Apposition (comma)
  - John, a fast-rising politician

Verb SRL is not Sufficient

- *John, a fast-rising politician, slept on the train to Chicago.*
- **Verb Predicate: sleep**
  - Sleeper: John, a fast-rising politician
  - Location: on the train to Chicago
- **Who was John?**
  - Relation: Apposition (comma)
  - John, a fast-rising politician
- **What was John's destination?**

Verb SRL is not Sufficient

- *John, a fast-rising politician, slept on the train to Chicago.*
- **Verb Predicate: sleep**
  - Sleeper: John, a fast-rising politician
  - Location: on the train to Chicago
- **Who was John?**
  - Relation: Apposition (comma)
  - John, a fast-rising politician
- **What was John's destination?**
  - Relation: Destination (preposition)
  - train to Chicago

Extended Semantic Role Labeling

- Improved sentence level analysis; dealing with more phenomena

BEIRUT, Lebanon — Lebanon's main opposition group called for widespread protests on Sunday in the wake of a powerful bomb attack for which it blamed Syria, posing a challenge to a shaky coalition government that is led by pro-Syrian factions and intensifying fears that Syria's civil war is spilling over into this country.

BEIRUT, Lebanon — Lebanon's main opposition group called for widespread protests on Sunday in the wake of a powerful bomb attack for which it blamed Syria posing a challenge to a shaky coalition government that is led by pro-Syrian factions and intensifying fears that Syria's civil war is spilling over into this country.

[Beirut] is in [Lebanon].

[Lebanon] has a main opposition group.

[Lebanon's main opposition group] called for [widespread protests] [on Sunday].

There was [a powerful bomb attack].

[Lebanon's main opposition group] blamed [Syria].

[Pro-Syrian factions] lead [a shaky coalition government]

[Syria] has a [civil war].

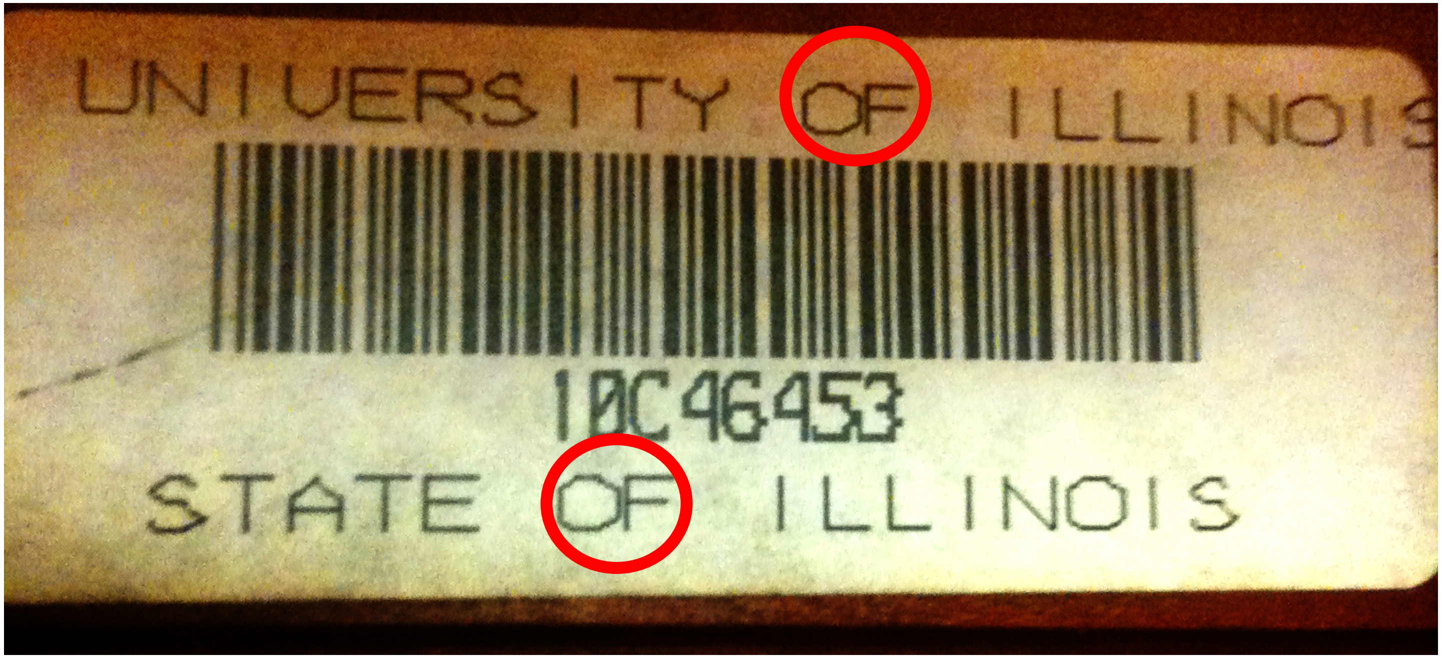
[Someone] fears that [Syria's civil war is spilling over into this country].

…

Sentence level analysis may be influenced by other sentences

Examples of Preposition Relations

Queen of England



City of Chicago

Predicates Expressed by Prepositions

live at Conway House
at:1

stopped at 9 PM
at:2

drive at 50 mph
at:5

look at the watch
at:9

cooler in the evening
in:3

the camp on the island
on:7

arrive on the 9th
on:17

Index of definition on Oxford English Dictionary

Predicates Expressed by Prepositions

Ambiguity & Variability

Index of definition on Oxford English Dictionary

live at Conway House	**at:1**
stopped at 9 PM	**at:2**
drive at 50 mph	**at:5**
look at the watch	**at:9**
cooler in the evening	**in:3**
the camp on the island	**on:7**
arrive on the 9th	**on:17**

Predicates Expressed by Prepositions

Ambiguity & Variability

live at Conway House
`at:1`

stopped at 9 PM
`at:2`

drive at 50 mph
`at:5`

look at the watch
`at:9`

cooler in the evening
`in:3`

the camp on the island
`on:7`

arrive on the 9th
`on:17`

Predicates Expressed by Prepositions

Ambiguity & Variability

live at Conway House
at:1

stopped at 9 PM
at:2

drive at 50 mph
at:5

look at the watch
at:9

cooler in the evening
in:3

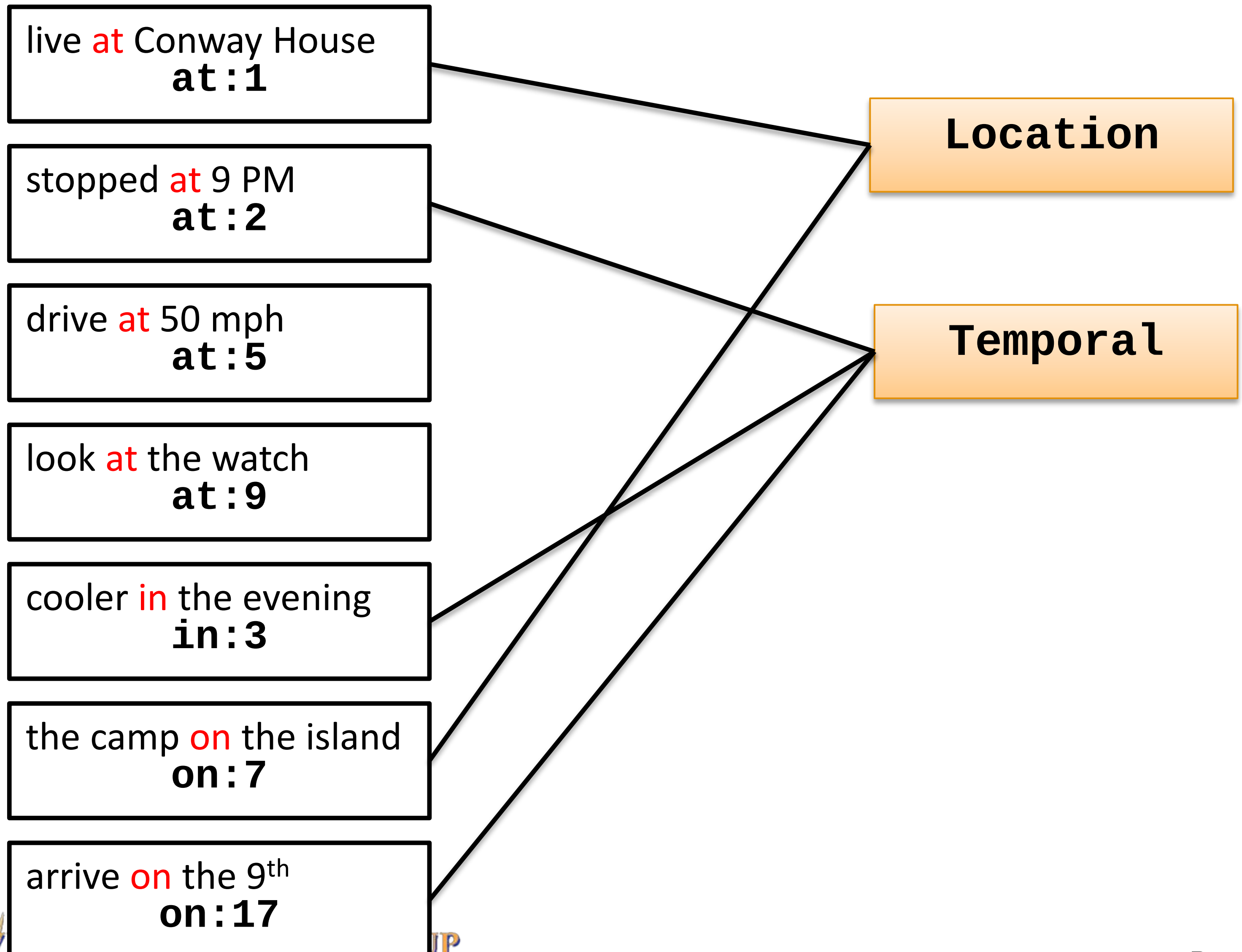
the camp on the island
on:7

arrive on the 9th
on:17

Location

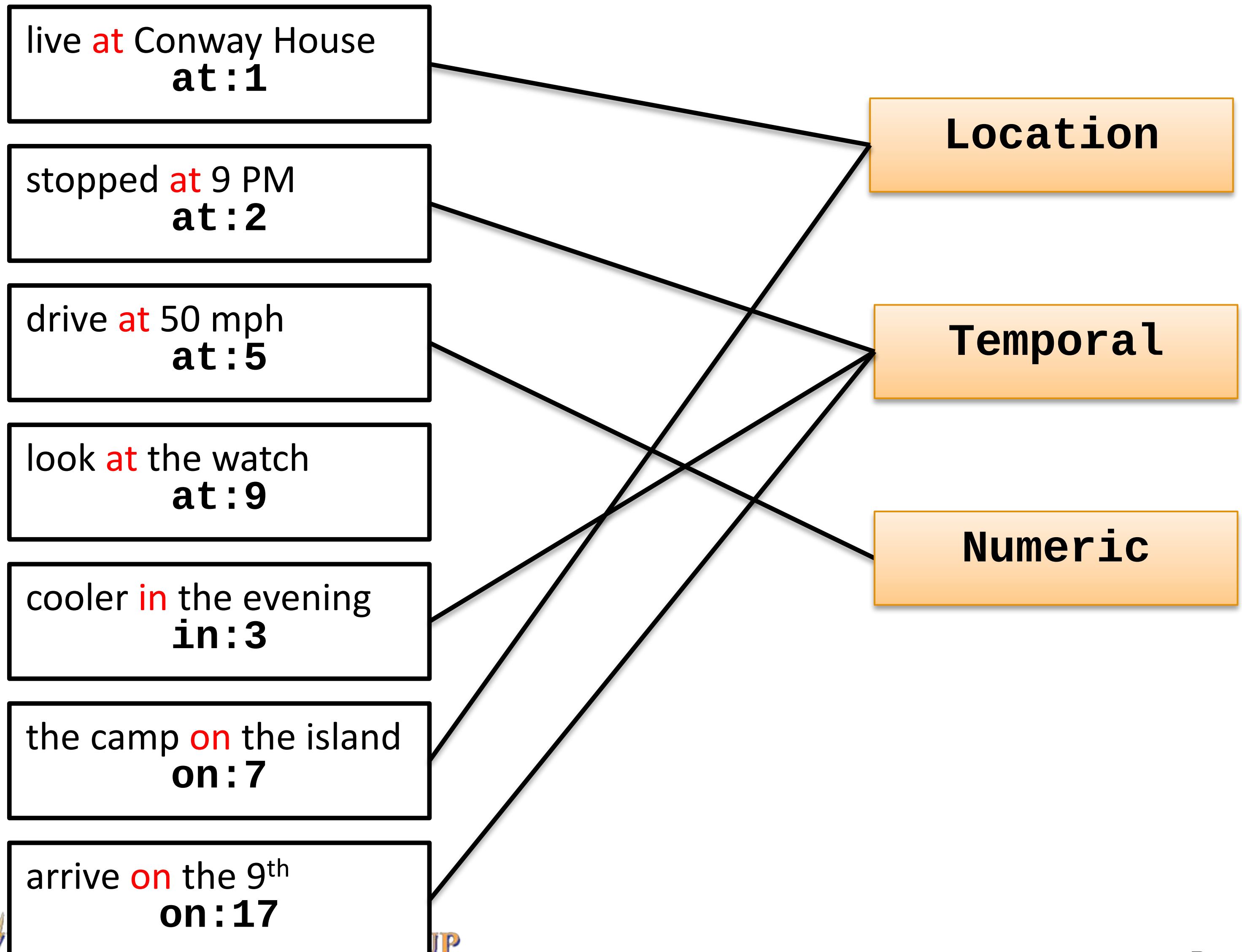
Predicates Expressed by Prepositions

Ambiguity & Variability



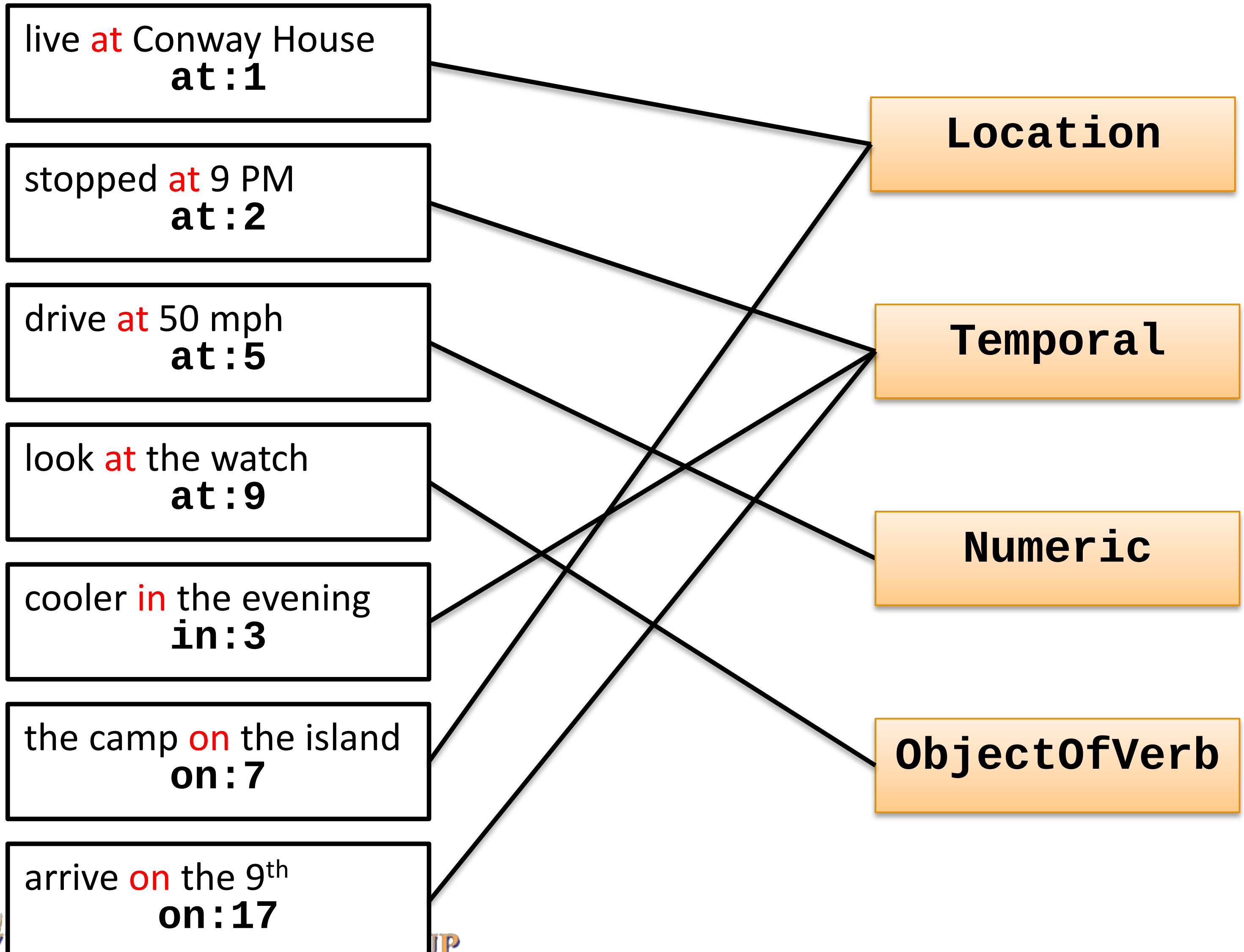
Predicates Expressed by Prepositions

Ambiguity & Variability



Predicates Expressed by Prepositions

Ambiguity & Variability



Computational Questions

- *John, a fast-rising politician, slept on the train to Chicago.*
- **Verb Predicate: sleep**
  - Sleeper: John, a fast-rising politician
  - Location: on the train to Chicago
- **Who was John?**
  - Relation: Apposition (comma)
  - John, a fast-rising politician
- **What was John's destination?**
  - Relation: Destination (preposition)
  - train to Chicago

Computational Questions

- *John, a fast-rising politician, slept on the train to Chicago.*
- **Verb Predicate: sleep**
  - Sleeper: John, a fast-rising politician
  - Location: on the train to Chicago
- **Who was John?**
  - Relation: Apposition (comma)
  - John, a fast-rising politician
- **What was John's destination?**
  - Relation: Destination (preposition)
  - train to Chicago

Identify the relation expressed by the predicate, and its arguments

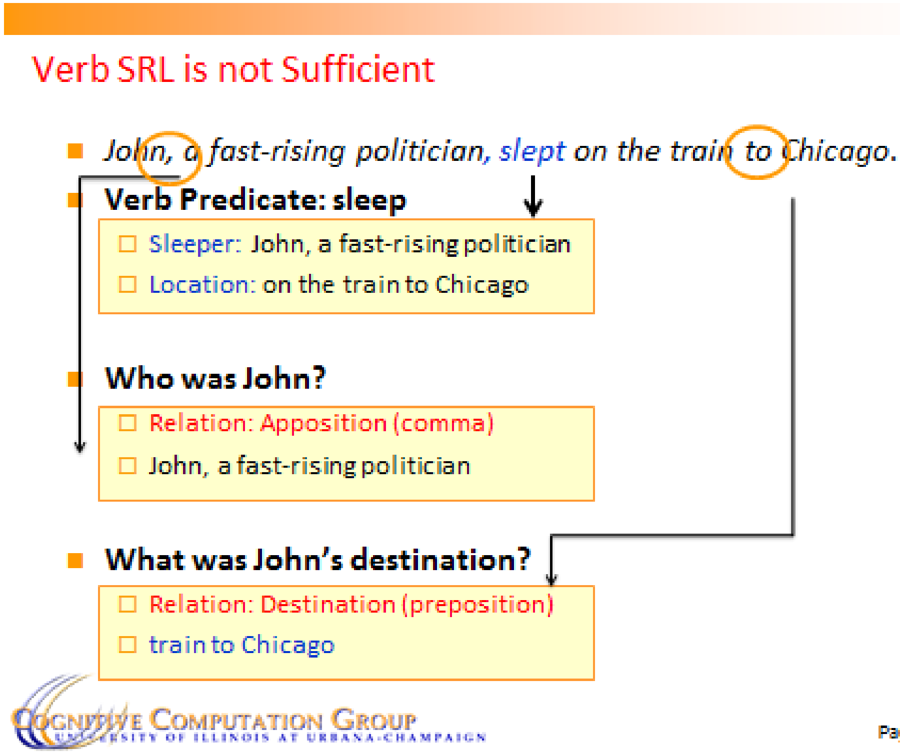
Computational Questions

- *John, a fast-rising politician, slept on the train to Chicago.*
- **Verb Predicate: sleep**
  - Sleeper: John, a fast-rising politician
  - Location: on the train to Chicago
- **Who was John?**
  - Relation: Apposition (comma)
  - John, a fast-rising politician
- **What was John's destination?**
  - Relation: Destination (preposition)
  - train to Chicago

Identify the relation expressed by the predicate, and its arguments

Computational Challenges

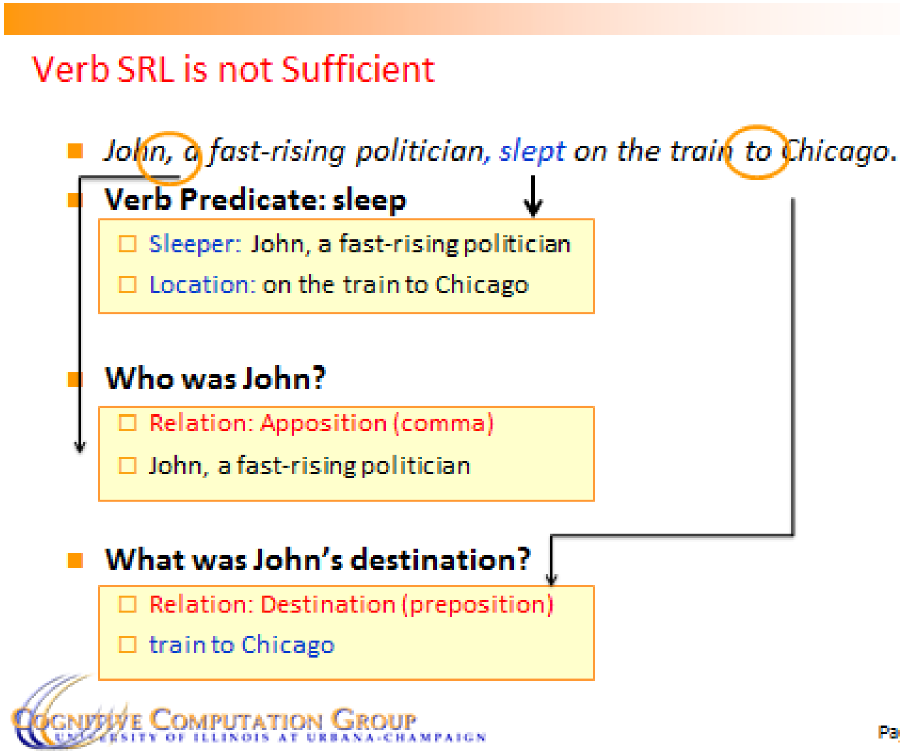
- Predict the preposition relations
  - [EMNLP, '11]
- Identify the relation's arguments
  - [Trans. Of ACL, '13]



Computational Challenges

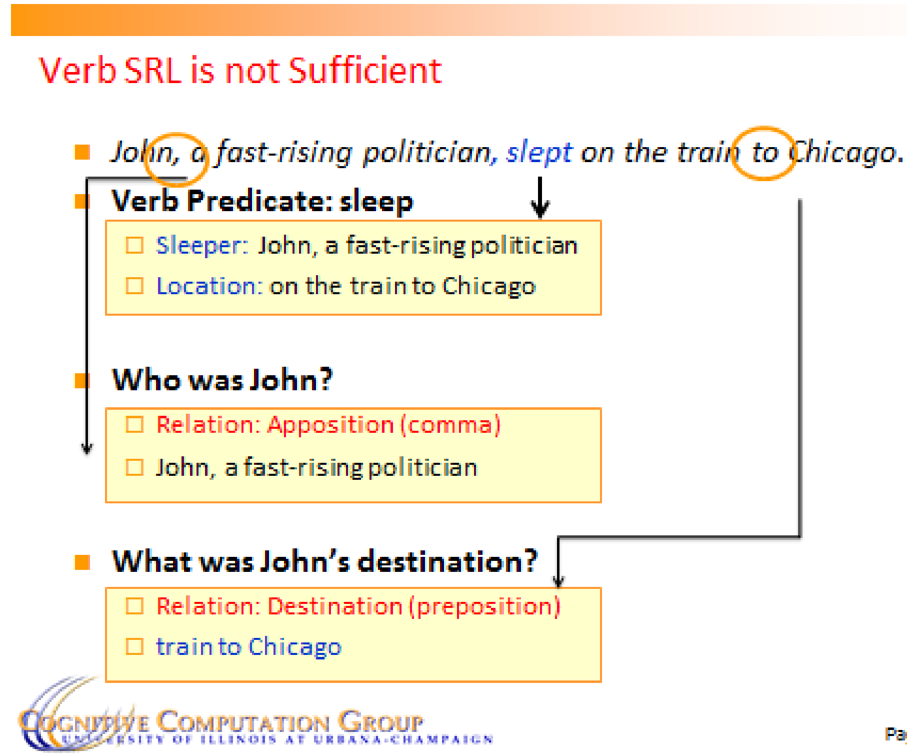
- Predict the preposition relations
  - [EMNLP, '11]
- Identify the relation's arguments
  - [Trans. Of ACL, '13]

- Very little supervised data
  - per phenomena
- Minimal annotation
  - only at the predicate level



Computational Challenges

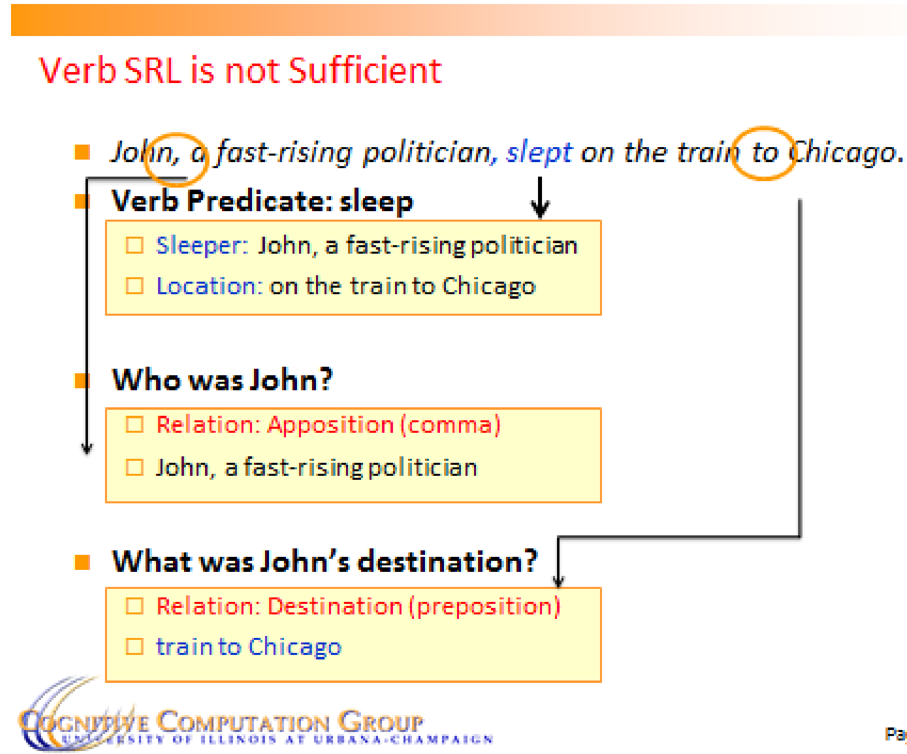
- Predict the preposition relations
  - [EMNLP, '11]
- Identify the relation's arguments
  - [Trans. Of ACL, '13]



- Very little supervised data
  - per phenomena
- Minimal annotation
  - only at the predicate level
- The Learning & Inference paradigm exploits two principles:
  - Coherency among multiple phenomena
  - Constraining latent structures (relating observed and latent variables)

Computational Challenges

- Predict the preposition relations
  - [EMNLP, '11]
- Identify the relation's arguments
  - [Trans. Of ACL, '13]



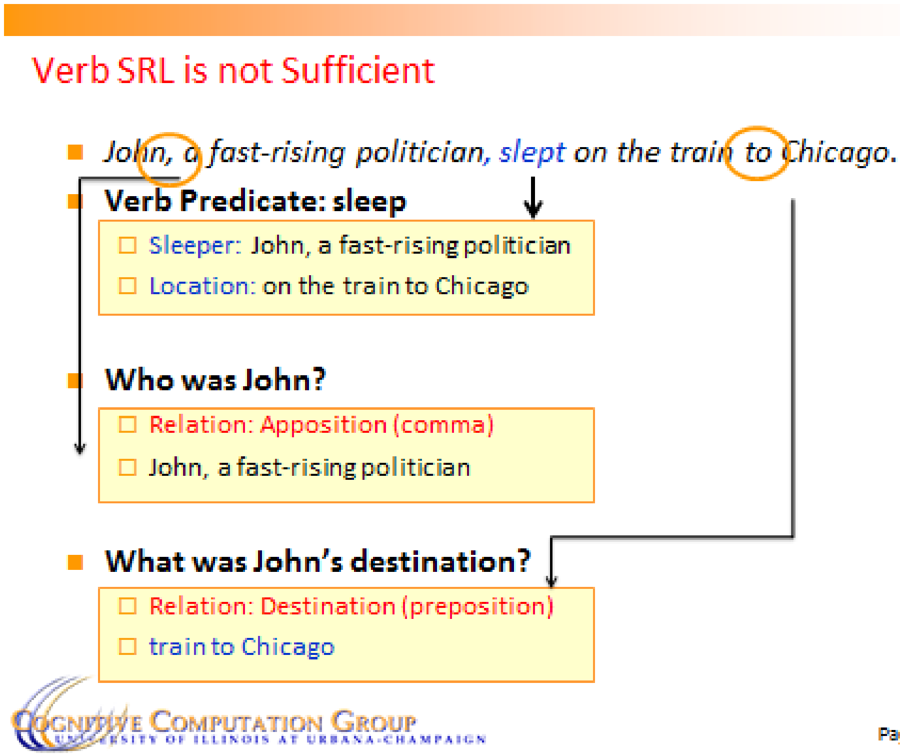
- Very little supervised data
  - per phenomena
- Minimal annotation
  - only at the predicate level
- The Learning & Inference paradigm exploits two principles:
  - Coherency among multiple phenomena
  - Constraining latent structures (relating observed and latent variables)

Input & relation

Argument & their types

Computational Challenges

- Predict the preposition relations
  - [EMNLP, '11]
- Identify the relation's arguments
  - [Trans. Of ACL, '13]



- Very little supervised data
  - per phenomena
- Minimal annotation
  - only at the predicate level
- The Learning & Inference paradigm exploits two principles:
  - Coherency among multiple phenomena
  - Constraining latent structures (relating observed and latent variables)
  - Skip

Input & relation

Argument & their types

Prepositional Relations: Coherence of predictions

The bus was heading for Nairobi in Kenya.

Location

Destination

Prepositional Relations: Coherence of predictions

The bus was heading for Nairobi in Kenya.

Location

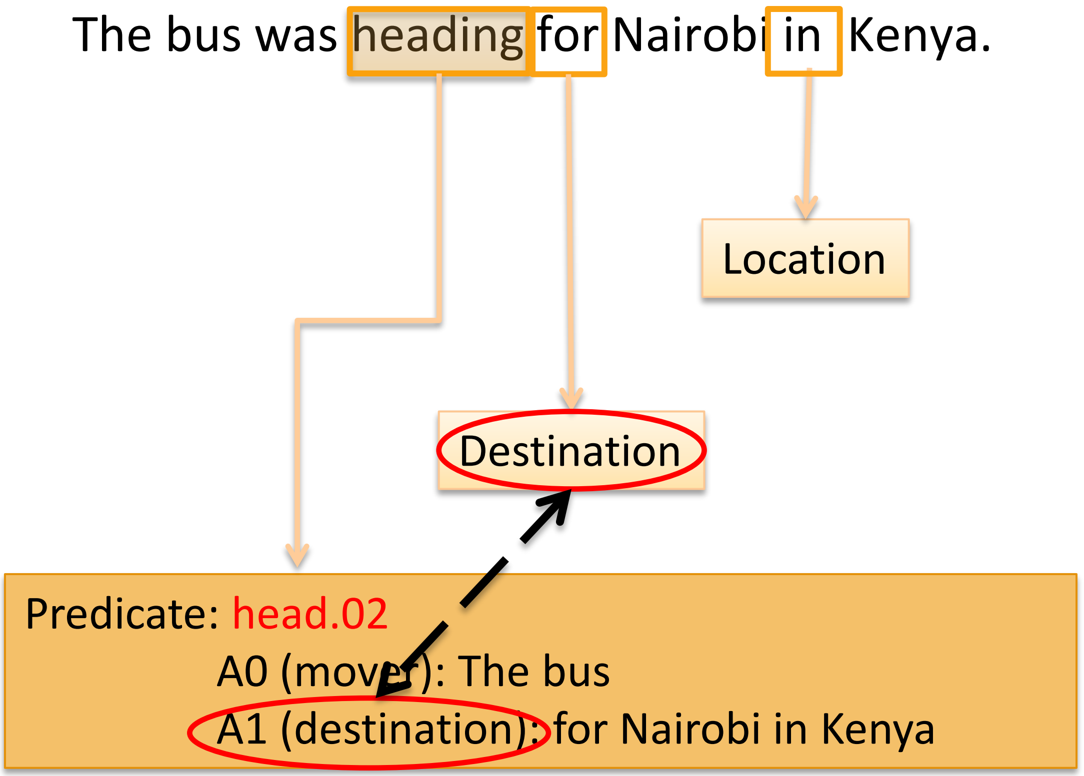
Destination

Predicate: head.02

A0 (mover): The bus

A1 (destination): for Nairobi in Kenya

Prepositional Relations: Coherence of predictions



Prepositional Relations: Coherence of predictions

Predicate arguments from different triggers should be consistent

The bus was heading for Nairobi in Kenya.

Joint constraints
linking the two tasks.

Destination ⇔ A1

Location

Destination

Predicate: head.02
A0 (mover): The bus
A1 (destination): for Nairobi in Kenya

Joint inference (CCMs)

$\boxed{\text{Verb arguments}}$

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

Joint inference (CCMs)

Variable $y^{a,t}$ indicates whether candidate argument a is assigned a label t.
$c^{a,t}$ is the corresponding model score

Verb arguments

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

Joint inference (CCMs)

Variable $y^{a,t}$ indicates whether candidate argument a is assigned a label t.
$c^{a,t}$ is the corresponding model score

Verb arguments

$$\max_{\mathbf{y}} \sum_{\boxed{t}} \sum_{\boxed{a}} y^{a,t} c^{a,t}$$

Argument candidates

Each argument label

Joint inference (CCMs)

Variable $y^{a,t}$ indicates whether candidate argument a is assigned a label t.
$c^{a,t}$ is the corresponding model score

Verb arguments

$$\max_{\mathbf{y}} \sum_t \sum_a y^{a,t} c^{a,t}$$

Constraints:

Verb SRL constraints

Joint inference (CCMs)

Variable $y^{a,t}$ indicates whether candidate argument a is assigned a label t.
$c^{a,t}$ is the corresponding model score

Verb arguments

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

Preposition relations

$$\max_{\mathbf{y}} \sum_{r} \sum_{p} y^{r,p} c^{r,p}$$

Constraints:

Verb SRL constraints

Joint inference (CCMs)

Variable $y^{a,t}$ indicates whether candidate argument a is assigned a label t.
$c^{a,t}$ is the corresponding model score

Verb arguments

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

Preposition relations

$$\max_{\mathbf{y}} \sum_{\boxed{r}} \sum_{\boxed{p}} y^{r,p} c^{r,p}$$

Preposition relation label

Preposition

Constraints:

Verb SRL constraints

Joint inference (CCMs)

Variable $y^{a,t}$ indicates whether candidate argument a is assigned a label t.
$c^{a,t}$ is the corresponding model score

Verb arguments

$$\max_{\mathbf{y}} \sum_t \sum_a y^{a,t} c^{a,t}$$

Preposition relations

$$\max_{\mathbf{y}} \sum_r \sum_p y^{r,p} c^{r,p}$$

Constraints:

Verb SRL constraints

Preposition SRL Constraints

Joint inference (CCMs)

Variable $y^{a,t}$ indicates whether candidate argument a is assigned a label t.
$c^{a,t}$ is the corresponding model score

Verb arguments

Preposition relations

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

Constraints:

Verb SRL constraints

Preposition SRL Constraints

Joint inference (CCMs)

Variable $y^{a,t}$ indicates whether candidate argument a is assigned a label t.
$c^{a,t}$ is the corresponding model score

Verb arguments

Preposition relations

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

Constraints:

Verb SRL constraints

Preposition SRL Constraints

+ Joint constraints between tasks; easy with ILP formulations

Joint inference (CCMs)

Variable $y^{a,t}$ indicates whether candidate argument a is assigned a label t.
$c^{a,t}$ is the corresponding model score

Verb arguments

Preposition relations

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

Constraints:

Verb SRL constraints

Preposition SRL Constraints

+ Joint constraints between tasks; easy with ILP formulations

Joint Inference – no (or minimal) joint learning

Joint inference (CCMs)

Variable $y^{a,t}$ indicates whether candidate argument a is assigned a label t.
$c^{a,t}$ is the corresponding model score

Verb arguments

Preposition relations

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

Constraints:

Verb SRL constraints

Preposition SRL Constraints

+ Joint constraints between tasks; easy with ILP formulations

Joint Inference – no (or minimal) joint learning

Predicate-Argument Structure of Prepositions

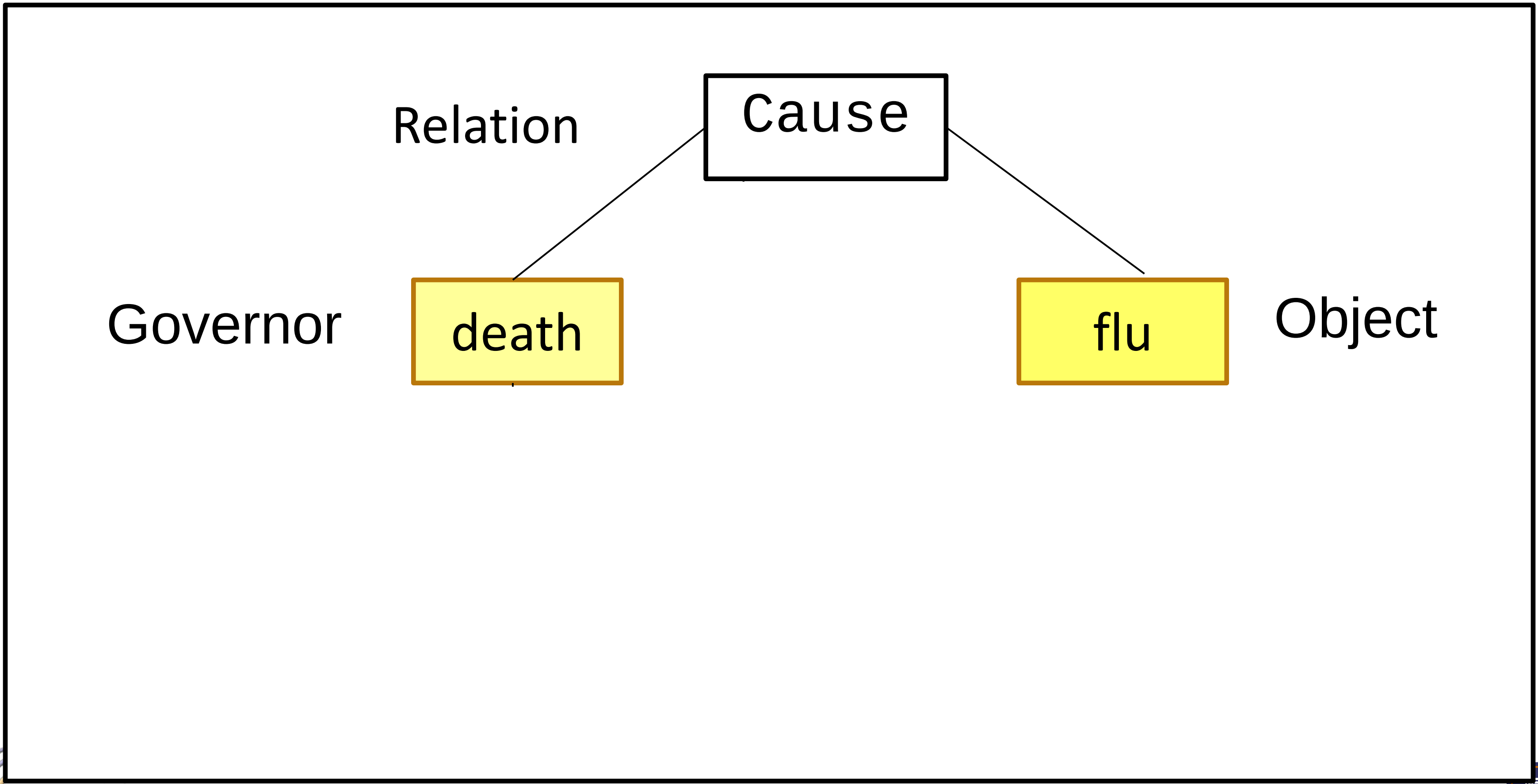
Poor care led to her death **from** flu.

Predicate-Argument Structure of Prepositions

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

Predicate-Argument Structure of Prepositions

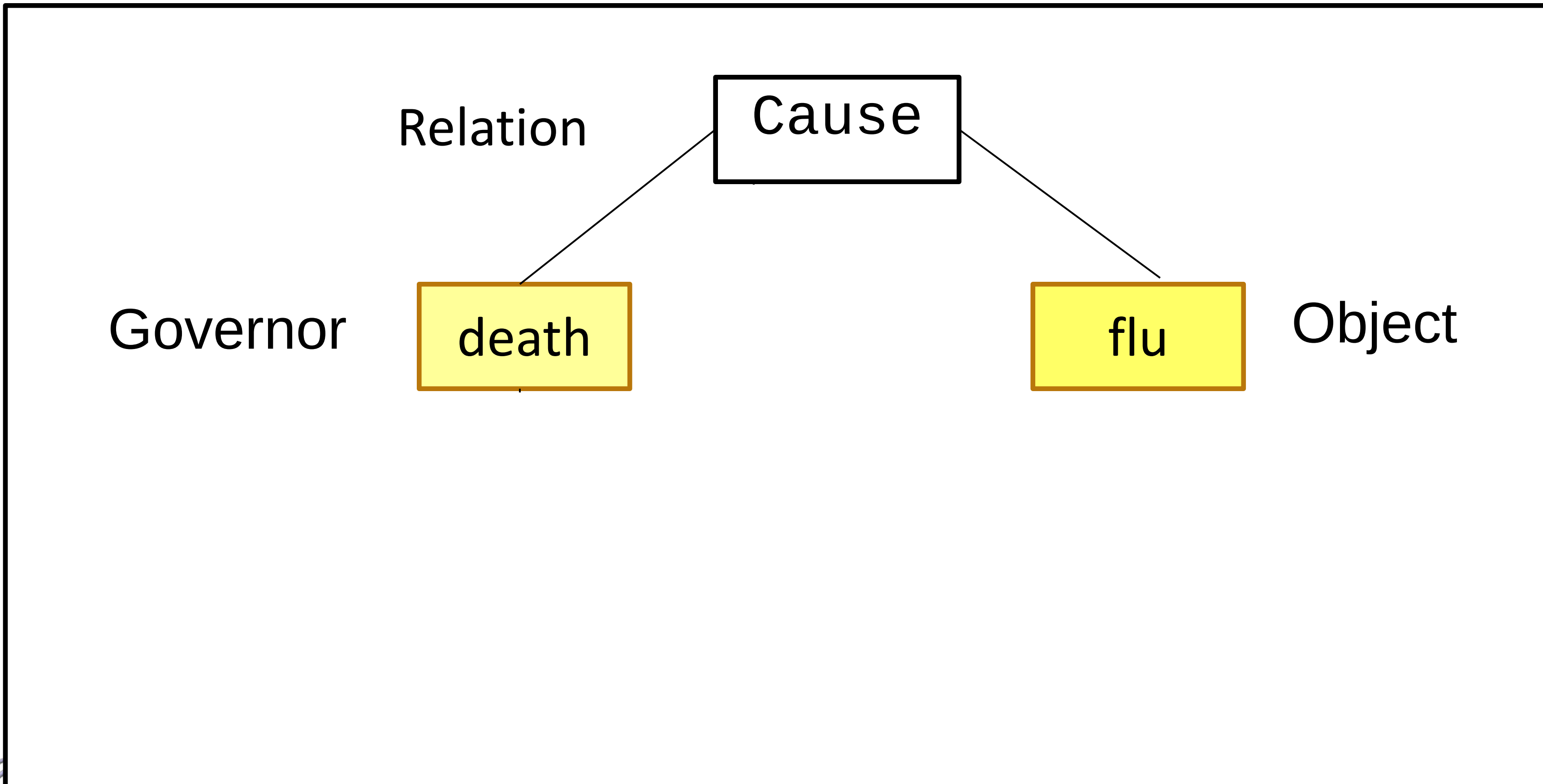
Poor care led to her death **from** flu.



Predicate-Argument Structure of Prepositions

……………her to suffer **from** infection.

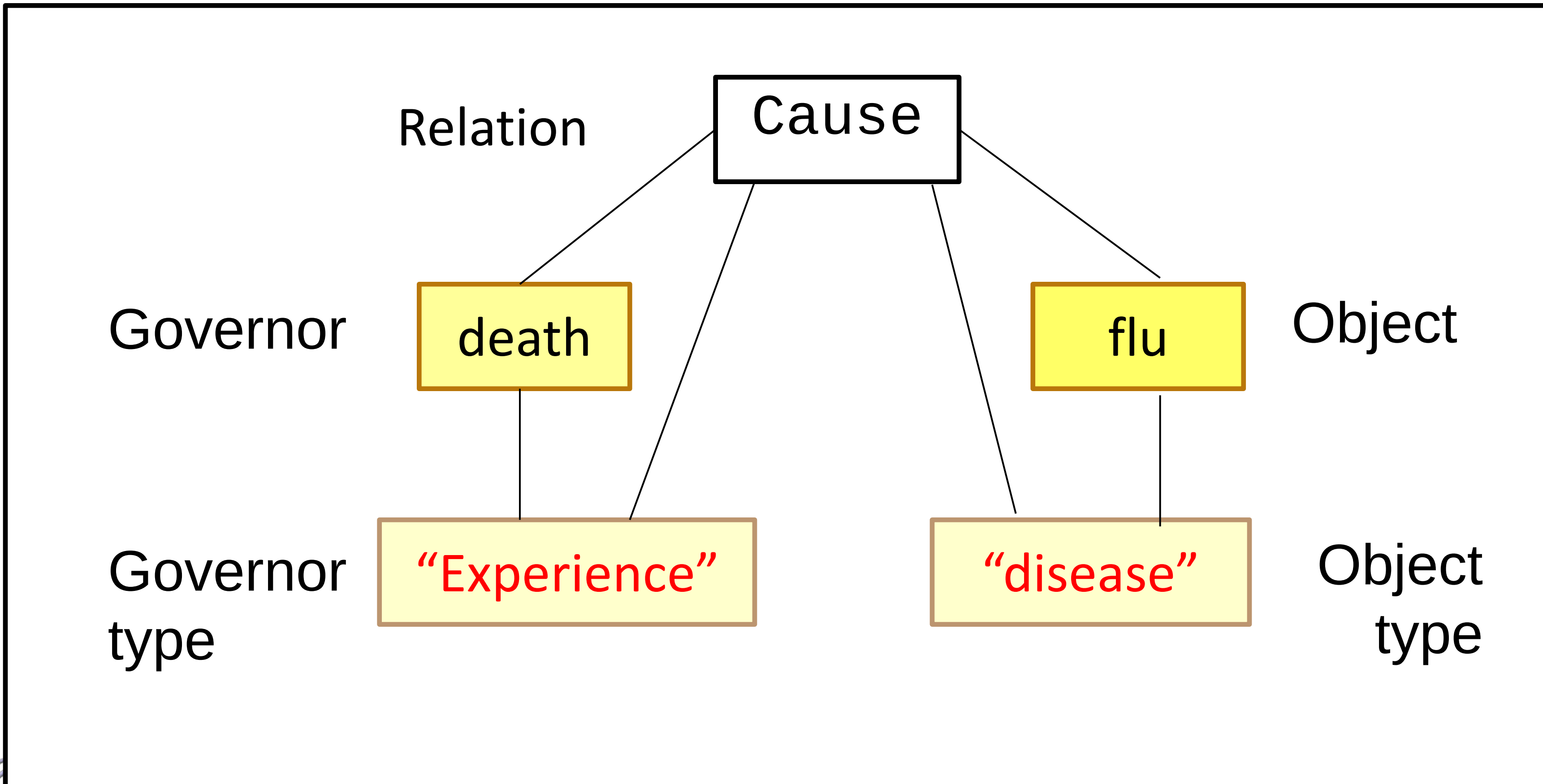
Poor care led to her death **from** flu.



Predicate-Argument Structure of Prepositions

……………her to suffer **from** infection.

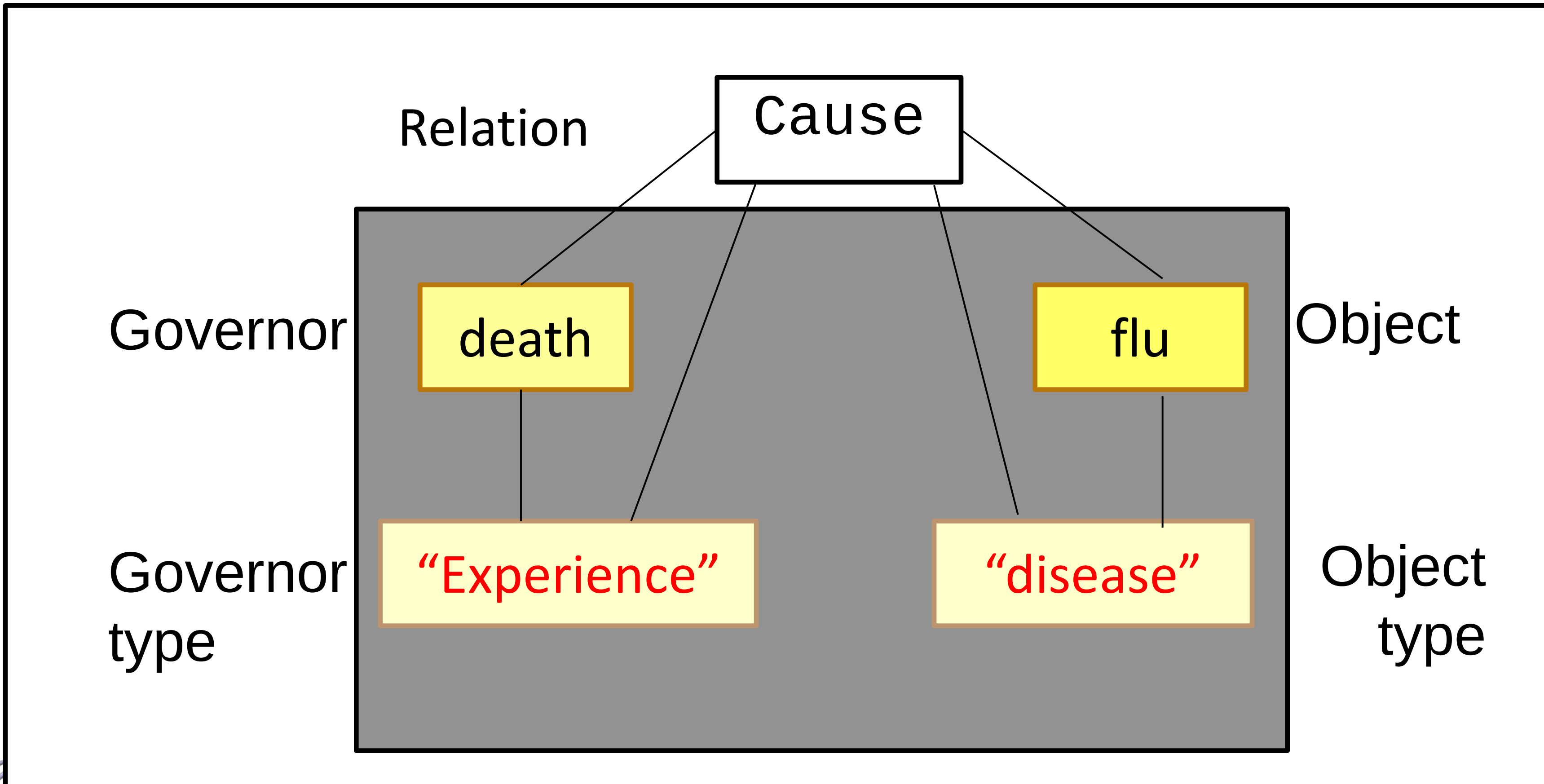
Poor care led to her death **from** flu.



Predicate-Argument Structure of Prepositions

…………….her to suffer **from** infection.

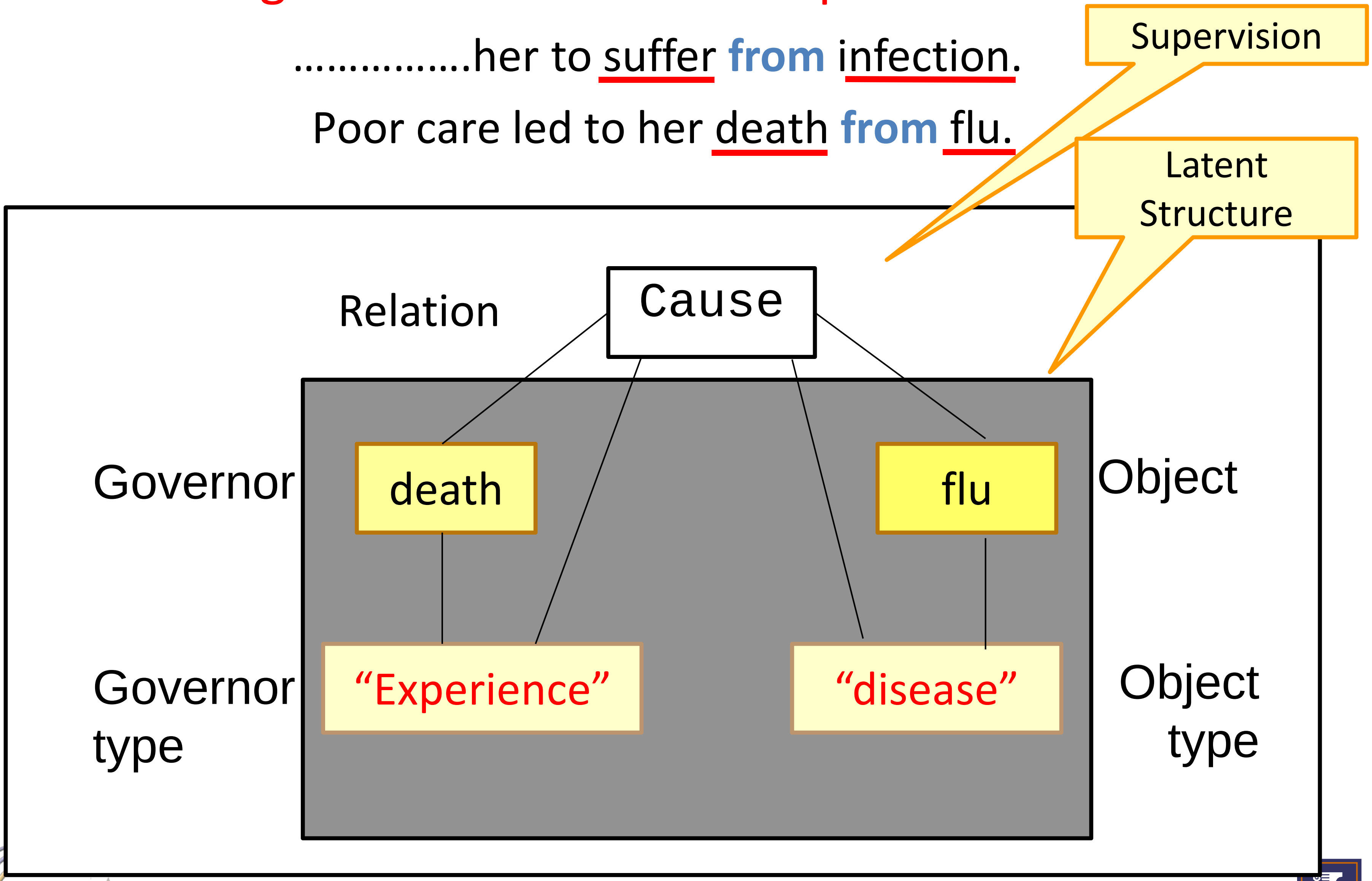
Poor care led to her death **from** flu.



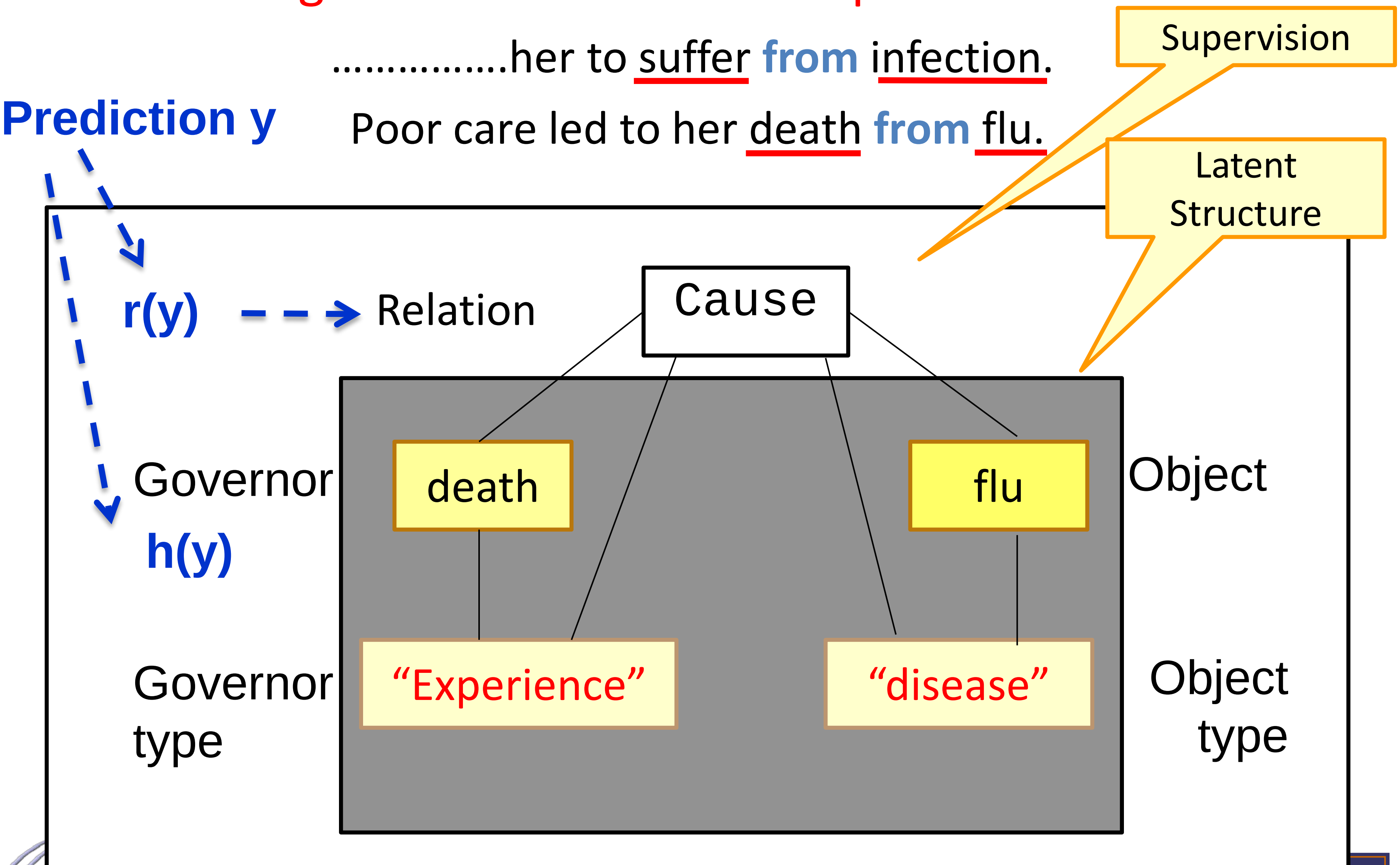
Predicate-Argument Structure of Prepositions

……………her to suffer **from** infection.

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



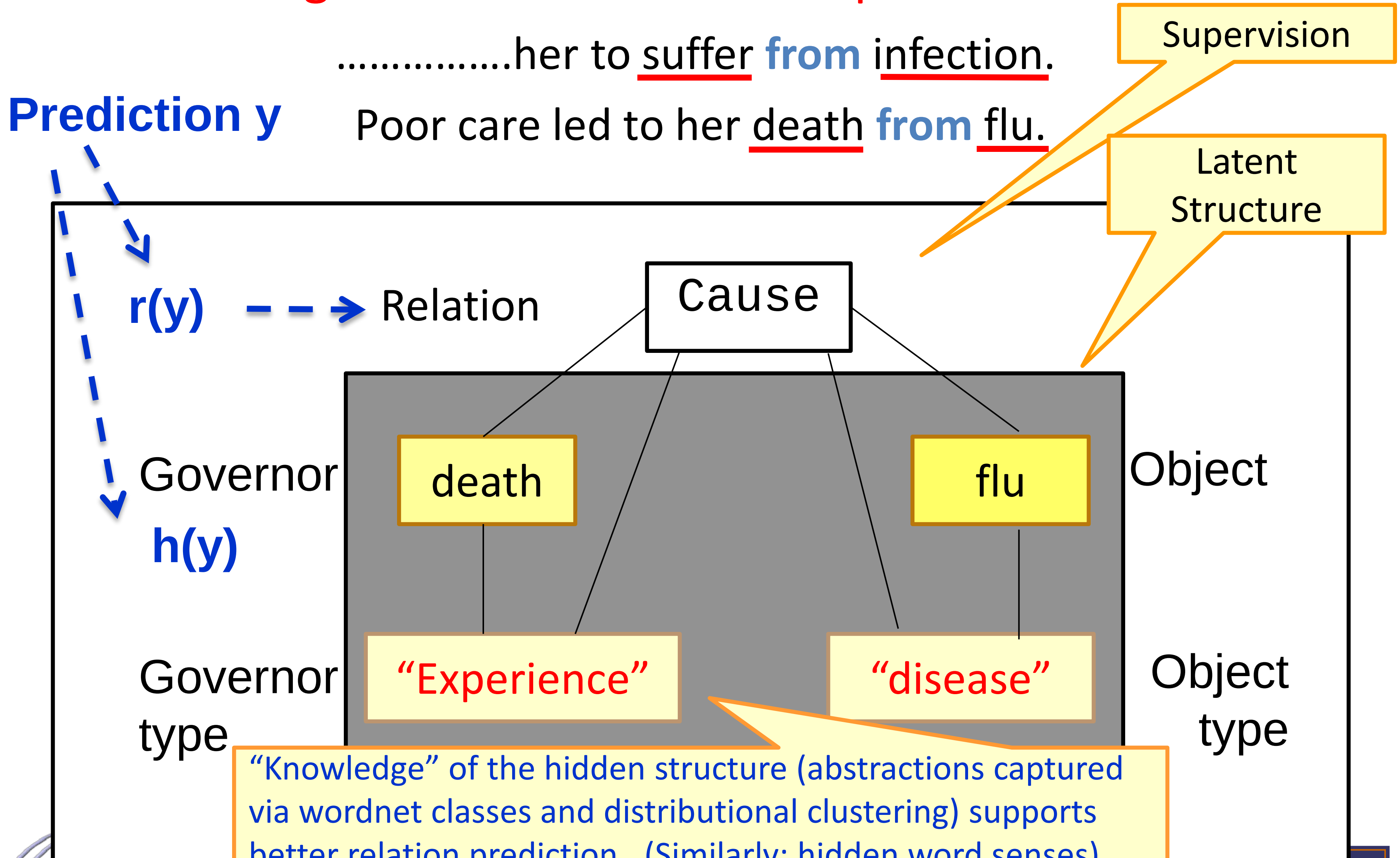
Predicate-Argument Structure of Prepositions



Predicate-Argument Structure of Prepositions

……………her to suffer **from** infection.

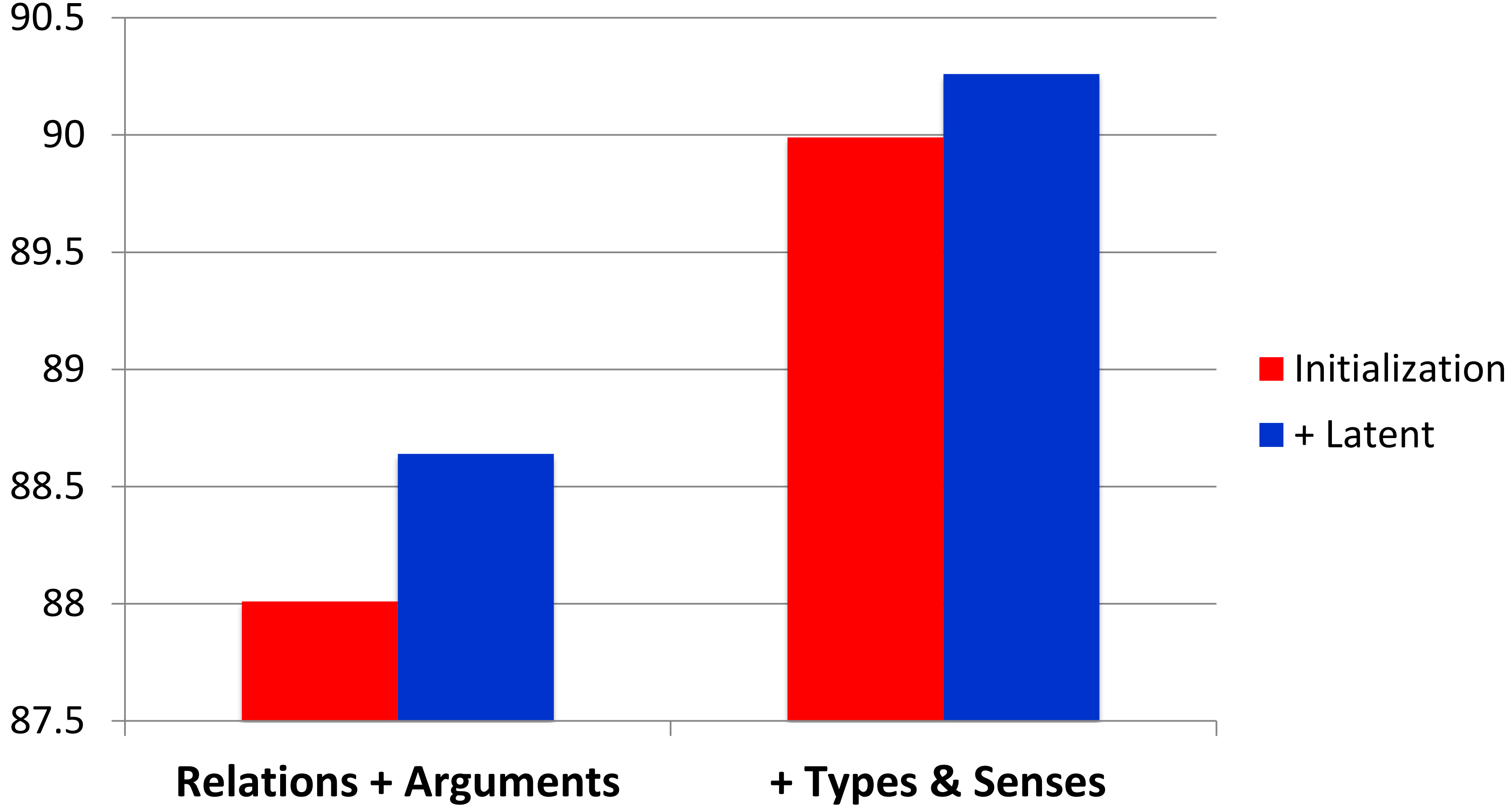
Poor care led to her death **from** flu.



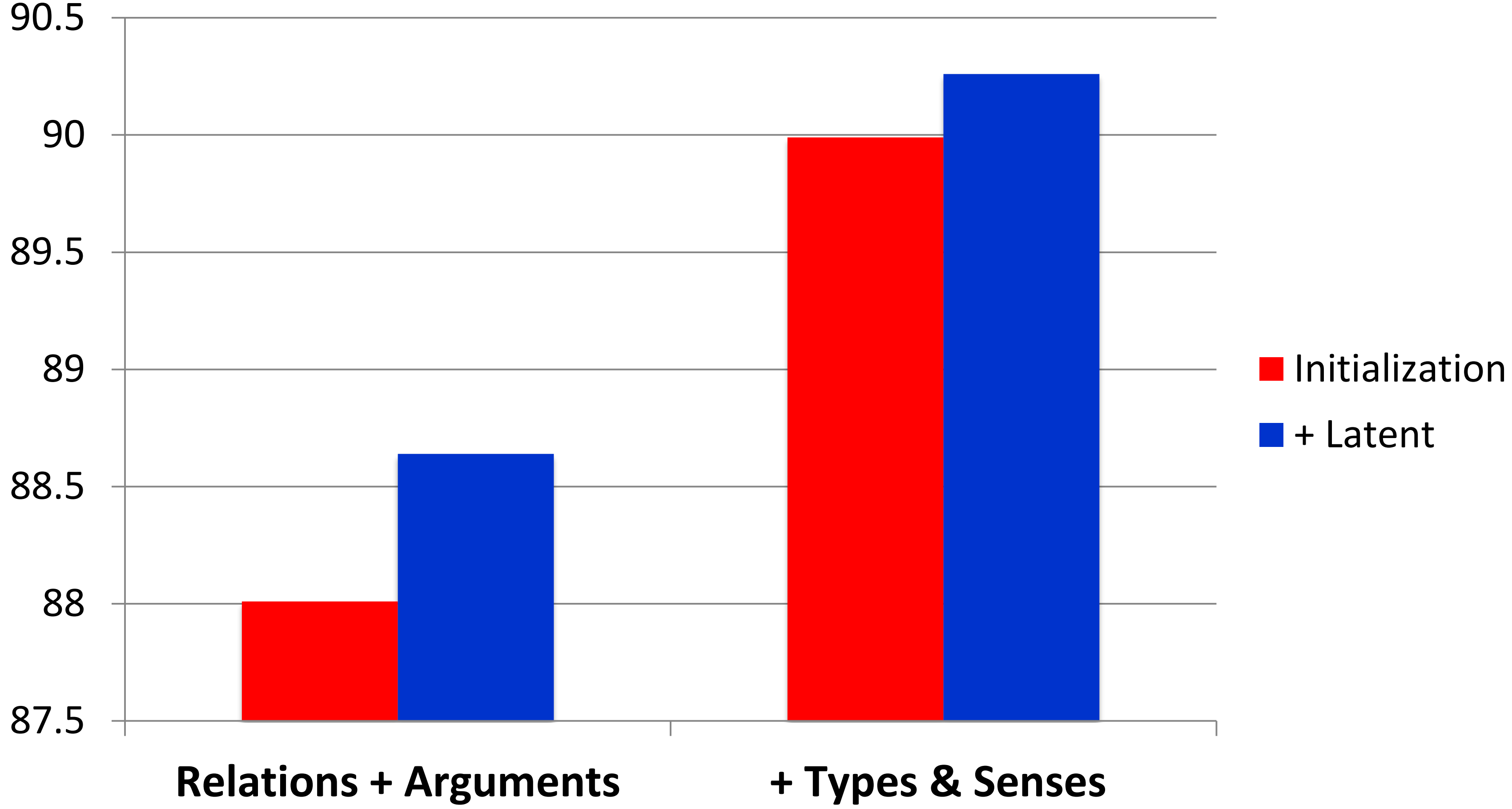
"Knowledge" of the hidden structure (abstractions captured via wordnet classes and distributional clustering) supports better relation prediction. (Similarly: hidden word senses) Inference relating latent and observed variables is a CCM

Performance

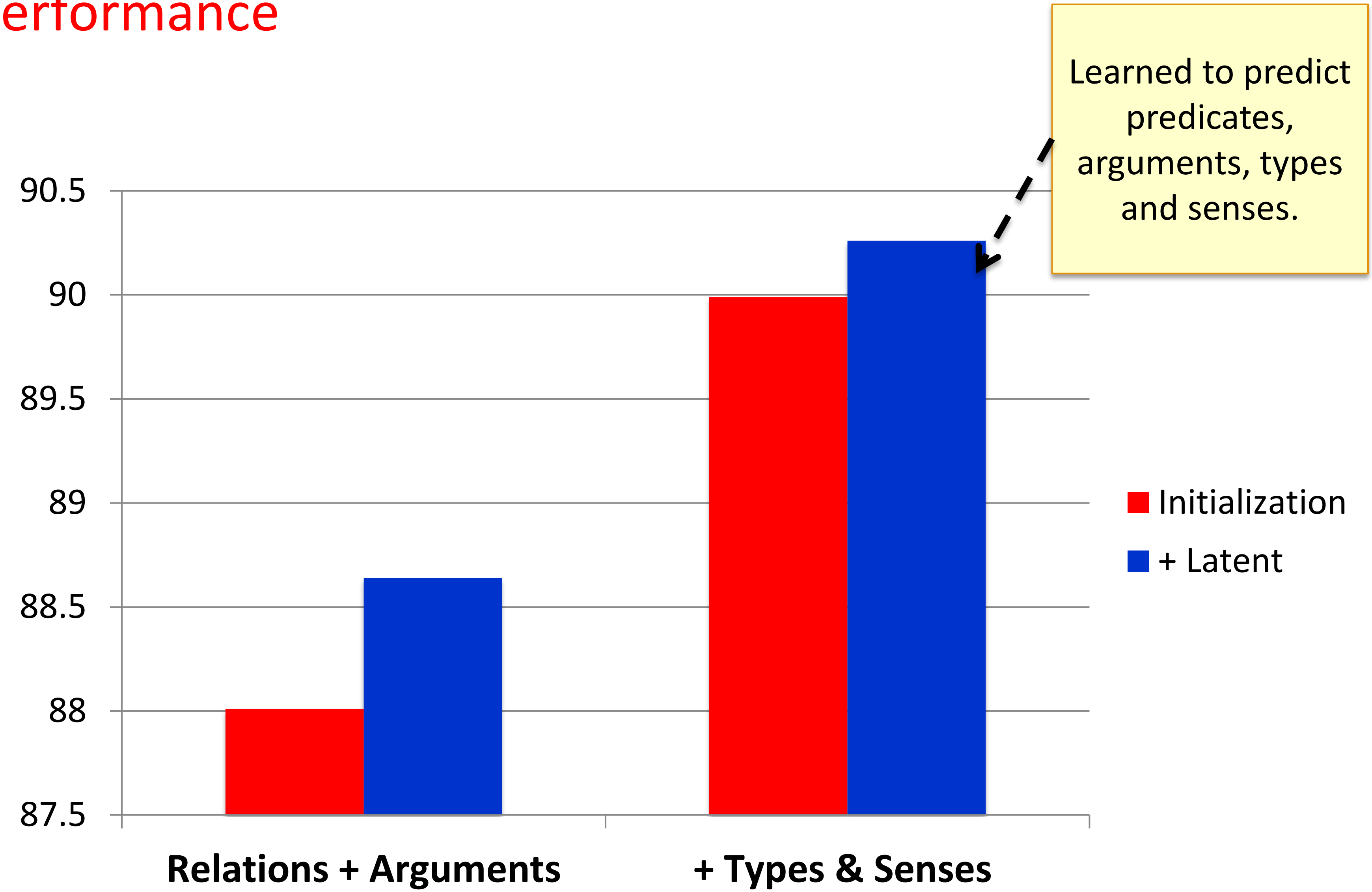
Performance



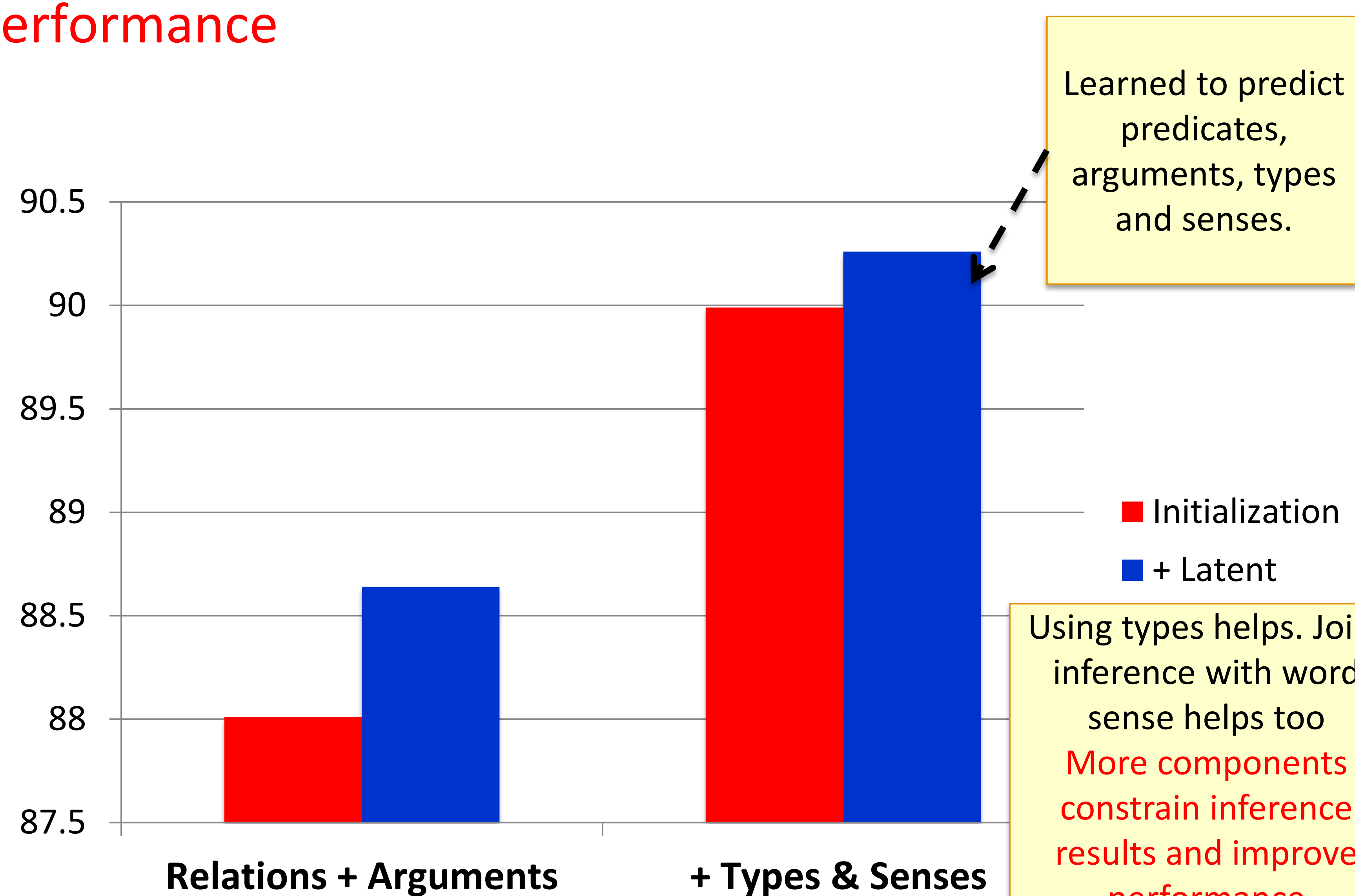
Performance



Performance



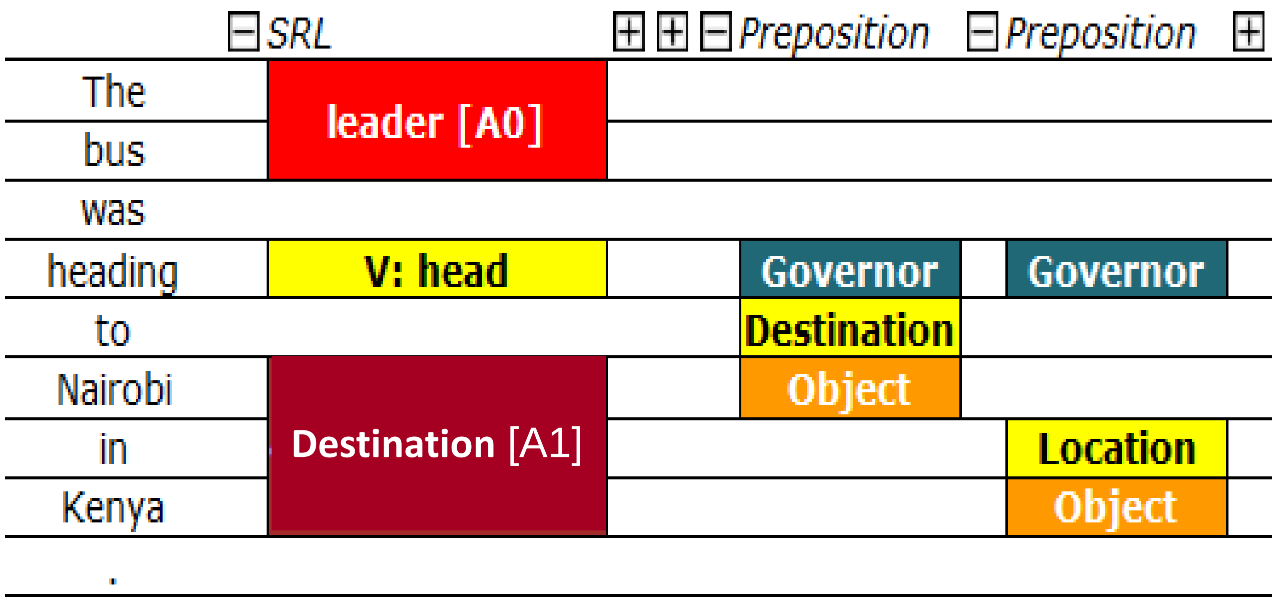
Performance



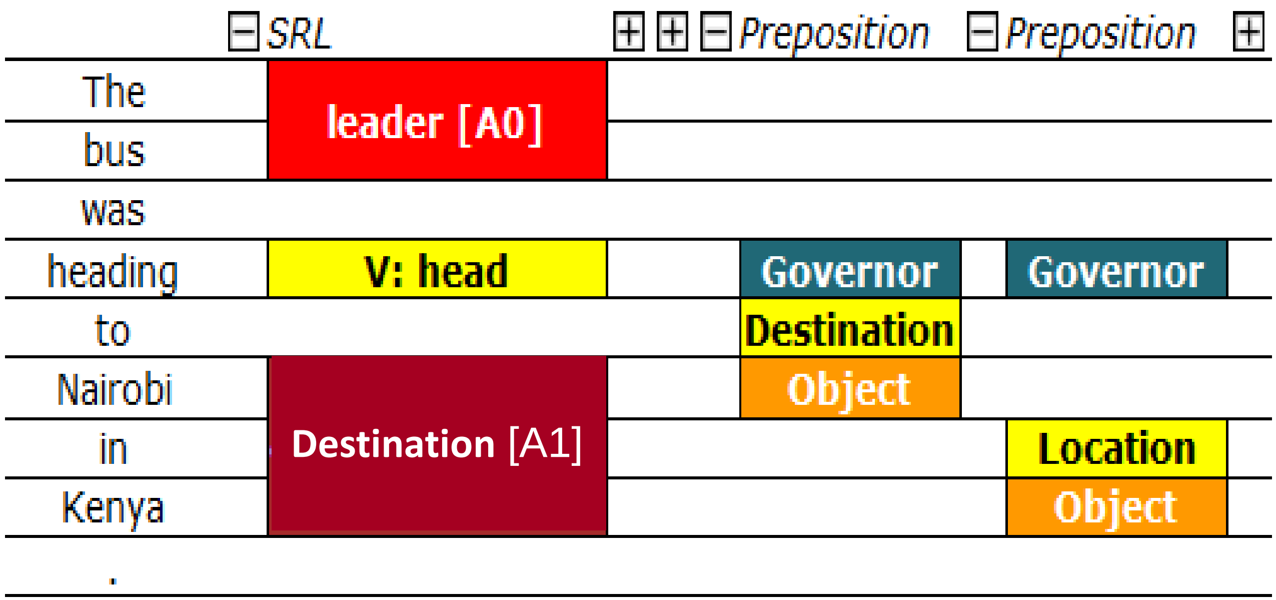
Using types helps. Joint inference with word sense helps too

More components constrain inference results and improve performance

Extended SRL [Demo]



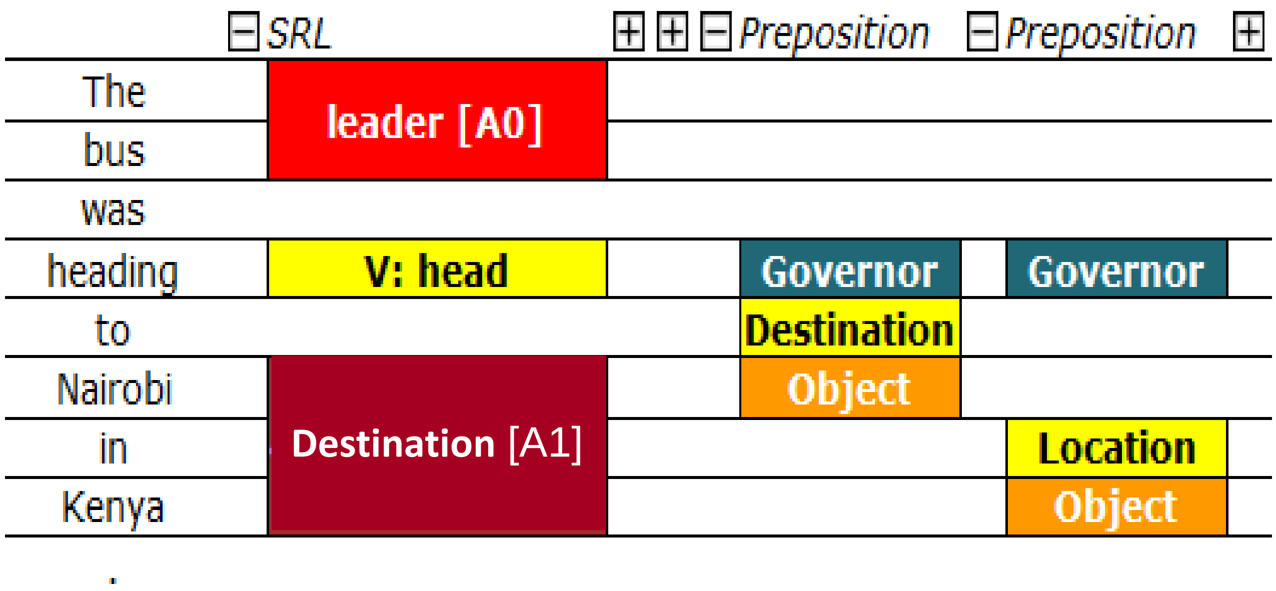
Extended SRL [Demo]



Joint inference over phenomena specific models to enforce consistency

Models trained with latent structure: senses, types, arguments

Extended SRL [Demo]



Joint inference over phenomena specific models to enforce consistency

Models trained with latent structure: senses, types, arguments

- More to do with other relations, discourse phenomena,…

Constrained Conditional Models—ILP Formulations

- **Have been shown useful in the context of many NLP problems**
- [Roth&Yih, 04,07: Entities and Relations; Punyakanok et. al: SRL …]
  - Summarization; Co-reference; Information & Relation Extraction; Event Identifications and causality ; Transliteration; Textual Entailment; Knowledge Acquisition; Sentiments; Temporal Reasoning, Parsing,…
- Some theoretical work on training paradigms [Punyakanok et. al., 05 more; Constraints Driven Learning, PR, Constrained EM…]
- Some work on Inference, mostly approximations, bringing back ideas on Lagrangian relaxation, etc.

Constrained Conditional Models—ILP Formulations

- **Have been shown useful in the context of many NLP problems**

- [Roth&Yih, 04,07: Entities and Relations; Punyakanok et. al: SRL …]
  - Summarization; Co-reference; Information & Relation Extraction; Event Identifications and causality ; Transliteration; Textual Entailment; Knowledge Acquisition; Sentiments; Temporal Reasoning, Parsing,…

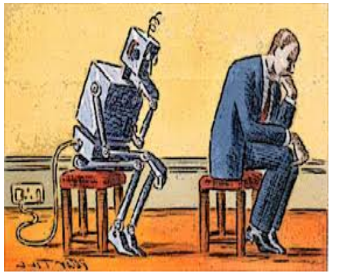
- Some theoretical work on training paradigms [Punyakanok et. al., 05 more; Constraints Driven Learning, PR, Constrained EM…]
- Some work on Inference, mostly approximations, bringing back ideas on Lagrangian relaxation, etc.

- Good summary and description of training paradigms: [Chang, Ratinov & Roth, Machine Learning Journal 2012]

- **Summary of work & a bibliography: http://L2R.cs.uiuc.edu/tutorials.html**

Outline

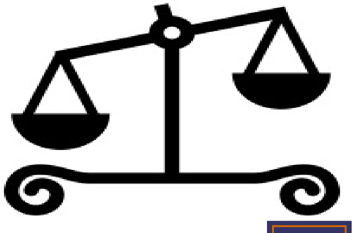
- Knowledge and Inference
  - Combining the soft with the logical/declarative nature of Natural Language
    - **Constrained Conditional Models: A formulation for global inference with knowledge modeled as expressive structural constraints**
    - **Some examples**



- Cycles of Knowledge
  - Grounding/Acquisition – knowledge – inference

- Learning with Indirect Supervision
  - Response Based Learning: learning from the world's feedback

- Scaling Up: Amortized Inference
  - Can the k-th inference problem be cheaper than the 1st?



Wikification: The Reference Problem

Blumenthal (D) is a candidate for the U.S. Senate seat now held by Christopher Dodd (D), and he has held a commanding lead in the race since he entered it. But the Times report has the potential to fundamentally reshape the contest in the Nutmeg State.

Wikification: The Reference Problem

Blumenthal (D) is a candidate for the U.S. Senate seat now held by Christopher Dodd (D), and he has held a commanding lead in the race since he entered it. But the Times report has the potential to fundamentally reshape the contest in the Nutmeg State.

Richard Blumenthal
From Wikipedia, the free encyclopedia

Democratic Party (United States)
From Wikipedia, the free encyclopedia

United States Senate
From Wikipedia, the free encyclopedia

Blumenthal (D) is a candidate for the U.S. Senate seat now held by Christopher Dodd (D), and he has held a commanding lead in the race since he entered it. But the Times report has the potential to fundamentally reshape the contest in the Nutmeg State.

Chris Dodd
From Wikipedia, the free encyclopedia

The New York Times
From Wikipedia, the free encyclopedia

Connecticut
From Wikipedia, the free encyclopedia

Wikification: The Reference Problem

Blumenthal (D) is a candidate for the U.S. Senate seat now held by Christopher Dodd (D), and he has held a commanding lead in the race since he entered it. But the Times report has the potential to fundamentally reshape the contest in the Nutmeg State.

Richard Blumenthal
From Wikipedia, the free encyclopedia

Democratic Party (United States)
From Wikipedia, the free encyclopedia

United States Senate
From Wikipedia, the free encyclopedia

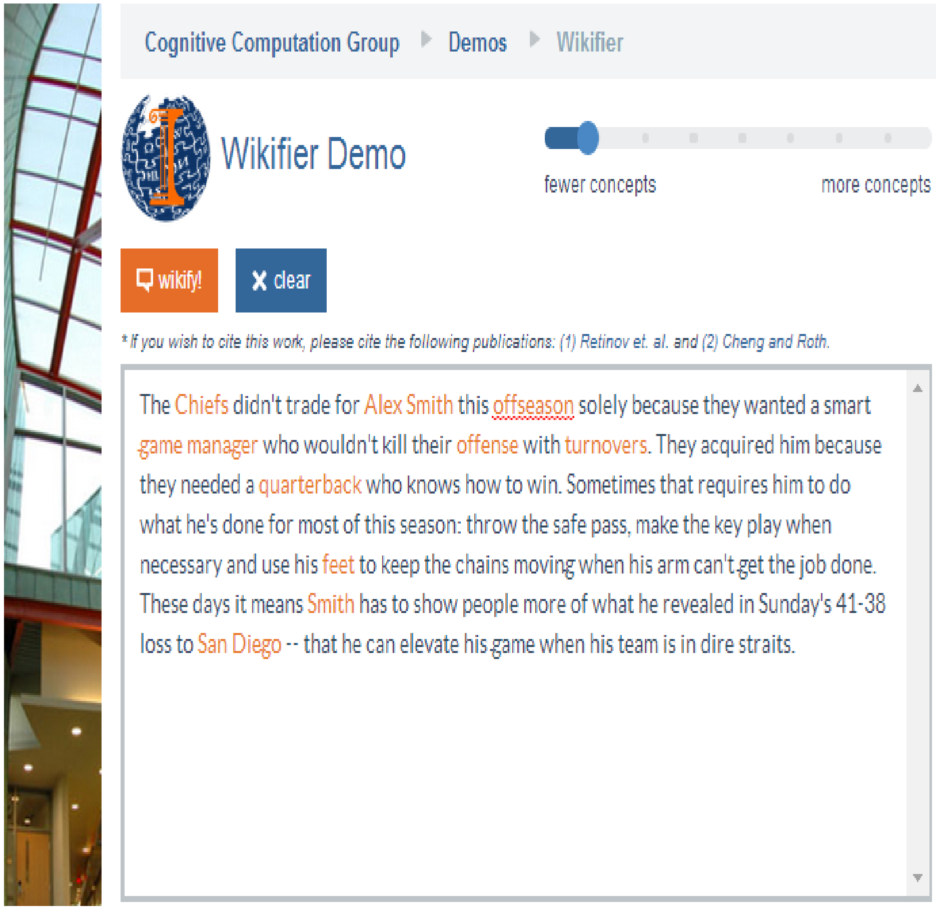
Blumenthal (D) is a candidate for the U.S. Senate seat now held by Christopher Dodd (D), and he has held a commanding lead in the race since he entered it. But the Times report has the potential to fundamentally reshape the contest in the Nutmeg State.

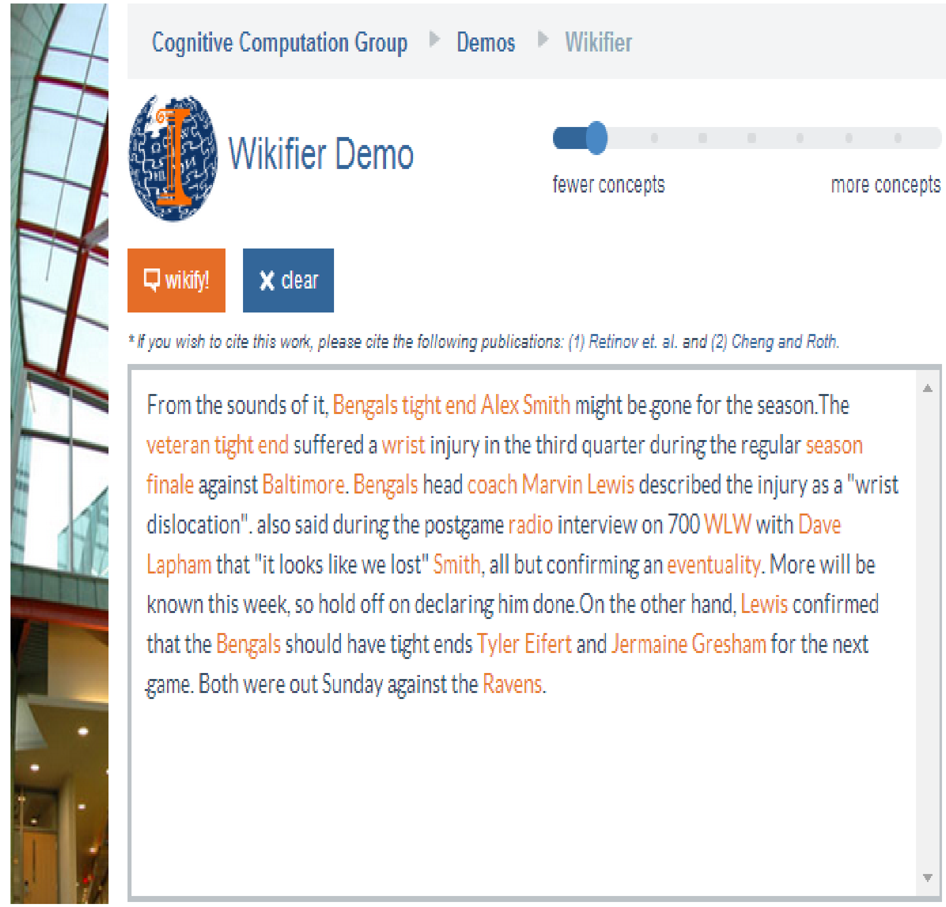
Chris Dodd
From Wikipedia, the free encyclopedia

The New York Times
From Wikipedia, the free encyclopedia

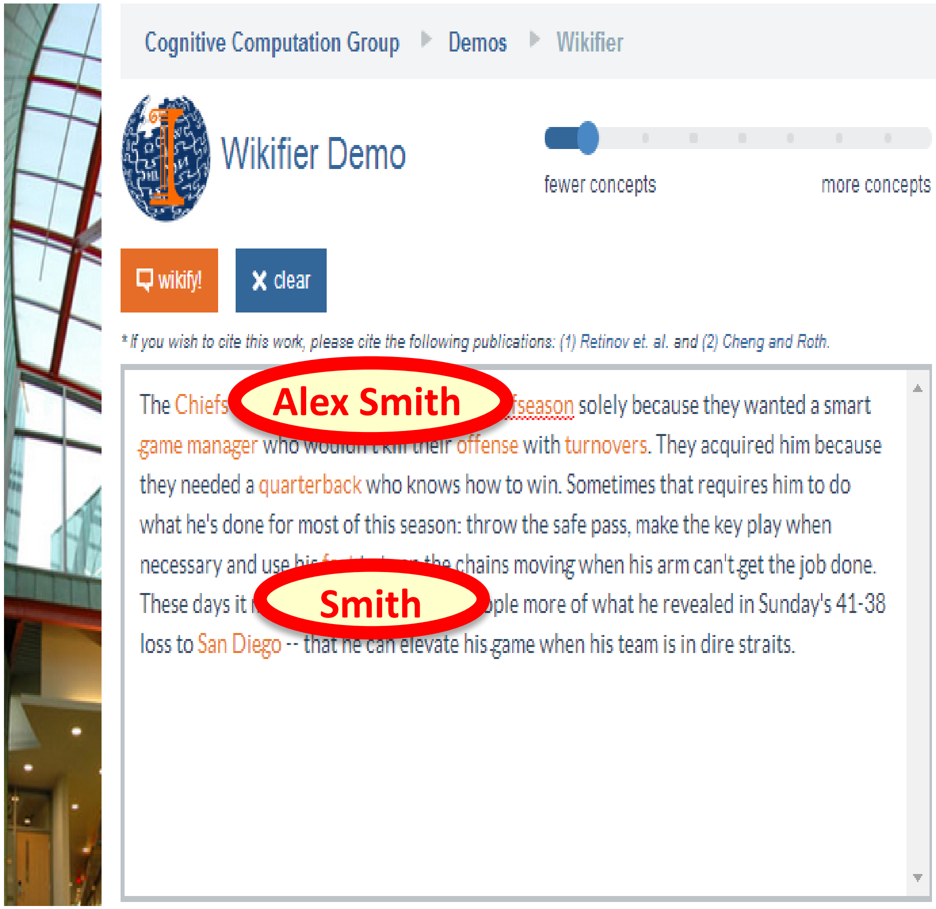
Connecticut
From Wikipedia, the free encyclopedia

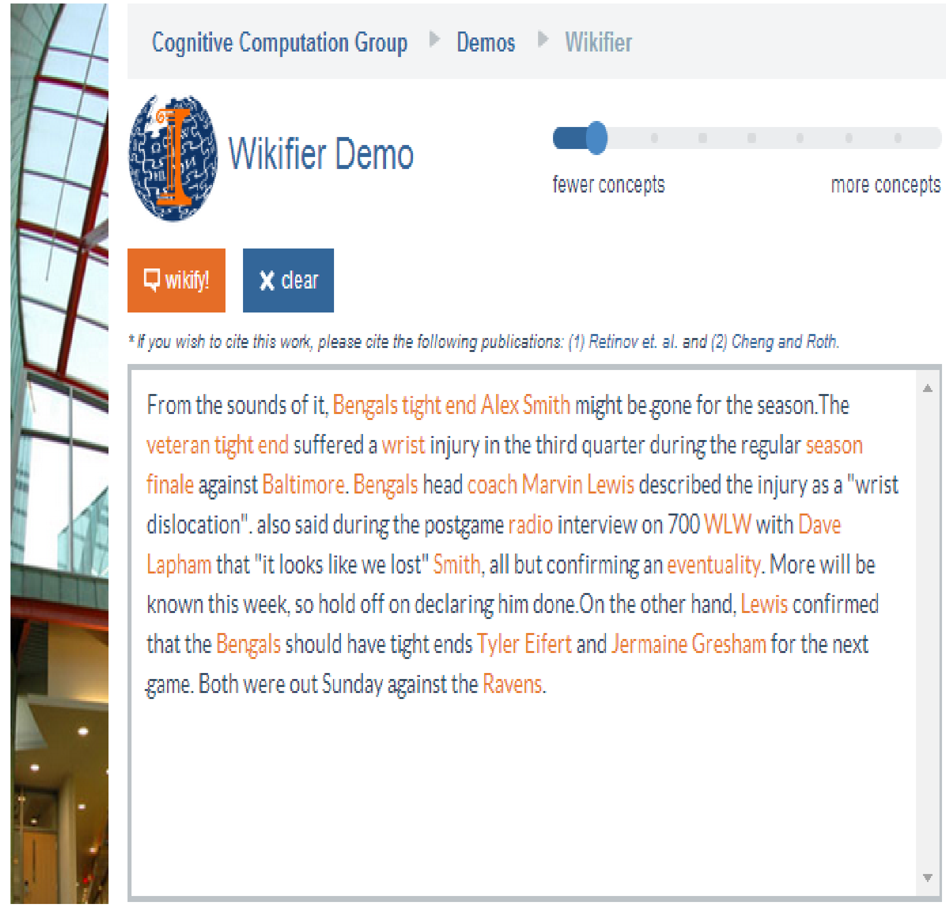
Who is Alex Smith?



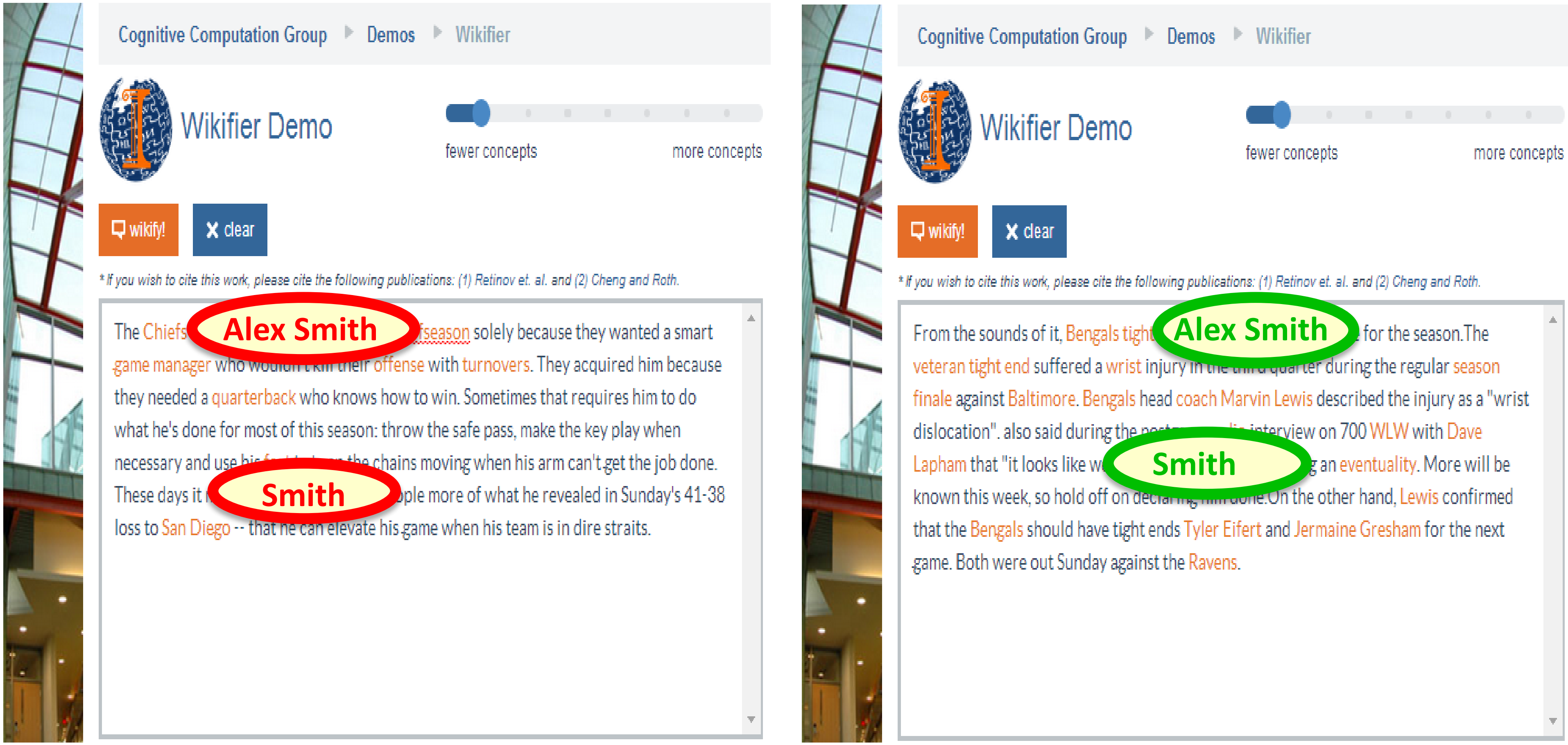


Who is Alex Smith?

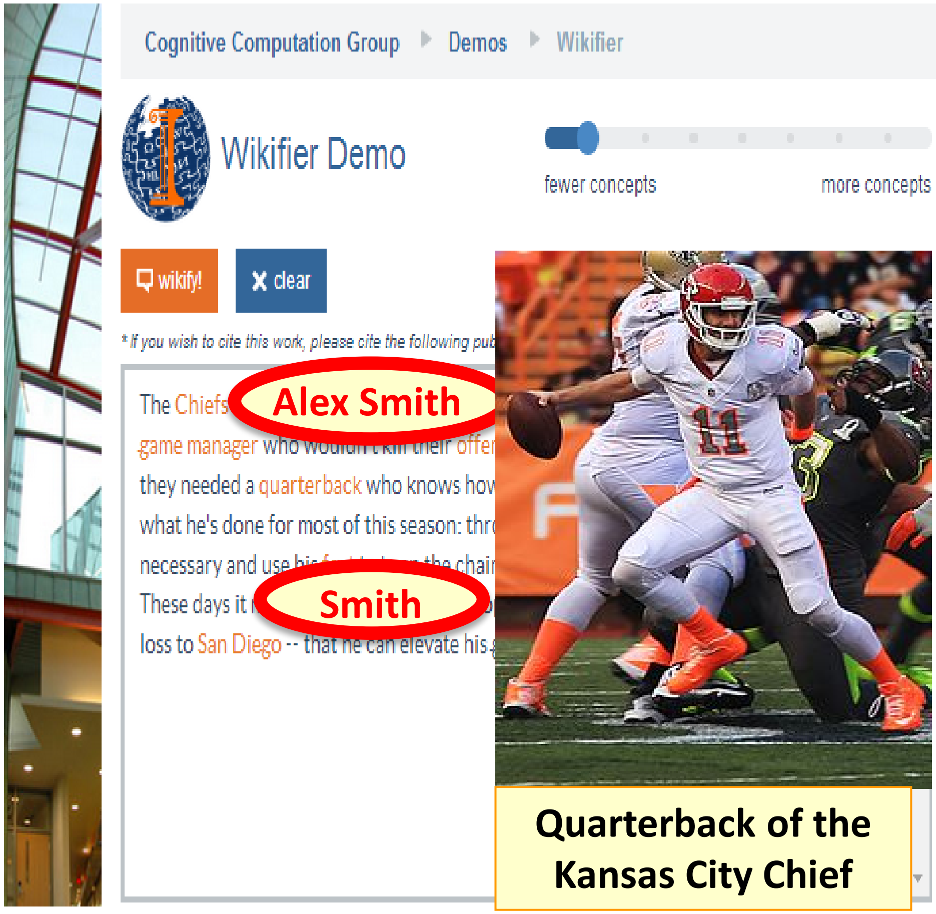


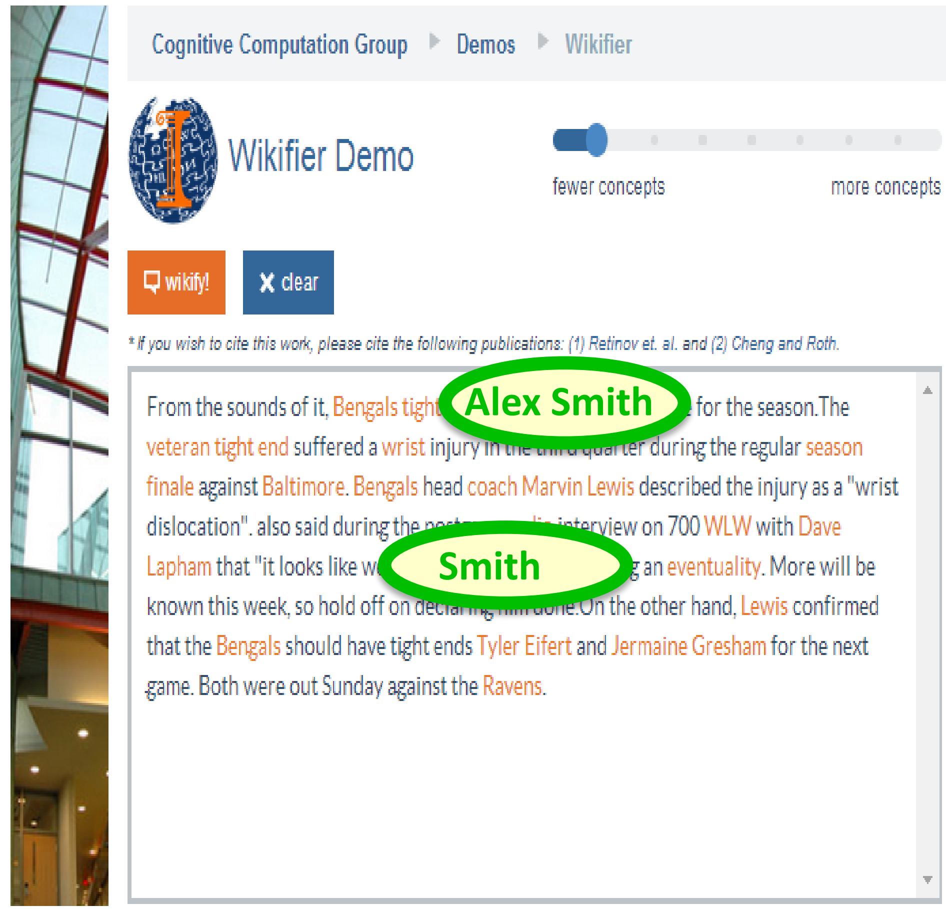


Who is Alex Smith?

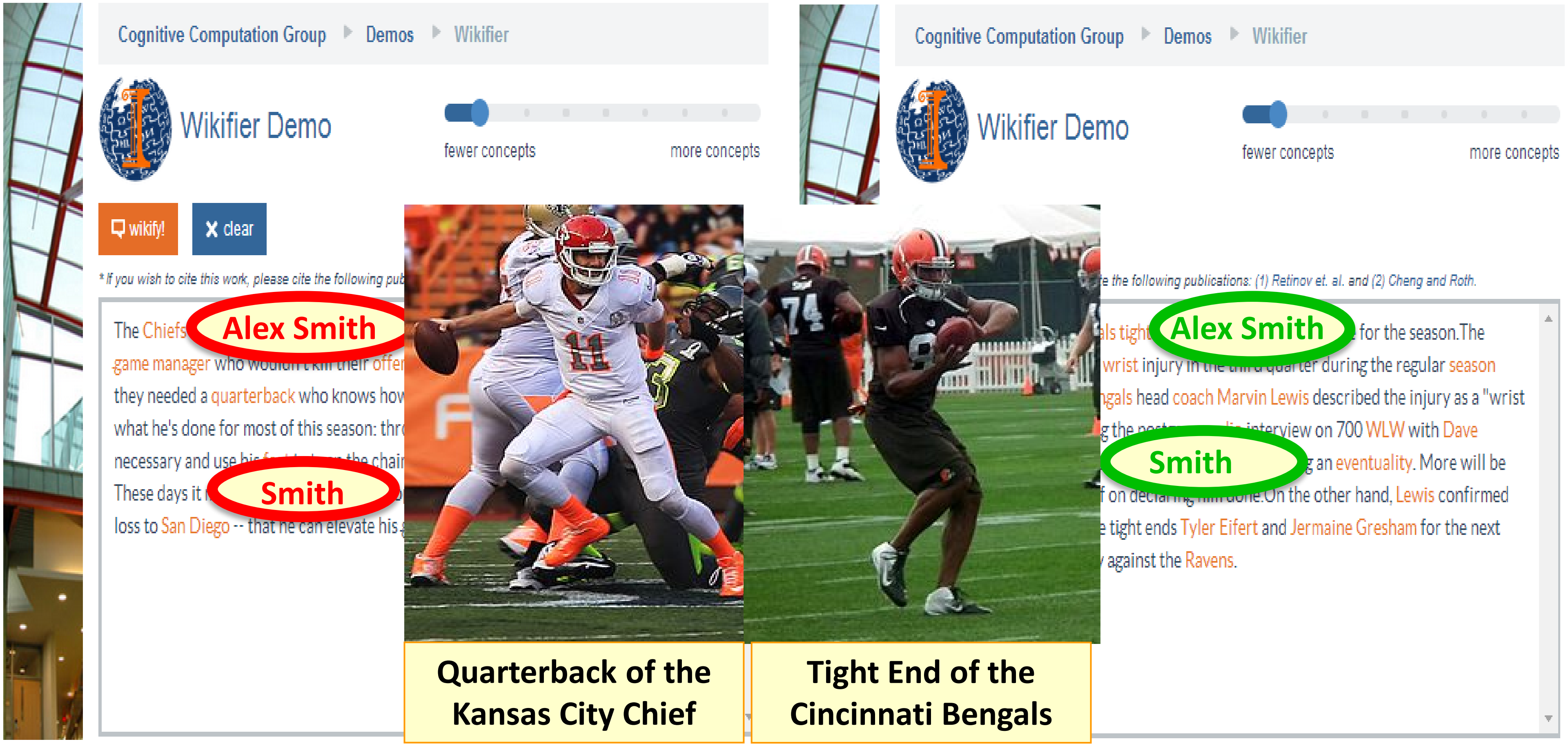


Who is Alex Smith?





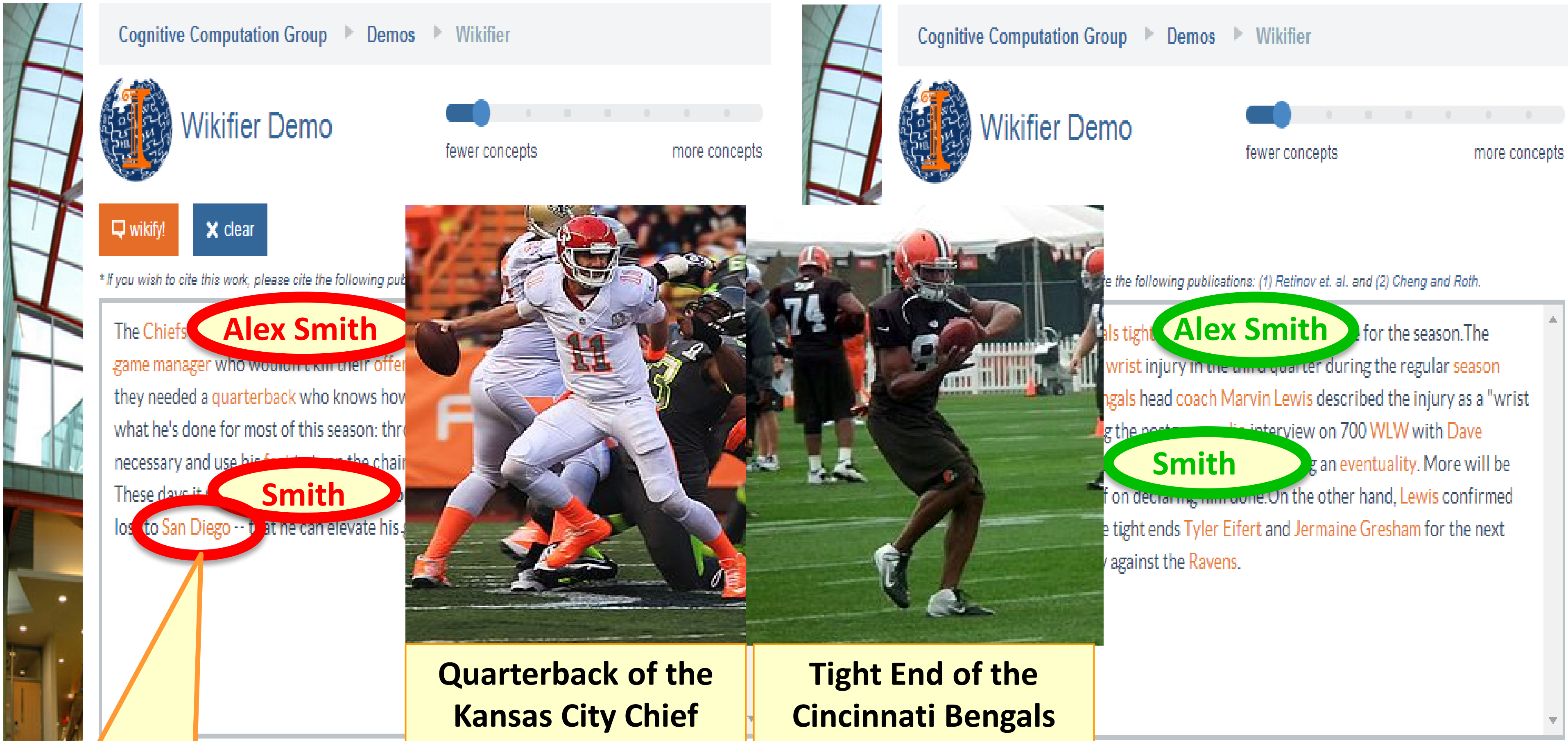
Who is Alex Smith?



Quarterback of the Kansas City Chief

Tight End of the Cincinnati Bengals

Who is Alex Smith?



San Diego: The San Diego Chargers (A Football team)

Who is Alex Smith?

Cognitive Computation Group ▸ Demos ▸ Wikifier

Wikifier Demo

fewer concepts — more concepts

wikify! | clear

* If you wish to cite this work, please cite the following pub... e the following publications: (1) Retinov et. al. and (2) Cheng and Roth.

The Chiefs **Alex Smith** ... game manager who wouldn't kill their offen... they needed a quarterback who knows how ... what he's done for most of this season: thr... necessary and use his ... the chai... These days it **Smith** ... lost to San Diego -- that he can elevate his ...

...als tight **Alex Smith** for the season.The ... wrist injury in the third quarter during the regular season ... ngals head coach Marvin Lewis described the injury as a "wrist ... g the post... interview on 700 WLW with Dave ... **Smith** ... an eventuality. More will be ... f on declaring him done.On the other hand, Lewis confirmed ... e tight end Tyler Eifert and Jermaine Gresham for the next ... y against the Ravens.

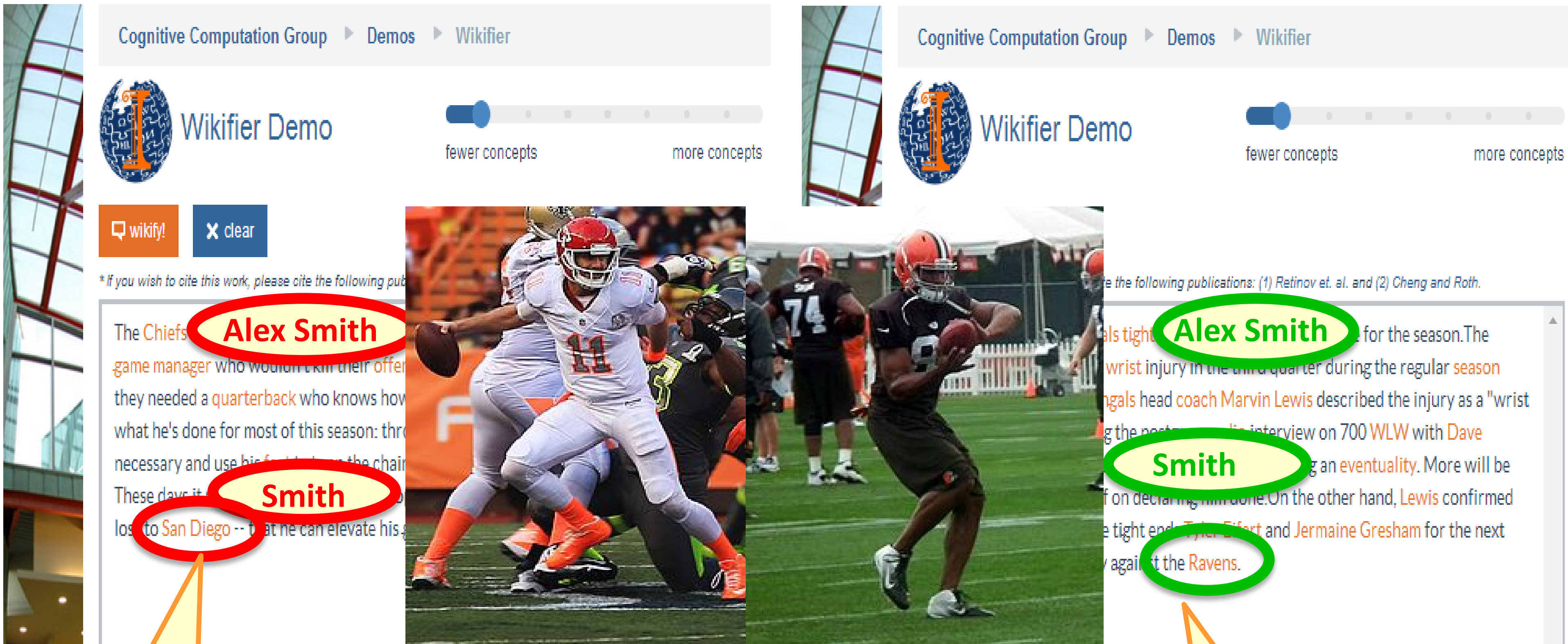
Quarterback of the Kansas City Chief

Tight End of the Cincinnati Bengals

San Diego: The San Diego Chargers (A Football team)

Ravens: The Baltimore Ravens (A Football team)

Who is Alex Smith?



Quarterback of the Kansas City Chief

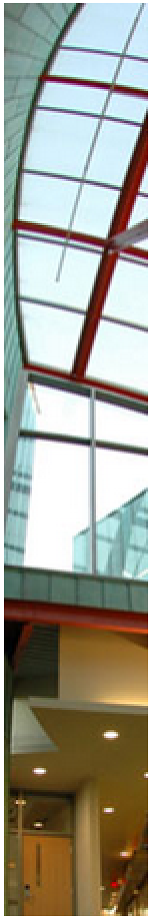
Tight End of the Cincinnati Bengals

San Diego: **The San Diego Chargers (A Football team)**

Contextual Disambiguation

Ravens: **The Baltimore Ravens (A Football team)**

Middle Eastern Politics



Cognitive Computation Group ▶ Demos ▶ Wikifier

Wikifier Demo

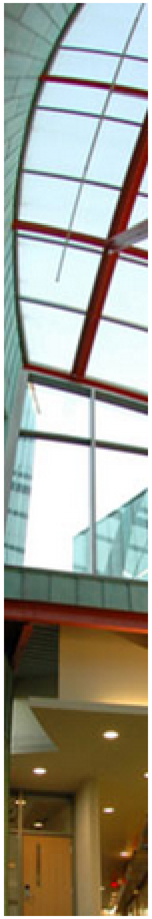
fewer concepts   more concepts

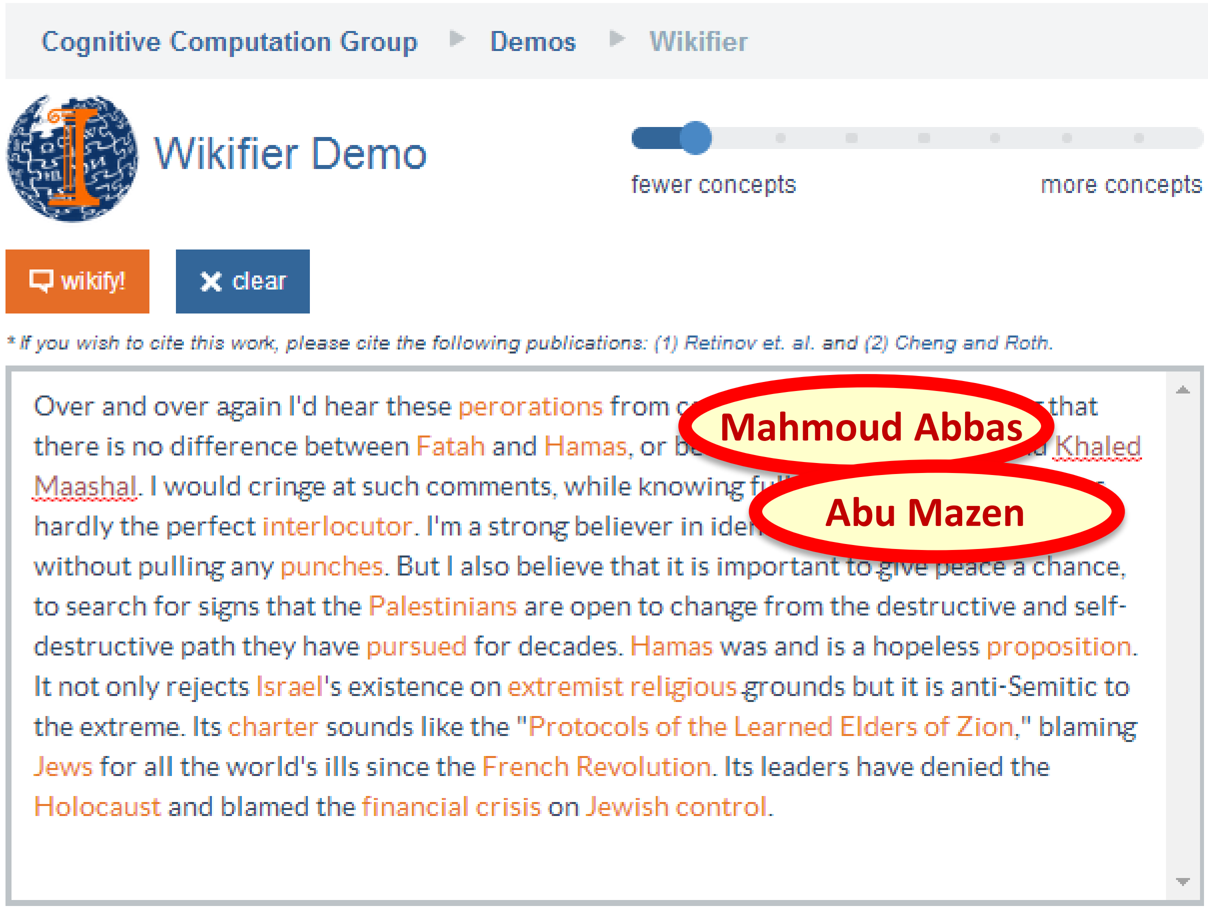
wikify!   clear

* If you wish to cite this work, please cite the following publications: (1) Retinov et. al. and (2) Cheng and Roth.

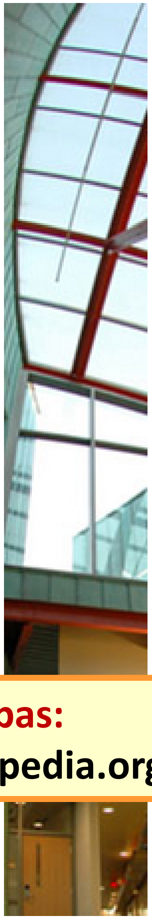
Over and over again I'd hear these perorations from certain Jewish circles arguing that there is no difference between Fatah and Hamas, or between Mahmoud Abbas and Khaled Maashal. I would cringe at such comments, while knowing full well that Abu Mazen was hardly the perfect interlocutor. I'm a strong believer in identifying the threats to Israel without pulling any punches. But I also believe that it is important to give peace a chance, to search for signs that the Palestinians are open to change from the destructive and self-destructive path they have pursued for decades. Hamas was and is a hopeless proposition. It not only rejects Israel's existence on extremist religious grounds but it is anti-Semitic to the extreme. Its charter sounds like the "Protocols of the Learned Elders of Zion," blaming Jews for all the world's ills since the French Revolution. Its leaders have denied the Holocaust and blamed the financial crisis on Jewish control.

Middle Eastern Politics





Middle Eastern Politics



Cognitive Computation Group ▶ Demos ▶ Wikifier

Wikifier Demo

fewer concepts   more concepts

wikify!   clear

* If you wish to cite this work, please cite the following publications: (1) Retinov et. al. and (2) Cheng and Roth.

Over and over again I'd hear these perorations from c… that there is no difference between Fatah and Hamas, or be… Khaled Maashal. I would cringe at such comments, while knowing f… hardly the perfect interlocutor. I'm a strong bel… in iden… without pulling any punches. But I al… e that it is important to give peace a chance, to search for signs that … ans are open to change from the des… uctive and self-destructive … pursued for decades. Hamas was and is a ho… ess proposition. … e on extremist religious grounds b… anti-Semitic to … e the "Protocols of the Learned E… Zion," blaming … e French Revolution. Its leade… nied the Holocaust and blamed the financial crisis on Jewish control.

Mahmoud Abbas

Abu Mazen

Mahmoud Abbas:
http://en.wikipedia.org/wiki/Mahmoud_Abbas

Abu Mazen:
http://en.wikipedia.org/wiki/Mahmoud_Abbas

Middle Eastern Politics

Mahmoud Abbas

Abu Mazen

Mahmoud Abbas:
http://en.wikipedia.org/wiki/Mahmoud_Abbas

Variability: Getting around multiple **surface representations. Co-reference resolution** within & across documents, **with grounding**

Abu Mazen:
http://en.wikipedia.org/wiki/Mahmoud_Abbas

Knowledge Acquisition requires Knowledge (& inference)

- Mubarak,

Knowledge Acquisition requires Knowledge (& inference)

- Mubarak,

Knowledge Acquisition requires Knowledge (& inference)

- Mubarak,

Knowledge Acquisition requires Knowledge (& inference)

- Mubarak, the wife of deposed Egyptian President Hosni Mubarak,…

COGNITIVE COMPUTATION GROUP
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Knowledge Acquisition requires Knowledge (& inference)

- Mubarak, the wife of deposed Egyptian President Hosni Mubarak,…

Knowledge Acquisition requires Knowledge (& inference)

- Mubarak, the wife of deposed Egyptian President Hosni Mubarak,…

Knowledge Acquisition requires Knowledge (& inference)

- Mubarak, the wife of deposed Egyptian President Hosni Mubarak,…

Relational Inference

Mubarak, the wife of deposed Egyptian President Hosni Mubarak , …

Relational Inference

wife Egyptian President

Mubarak Hosni Mubarak

Relational Inference

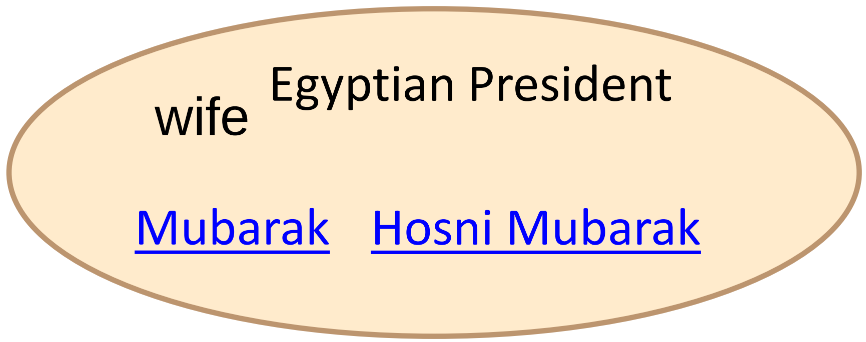
wife Egyptian President

Mubarak Hosni Mubarak

- What are we missing with Bag of Words (BOW) models?
  - Who is Mubarak?

Relational Inference

Mubarak, the wife of deposed Egyptian President Hosni Mubarak, …



- What are we missing with Bag of Words (BOW) models?
  - Who is Mubarak?
- Textual relations provide another dimension of text understanding
- Can be used to constrain interactions between concepts
  - (Mubarak, wife, Hosni Mubarak)
- Has impact in several steps in the Wikification process:
  - From candidate selection to ranking and global decision

Formulation

- Goal: Promote concepts that are coherent with textual relations
- Formulate as an Integer Linear Program (ILP):

$$\Gamma_D = \underbrace{\arg\max_{\Gamma} \sum_i \sum_k s_i^k e_i^k}_{} + \sum_{i,j} \sum_{k,l} w_{ij}^{(k,l)} r_{ij}^{(k,l)}$$

$$\begin{array}{lll} s.t. & r_{ij}^{(k,l)} \in \{0,1\} & \text{Integral constraints} \\ & e_i^k \in \{0,1\} & \text{Integral constraints} \\ & \forall i \sum_k e_i^k = 1 & \text{Unique solution} \\ & 2r_{ij}^{(k,l)} \leq e_i^k + e_j^l & \text{Relation definition} \end{array}$$

- If no relation exists, collapses to the non-structured decision

Formulation

- Goal: Promote concepts that are coherent with textual relations
- Formulate as an Integer Linear Program (ILP):

weight of variable e_i^k

Boolean variable: kth candidate corresponds to ith mention (or not)

$$\Gamma_D = \underline{\arg\max_{\Gamma} \sum_i \sum_k s_i^k e_i^k} + \sum_{i,j} \sum_{k,l} w_{ij}^{(k,l)} r_{ij}^{(k,l)}$$

$$\begin{array}{lll} s.t. & r_{ij}^{(k,l)} \in \{0,1\} & \text{Integral constraints} \\ & e_i^k \in \{0,1\} & \text{Integral constraints} \\ & \forall i \sum_k e_i^k = 1 & \text{Unique solution} \\ & 2r_{ij}^{(k,l)} \leq e_i^k + e_j^l & \text{Relation definition} \end{array}$$

- If no relation exists, collapses to the non-structured decision

Formulation

- Goal: Promote concepts that are coherent with textual relations
- Formulate as an Integer Linear Program (ILP):

weight of variable e_i^k

Boolean variable: kth candidate corresponds to ith mention (or not)

$$\Gamma_D = \arg\max_{\Gamma} \sum_i \sum_k s_i^k e_i^k + \sum_{i,j} \sum_{k,l} w_{ij}^{(k,l)} r_{ij}^{(k,l)}$$

weight of a relation $r_{ij}^{(k,l)}$

Boolean variable: a relation exists between e_i^k and e_j^l (or not)

$s.t.$	$r_{ij}^{(k,l)} \in \{0,1\}$	Integral constraints
	$e_i^k \in \{0,1\}$	Integral constraints
	$\forall i \sum_k e_i^k = 1$	Unique solution
	$2r_{ij}^{(k,l)} \leq e_i^k + e_j^l$	Relation definition

- If no relation exists, collapses to the non-structured decision

Formulation

Knowledge + Ability to use it (Inference) facilitates additional knowledge acquisition

- Goal: Promote concepts that are coherent with textual relations
- Formulate as an Integer Linear Program (ILP):

weight of variable e_i^k

Boolean variable: kth candidate corresponds to ith mention (or not)

$$\Gamma_D = \arg\max_{\Gamma} \sum_i \sum_k s_i^k e_i^k + \sum_{i,j} \sum_{k,l} w_{ij}^{(k,l)} r_{ij}^{(k,l)}$$

weight of a relation $r_{ij}^{(k,l)}$

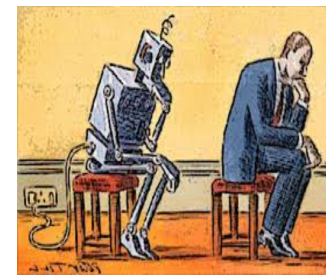
Boolean variable: a relation exists between e_i^k and e_j^l (or not)

$s.t.$	$r_{ij}^{(k,l)} \in \{0,1\}$	Integral constraints
	$e_i^k \in \{0,1\}$	Integral constraints
	$\forall i \sum_k e_i^k = 1$	Unique solution
	$2r_{ij}^{(k,l)} \leq e_i^k + e_j^l$	Relation definition

- If no relation exists, collapses to the non-structured decision

Outline

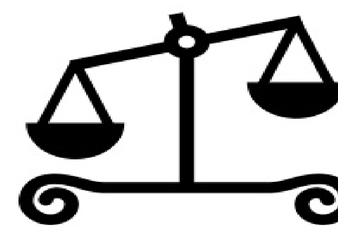
- Knowledge and Inference
  - Combining the soft with the logical/declarative nature of Natural Language
    - **Constrained Conditional Models: A formulation for global inference with knowledge modeled as expressive structural constraints**
    - **Some examples**



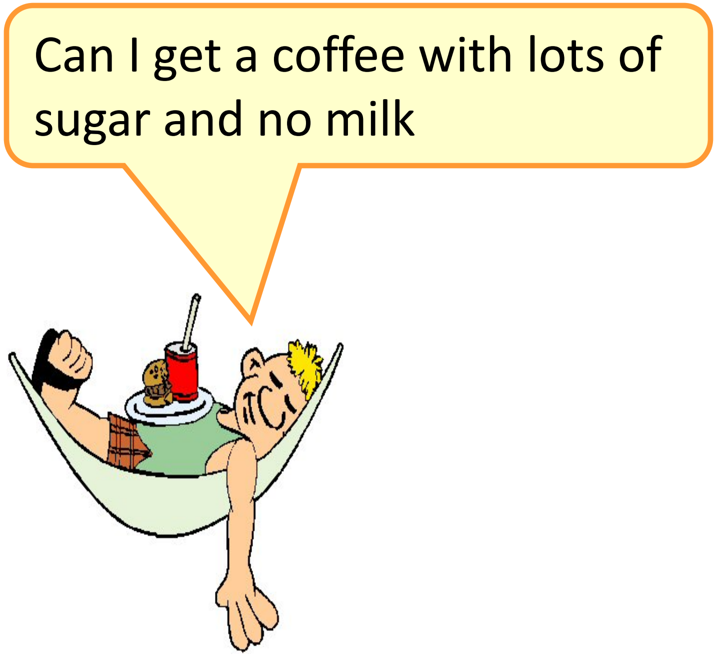
- Cycles of Knowledge
  - Grounding/Acquisition – knowledge – inference

- Learning with Indirect Supervision
  - Response Based Learning: learning from the world’s feedback

- Scaling Up: Amortized Inference
  - Can the k-th inference problem be cheaper than the 1st?

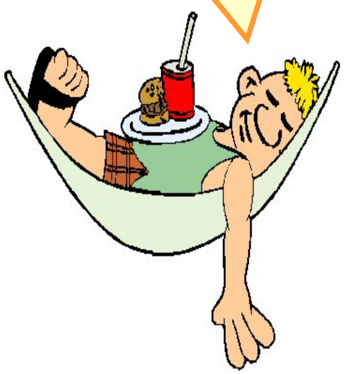


Understanding Language Requires Supervision



Understanding Language Requires Supervision

Can I get a coffee with lots of sugar and no milk

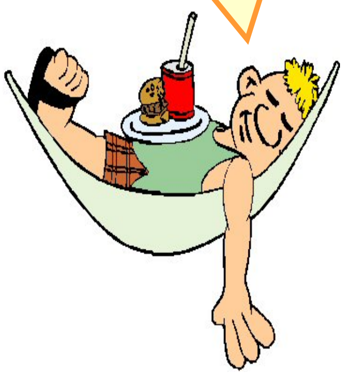


Semantic Parser

MAKE(COFFEE,SUGAR=YES,MILK=NO)

Understanding Language Requires Supervision

Can I get a coffee with lots of sugar and no milk



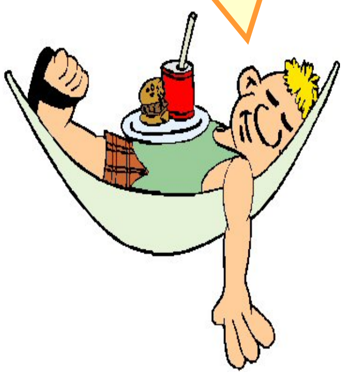
Semantic Parser

MAKE(COFFEE,SUGAR=YES,MILK=NO)

- How to recover meaning from text?

Understanding Language Requires Supervision

Can I get a coffee with lots of sugar and no milk



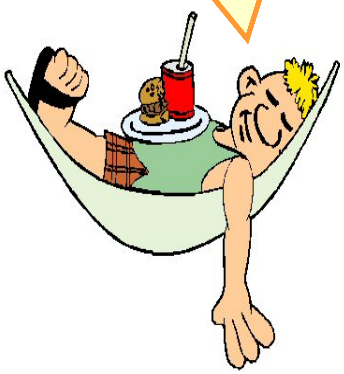
Semantic Parser

MAKE(COFFEE,SUGAR=YES,MILK=NO)

- How to recover meaning from text?
- Standard "example based" ML: annotate text with meaning representation
  - Teacher needs deep understanding of the learning agent ; not scalable.

Understanding Language Requires Supervision

Can I get a coffee with lots of sugar and no milk

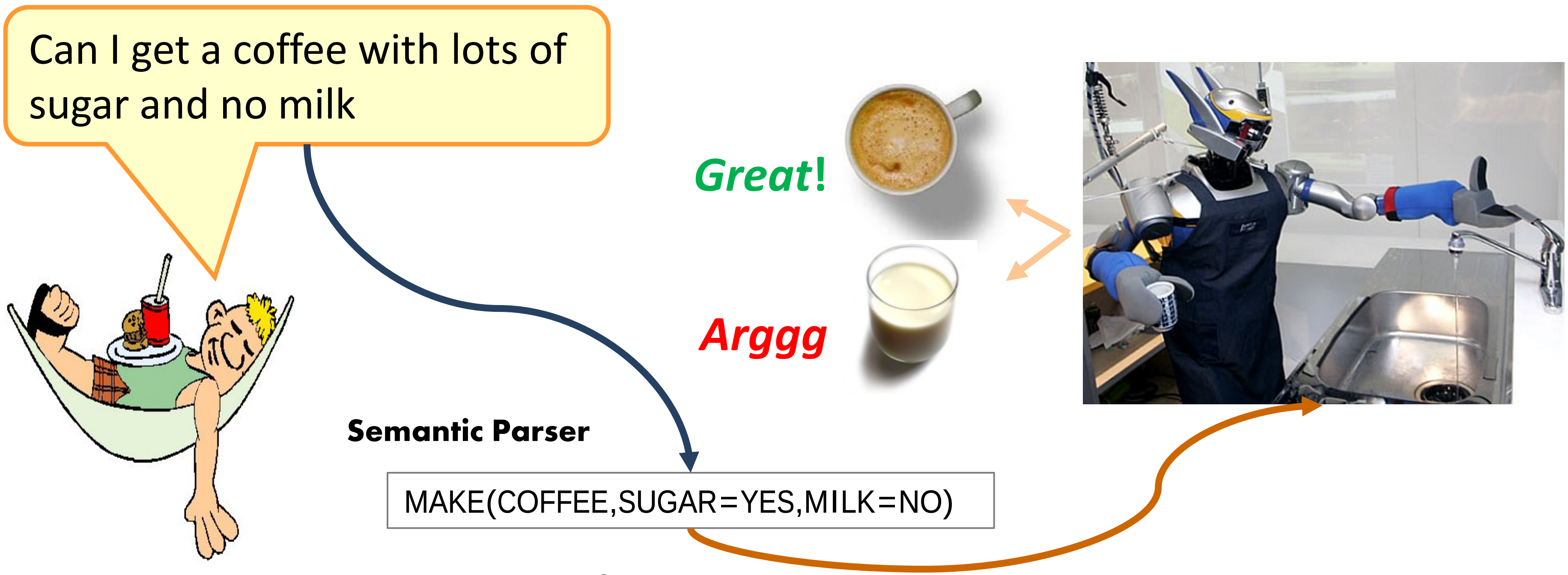


Semantic Parser

MAKE(COFFEE,SUGAR=YES,MILK=NO)

- How to recover meaning from text?
- Standard "example based" ML: annotate text with meaning representation
  - Teacher needs deep understanding of the learning agent ; not scalable.
- Response Driven Learning: Exploit indirect signals in the interaction between the learner and the teacher/environment

Understanding Language Requires Supervision

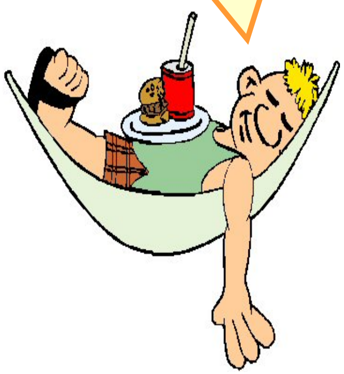


- How to recover meaning from text?
- Standard "example based" ML: annotate text with meaning representation
  - Teacher needs deep understanding of the learning agent ; not scalable.
- Response Driven Learning: Exploit indirect signals in the interaction between the learner and the teacher/environment

Understanding Language Requires Supervision

Can we rely on this interaction to provide supervision (and eventually, recover meaning) ?

Can I get a coffee with lots of sugar and no milk



Great!

Arggg

Semantic Parser

MAKE(COFFEE,SUGAR=YES,MILK=NO)

- How to recover meaning from text?
- Standard "example based" ML: annotate text with meaning representation
  - Teacher needs deep understanding of the learning agent ; not scalable.
- Response Driven Learning: Exploit indirect signals in the interaction between the learner and the teacher/environment

Response Based Learning

- We want to learn a model that transforms a natural language sentence to some meaning representation.

English Sentence → **Model** → Meaning Representation

- Instead of training with (Sentence, Meaning Representation) pairs

Response Based Learning

- We want to learn a model that transforms a natural language sentence to some meaning representation.

English Sentence → **Model** → Meaning Representation

- Instead of training with (Sentence, Meaning Representation) pairs
- Think about some simple derivatives of the models outputs,
  - Supervise the derivative [verifier] (easy!) and
  - Propagate it to learn the complex, structured, transformation model

Scenario I: Freecell with Response Based Learning

- We want to learn a model to transform a natural language sentence to some meaning representation.

English Sentence → **Model** → Meaning Representation

English Sentence	Meaning Representation
A top card can be moved to the tableau if it has a different color than the color of the top tableau card, and the cards have successive values.	Move (a1,a2) top(a1,x1) card(a1) tableau(a2) top(x2,a2) color(a1,x3) color(x2,x4) not-equal(x3,x4) value(a1,x5) value(x2,x6) successor(x5,x6)

Scenario I: Freecell with Response Based Learning

- We want to learn a model to transform a natural language sentence to some meaning representation.

English Sentence → **Model** → Meaning Representation

A top card can be moved to the tableau if it has a different color than the color of the top tableau card, and the cards have successive values.

Move (a1,a2) top(a1,x1) card(a1) tableau(a2) top(x2,a2) color(a1,x3) color(x2,x4) not-equal(x3,x4) value(a1,x5) value(x2,x6) successor(x5,x6)

- Simple derivatives of the models outputs: game API
  - Supervise the derivative and
  - Propagate it to learn the transformation model

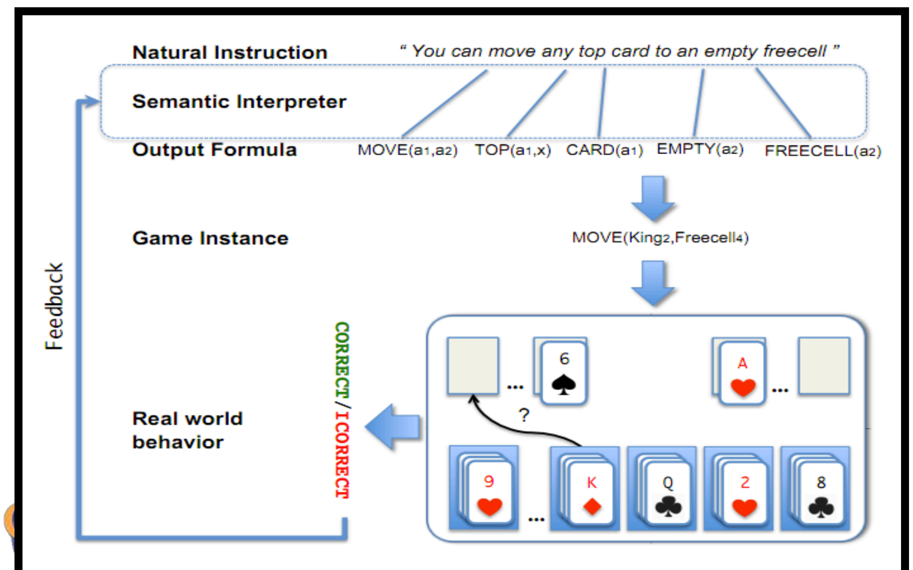
Scenario I: Freecell with Response Based Learning

- We want to learn a model to transform a natural language sentence to some meaning representation.

English Sentence → **Model** → Meaning Representation

A top card can be moved to the tableau if it has a different color than the color of the top tableau card, and the cards have successive values.

Move (a1,a2) top(a1,x1) card(a1) tableau(a2) top(x2,a2) color(a1,x3) color(x2,x4) not-equal(x3,x4) value(a1,x5) value(x2,x6) successor(x5,x6)



← Play Freecell (solitaire)

- Simple derivatives of the models outputs: game API
  - Supervise the derivative and
  - Propagate it to learn the transformation model

Scenario II: Geoquery with Response based Learning

- We want to learn a model to transform a natural language sentence to some formal representation.

English Sentence → **Model** → Meaning Representation

What is the largest state that borders NY? → largest(state(next_to(const(NY))))

Scenario II: Geoquery with Response based Learning

- We want to learn a model to transform a natural language sentence to some formal representation.

English Sentence → **Model** → Meaning Representation

What is the largest state that borders NY?

largest(state(next_to(const(NY))))

- Simple derivatives of the models outputs

Scenario II: Geoquery with Response based Learning

- We want to learn a model to transform a natural language sentence to some formal representation.

English Sentence → **Model** → Meaning Representation

What is the largest state that borders NY?

largest(state(next_to(const(NY))))

- Query a GeoQuery Database.

- Simple derivatives of the models outputs

Scenario II: Geoquery with Response based Learning

- We want to learn a model to transform a natural language sentence to some formal representation.

English Sentence → **Model** → Meaning Representation

What is the largest state that borders NY?

largest(state(next_to(const(NY))))

- Query a GeoQuery Database. ← - Simple derivatives of the models outputs

- "Guess" a semantic parse. **Is [DB response == Expected response] ?**
  - **Expected**: Pennsylvania **DB Returns**: Pennsylvania→ **Positive Response**
  - **Expected**: Pennsylvania **DB Returns**: NYC, or ???? → **Negative Response**

Response Based Learning: Using a Simple Feedback

- We want to learn a model to transform a natural language sentence to some formal representation.

English Sentence → **Model** → Meaning Representation

- Instead of training with (Sentence, Meaning Representation) pairs
- Think about some simple derivatives of the models outputs,
  - Supervise the derivative (easy!) and
  - Propagate it to learn the complex, structured, transformation model

Response Based Learning: Using a Simple Feedback

- We want to learn a model to transform a natural language sentence to some formal representation.

English Sentence → **Model** → Meaning Representation

- Instead of training with (Sentence, Meaning Representation) pairs
- Think about some simple derivatives of the models outputs,
  - Supervise the derivative (easy!) and
  - Propagate it to learn the complex, structured, transformation model

LEARNING:

- Train a structured predictor (semantic parse) with this binary supervision
  - Many challenges: e.g., how to make a better use of a negative response?
- Learning with a constrained latent representation, making used of CCM inference, exploiting knowledge on the structure of the meaning representation.

Geoquery: Response based Competitive with Supervised

Clarke, Goldwasser, Chang, Roth CoNLL'10; Goldwasser, Roth IJCAI'11, MLJ'14

Algorithm	Training Accuracy	Testing Accuracy	# Training Examples
NoLearn	22	--	-
Supervised	--	86.07	**600 structs.**

NoLearn :Initialization point

Supervised : Trained with annotated data

Response based Learning is gathering momentum:

- Liang, M.I. Jordan, D. Klein, Learning Dependency-Based Compositional Semantics, ACL'11.
- Berant et-al ' Semantic Parsing on Freebase from Question-Answer Pairs, EMNLP'13

Supervised: Y.-W. Wong and R. Mooney. Learning synchronous grammars for semantic parsing with lambda calculus. ACL'07

Geoquery: Response based Competitive with Supervised

Clarke, Goldwasser, Chang, Roth CoNLL'10; Goldwasser, Roth IJCAI'11, MLJ'14

Algorithm	Training Accuracy	Testing Accuracy	# Training Examples
NoLearn	22	--	-
Response-based (2010)	82.4	73.2	250 answers
Liang et-al 2011	--	78.9	250 answers
Response-based (2012)	**86.8**	**81.6**	250 answers
Supervised	--	86.07	**600 structs**.

NoLearn :Initialization point

Supervised : Trained with annotated data

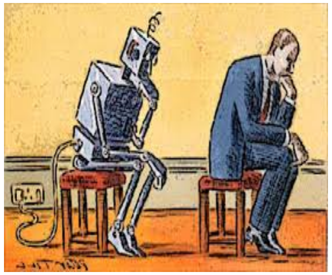
Response based Learning is gathering momentum:

- Liang, M.I. Jordan, D. Klein, Learning Dependency-Based Compositional Semantics, ACL'11.
- Berant et-al ' Semantic Parsing on Freebase from Question-Answer Pairs, EMNLP'13

Supervised: Y.-W. Wong and R. Mooney. Learning synchronous grammars for semantic parsing with lambda calculus. ACL'07

Outline

- Knowledge and Inference
  - Combining the soft with the logical/declarative nature of Natural Language
    - **Constrained Conditional Models: A formulation for global inference with knowledge modeled as expressive structural constraints**
    - **Some examples**



- Cycles of Knowledge
  - Grounding/Acquisition – knowledge – inference

- Learning with Indirect Supervision
  - Response Based Learning: learning from the world's feedback

- Scaling Up: Amortized Inference
  - Can the k-th inference problem be cheaper than the 1st?
  - Computational significance of ILP formulations

Inference for BIG TEXT

- In NLP, we typically don't solve a single inference problem.
- We solve one or more per sentence.
- Beyond improving the inference algorithm, what can be done?

Inference for BIG TEXT

- In NLP, we typically don't solve a single inference problem.
- We solve one or more per sentence.
- Beyond improving the inference algorithm, what can be done?

S1	S2	POS
He	She	PRP
is	is	VBZ
reading	watching	VBG
a	a	DT
book	movie	NN

S1 & S2 look very different but their output structures are the same

The inference outcomes are the same

Inference for BIG TEXT

- In NLP, we typically don't solve a single inference problem.
- We solve one or more per sentence.
- Beyond improving the inference algorithm, what can be done?

S1	S2	POS
He	She	PRP
is	is	VBZ
reading	watching	VBG
a	a	DT
book	movie	NN

S1 & S2 look very different but their output structures are the same

The inference outcomes are the same

After inferring the POS structure for S1,
Can we speed up inference for S2 ?

Amortized ILP based Inference

- Imagine that you already solved many structured output inference problems
  - Co-reference resolution; Semantic Role Labeling; Parsing citations; Summarization; dependency parsing; image segmentation,…
  - Your solution method doesn't matter either

Amortized ILP based Inference

- Imagine that you already solved many structured output inference problems
  - Co-reference resolution; Semantic Role Labeling; Parsing citations; Summarization; dependency parsing; image segmentation,…
  - Your solution method doesn’t matter either
- **How can we exploit this fact to save inference cost**?

After solving **n** inference problems, can we make the **(n+1)**th one faster?

Amortized ILP based Inference

- Imagine that you already solved many structured output inference problems
  - Co-reference resolution; Semantic Role Labeling; Parsing citations; Summarization; dependency parsing; image segmentation,…
  - Your solution method doesn't matter either
- **How can we exploit this fact to save inference cost**?

> After solving **n** inference problems, can we make the **(n+1)**th one faster?

- We will show how to do it when your problem is formulated as a 0-1 LP, Max $\mathbf{cx}$

$$A\mathbf{x} \leq \mathbf{b}$$

Amortized ILP based Inference

- Imagine that you already solved many structured output inference problems
  - Co-reference resolution; Semantic Role Labeling; Parsing citations; Summarization; dependency parsing; image segmentation,…
  - Your solution method doesn't matter either
- **How can we exploit this fact to save inference cost**?

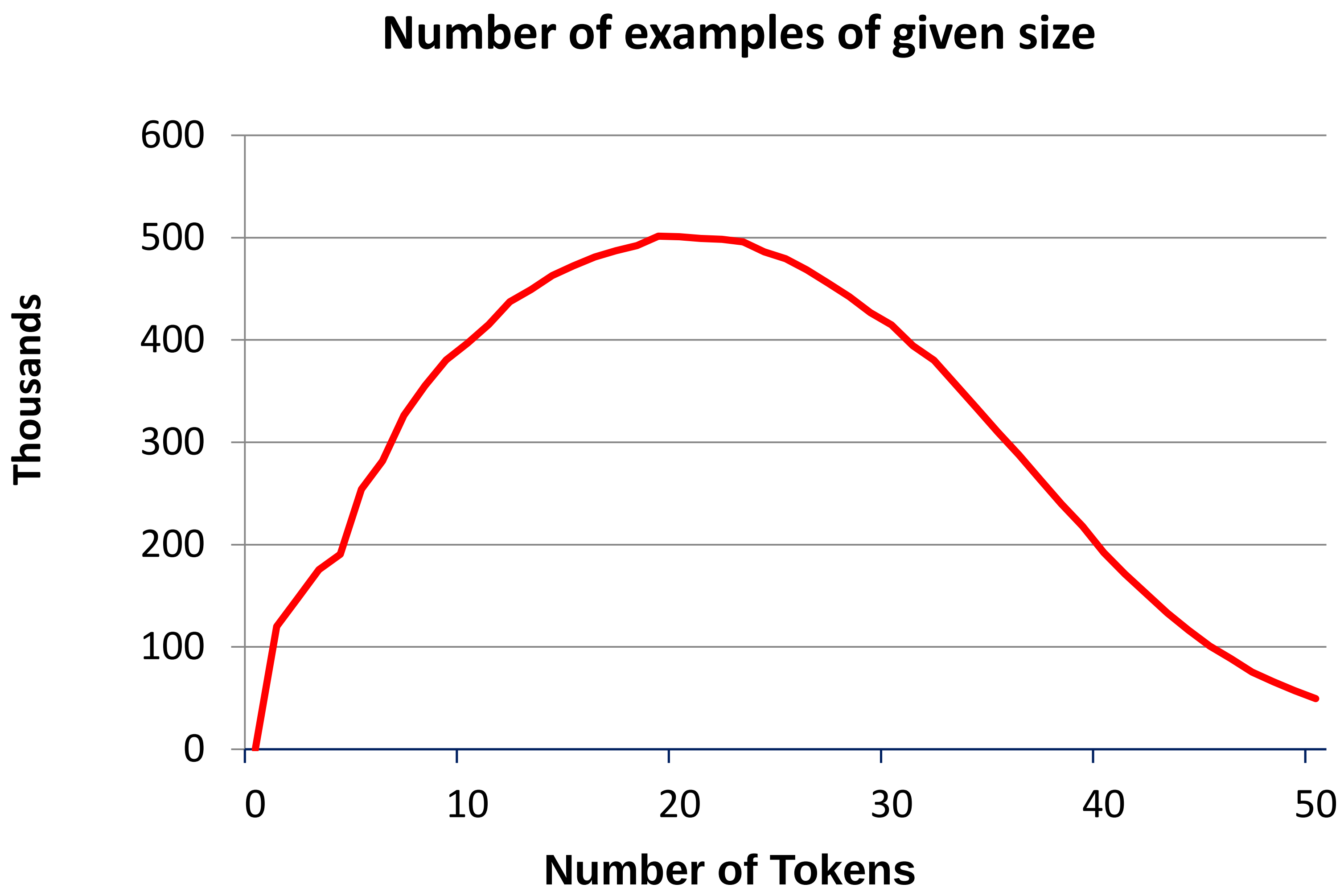
After solving **n** inference problems, can we make the **(n+1)**th one faster?

- We will show how to do it when your problem is formulated as a 0-1 LP, Max $\mathbf{cx}$

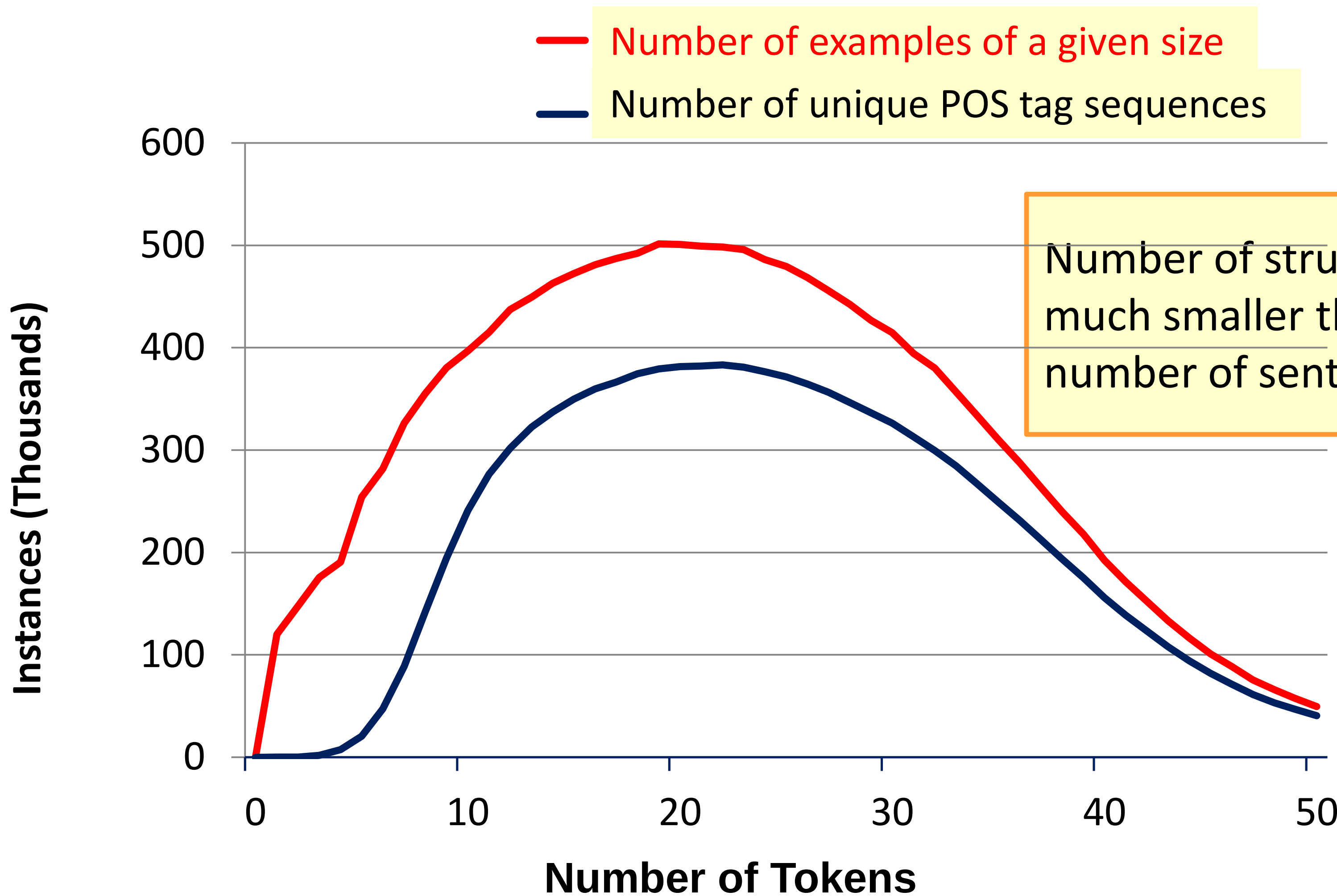
$$A\mathbf{x} \le \mathbf{b}$$

- Very general: All discrete MAP problems can be formulated as 0-1 LPs
- We only care about inference formulation, not algorithmic solution

The Hope: POS Tagging on Gigaword

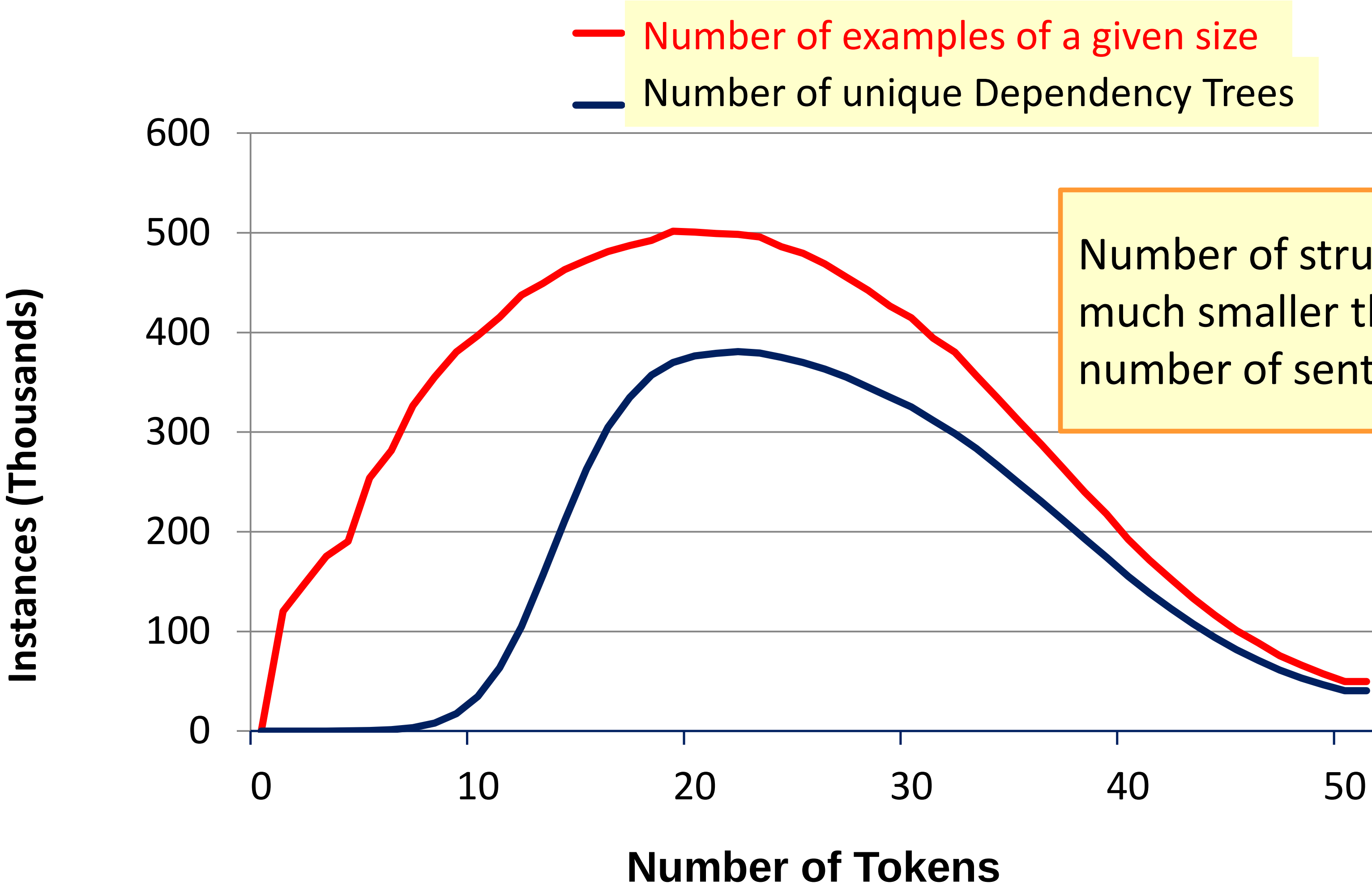


The Hope: POS Tagging on Gigaword



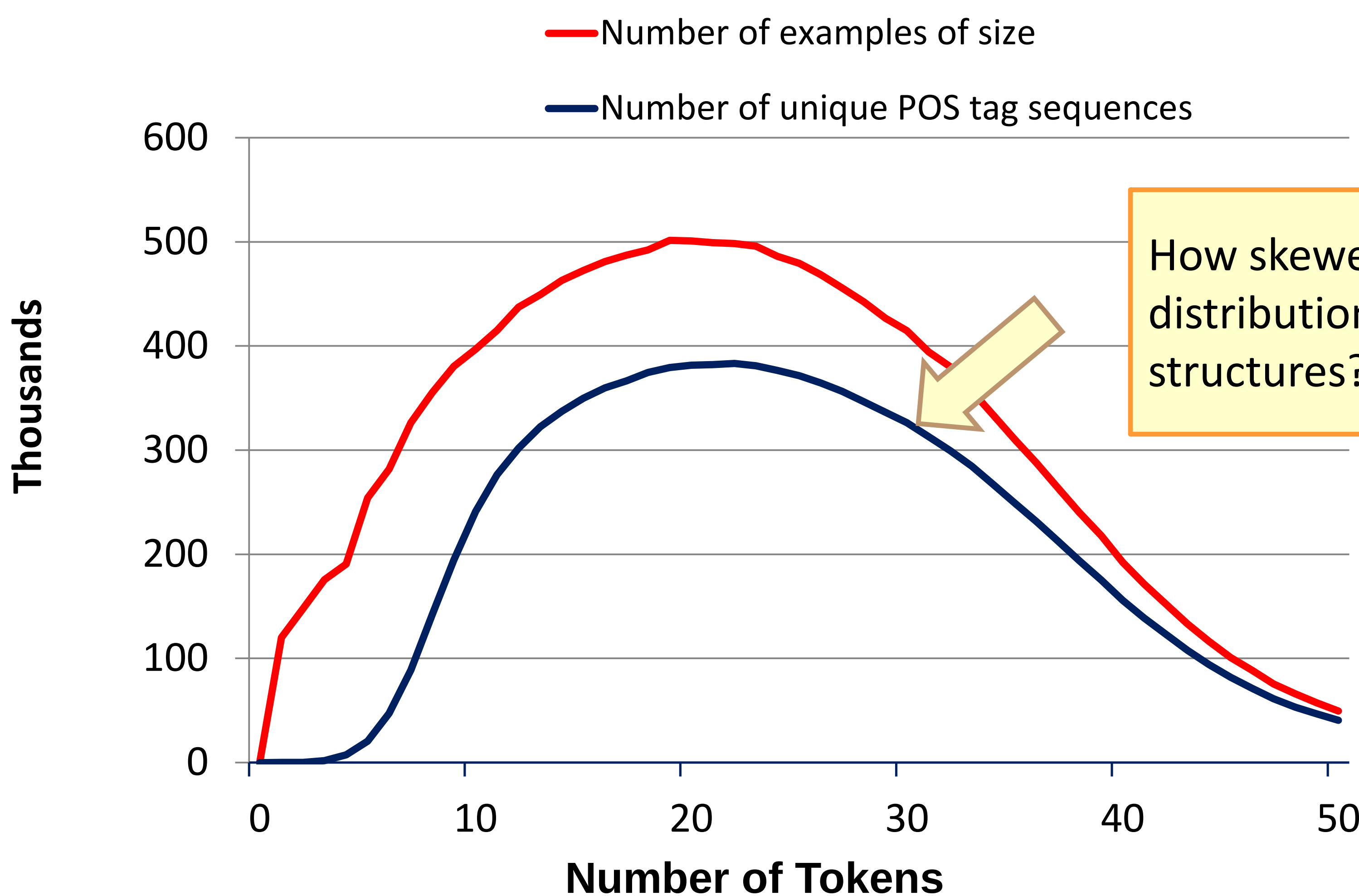
Number of structures is much smaller than the number of sentences

The Hope: Dependency Parsing on Gigaword



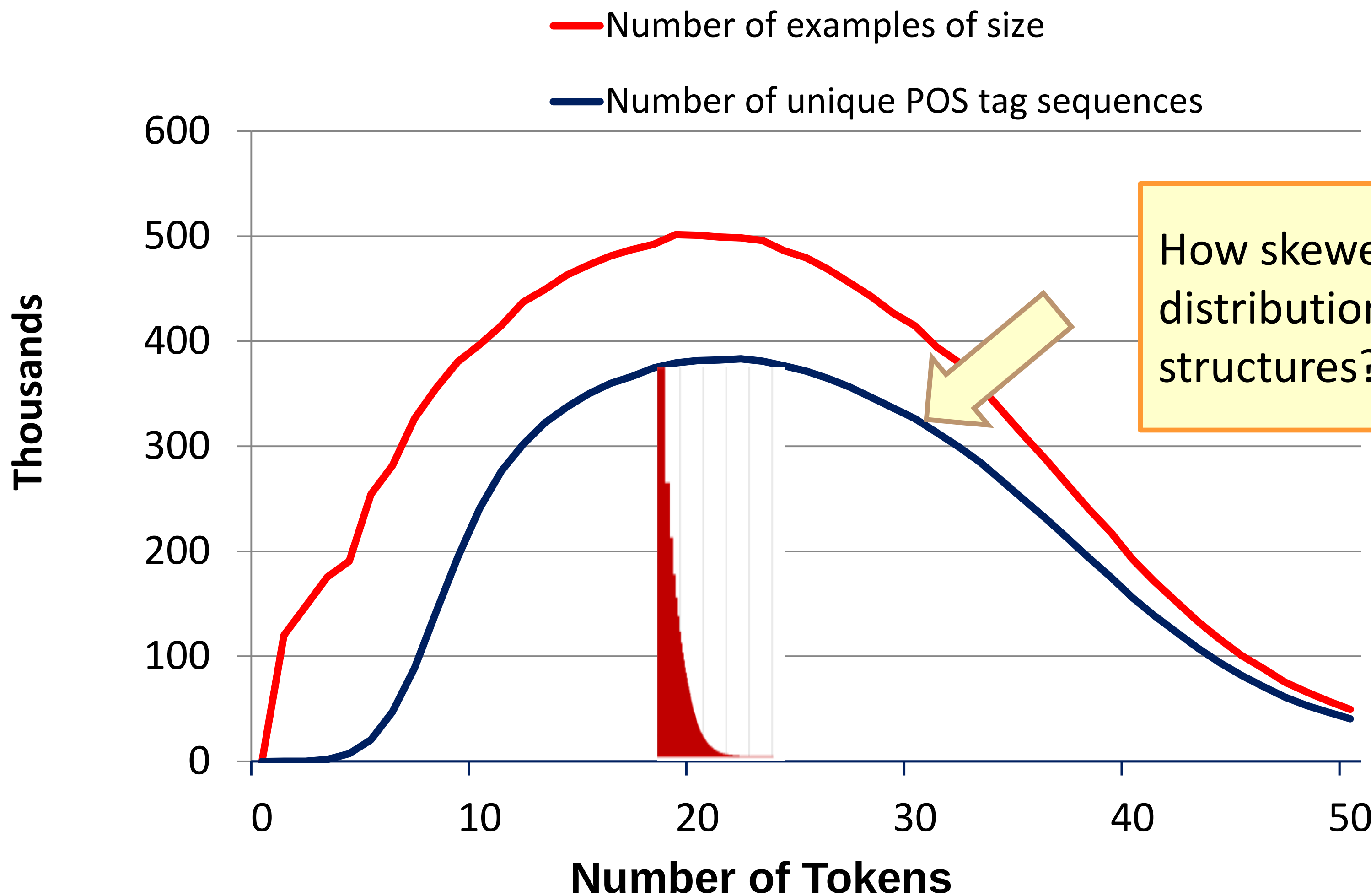
Number of structures is much smaller than the number of sentences

POS Tagging on Gigaword



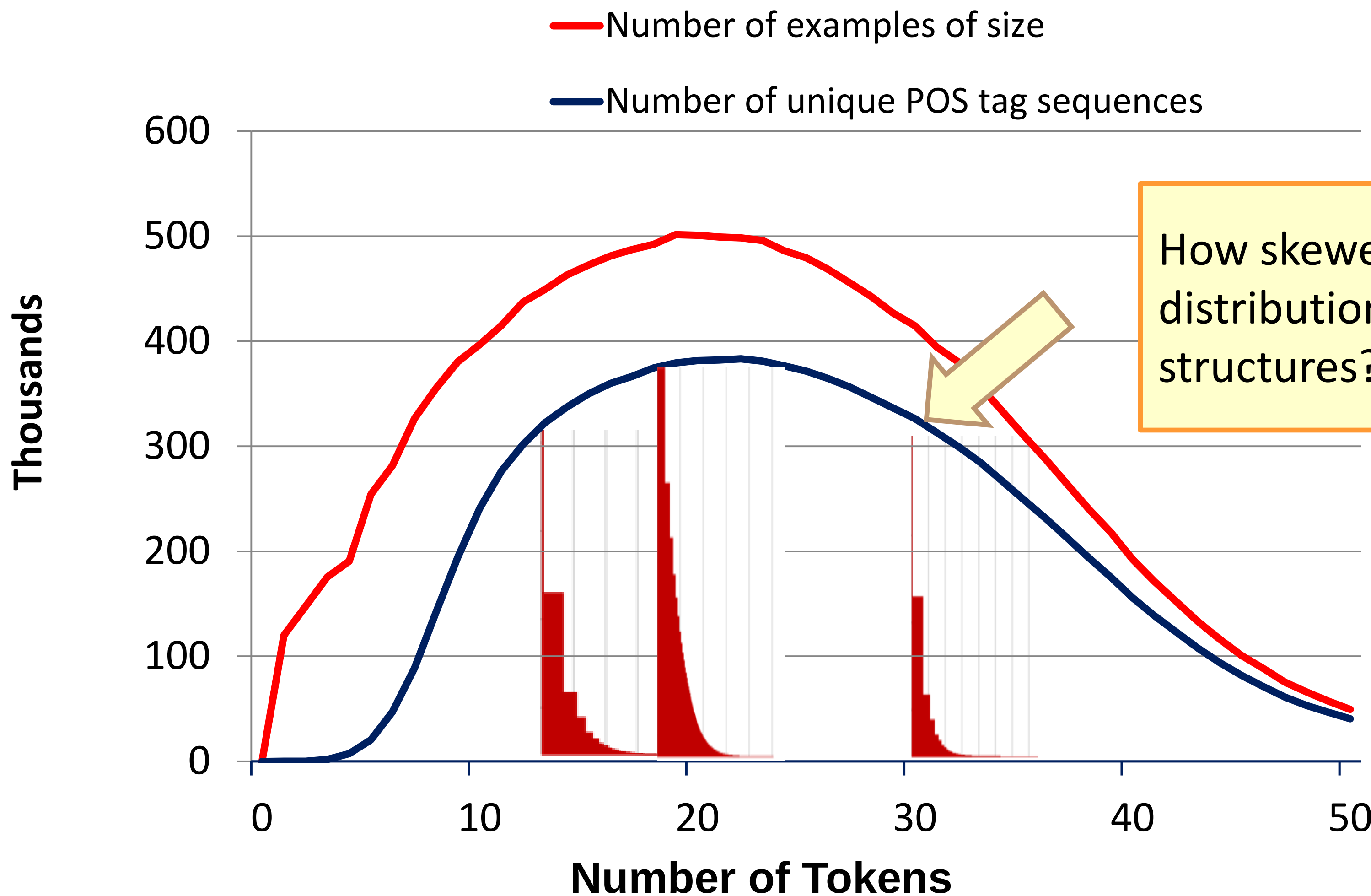
How skewed is the distribution of the structures?

POS Tagging on Gigaword



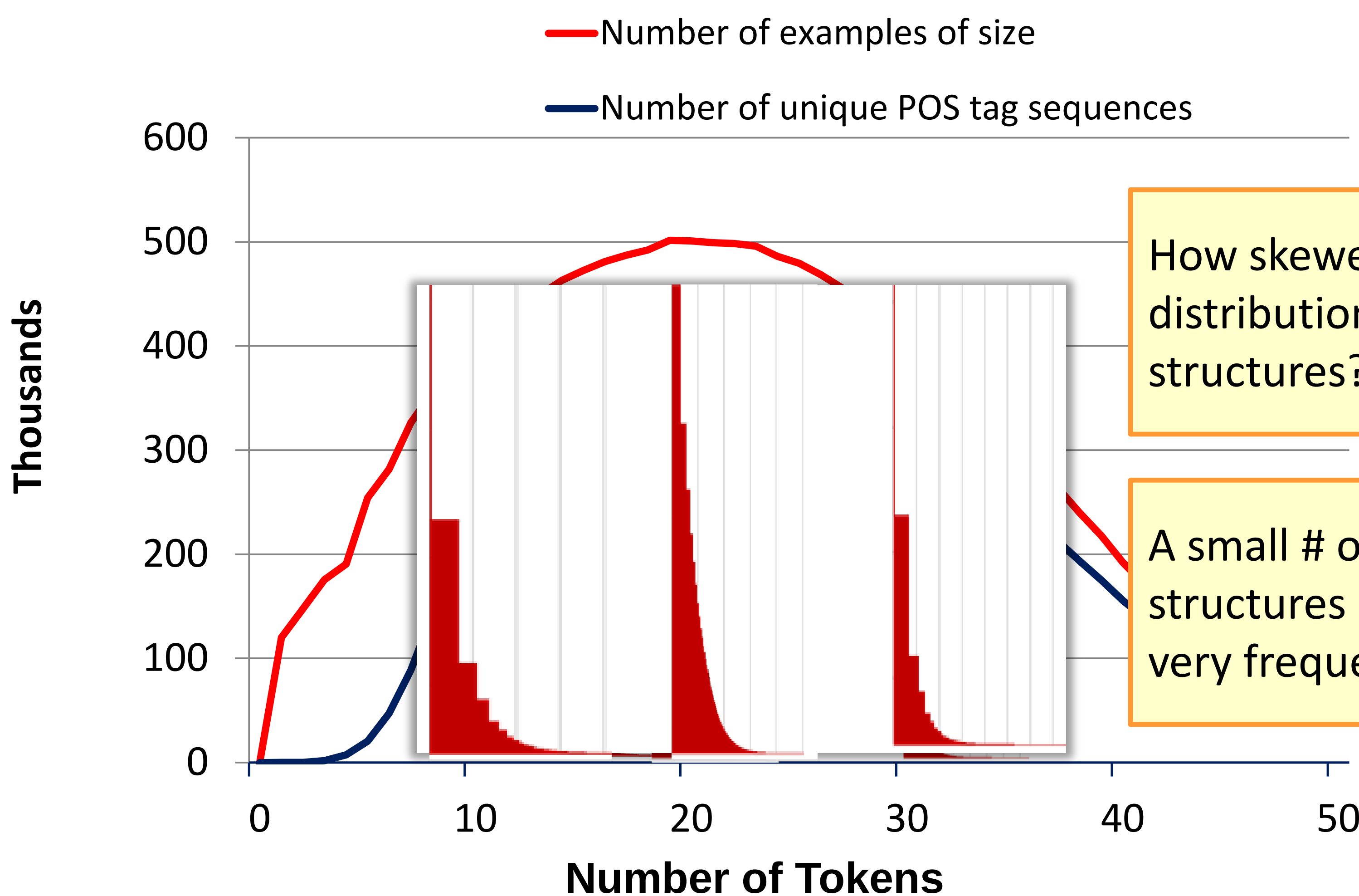
How skewed is the distribution of the structures?

POS Tagging on Gigaword



How skewed is the distribution of the structures?

POS Tagging on Gigaword



How skewed is the distribution of the structures?

A small # of structures occur very frequently

Amortized ILP Inference

- These statistics show that many different instances are mapped into identical inference outcomes.
  - Pigeon Hole Principle
- **How can we exploit this fact to save inference cost over the life time of the agent?** ?

Amortized ILP Inference

- These statistics show that many different instances are mapped into identical inference outcomes.
  - Pigeon Hole Principle
- **How can we exploit this fact to save inference cost over the life time of the agent?** ?

We give conditions on the objective functions
(for all objectives with the same # or variables and same feasible set),
under which the solution of a new problem Q is the same as the one of P (which we already cached)

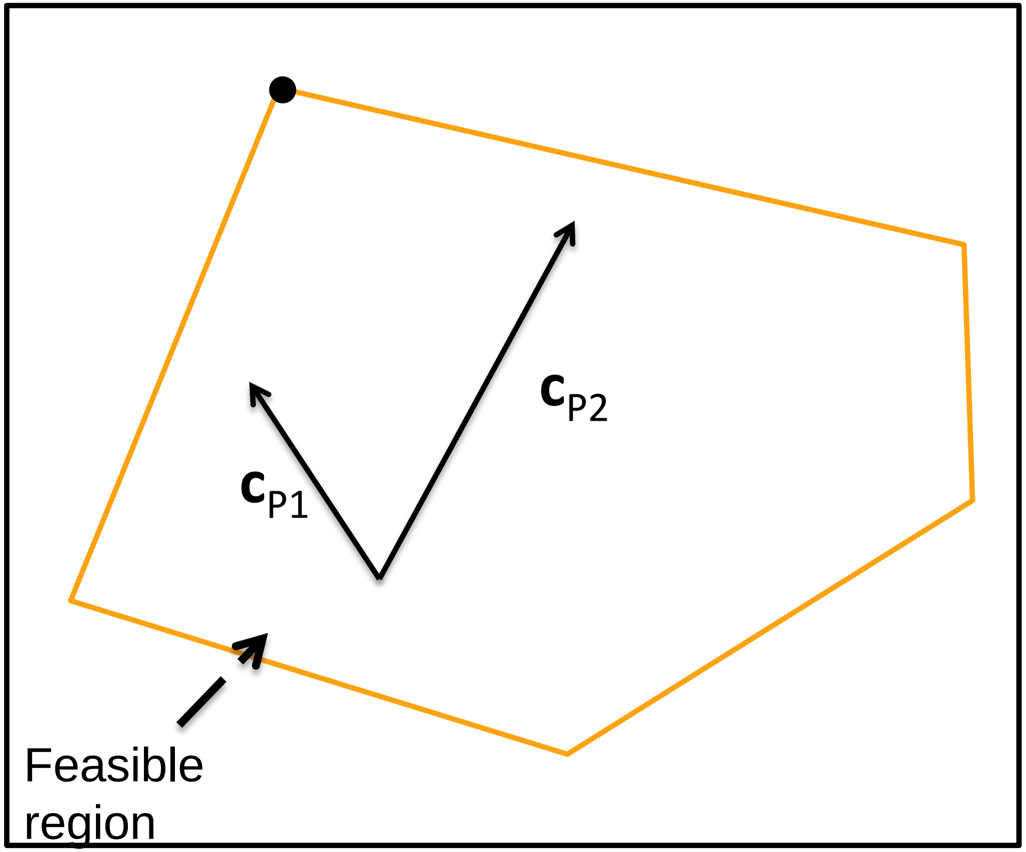
Amortized ILP Inference

We argue here that the inference formulation provides a new level of abstraction.

- These statistics show that many different instances are mapped into identical inference outcomes.
  - Pigeon Hole Principle
- **How can we exploit this fact to save inference cost over the life time of the agent?** ?

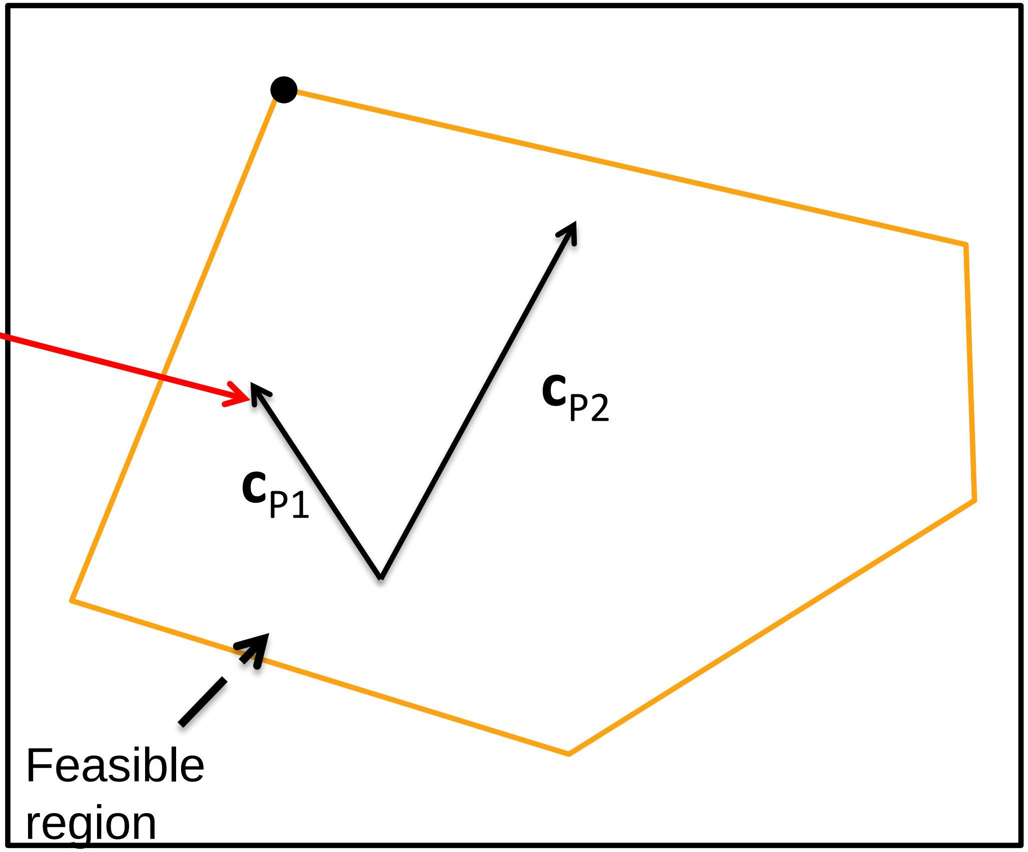
We give conditions on the objective functions
(for all objectives with the same # or variables and same feasible set),
under which the solution of a new problem Q is the same as the one of P (which we already cached)

Theorem II (Geometric Interpretation)



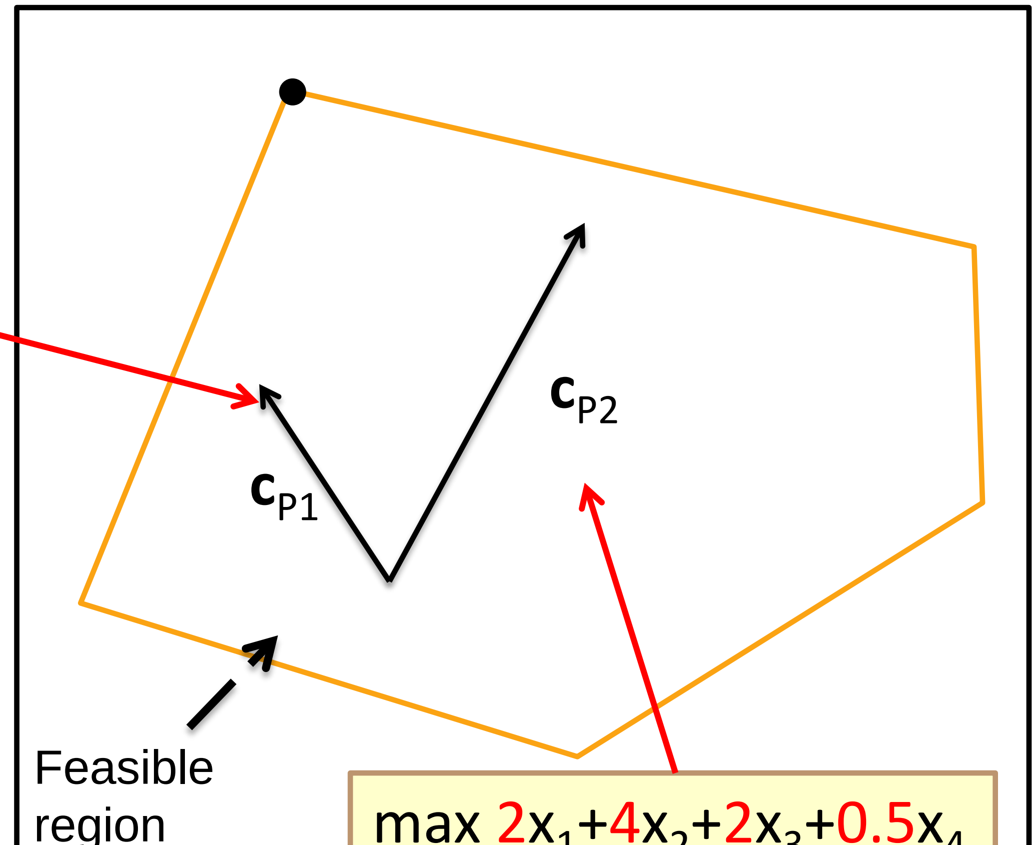
Theorem II (Geometric Interpretation)

$$\max\ 2x_1+3x_2+2x_3+1x_4$$
$$x_1 + x_2 \le 1$$
$$x_3 + x_4 \le 1$$



Theorem II (Geometric Interpretation)

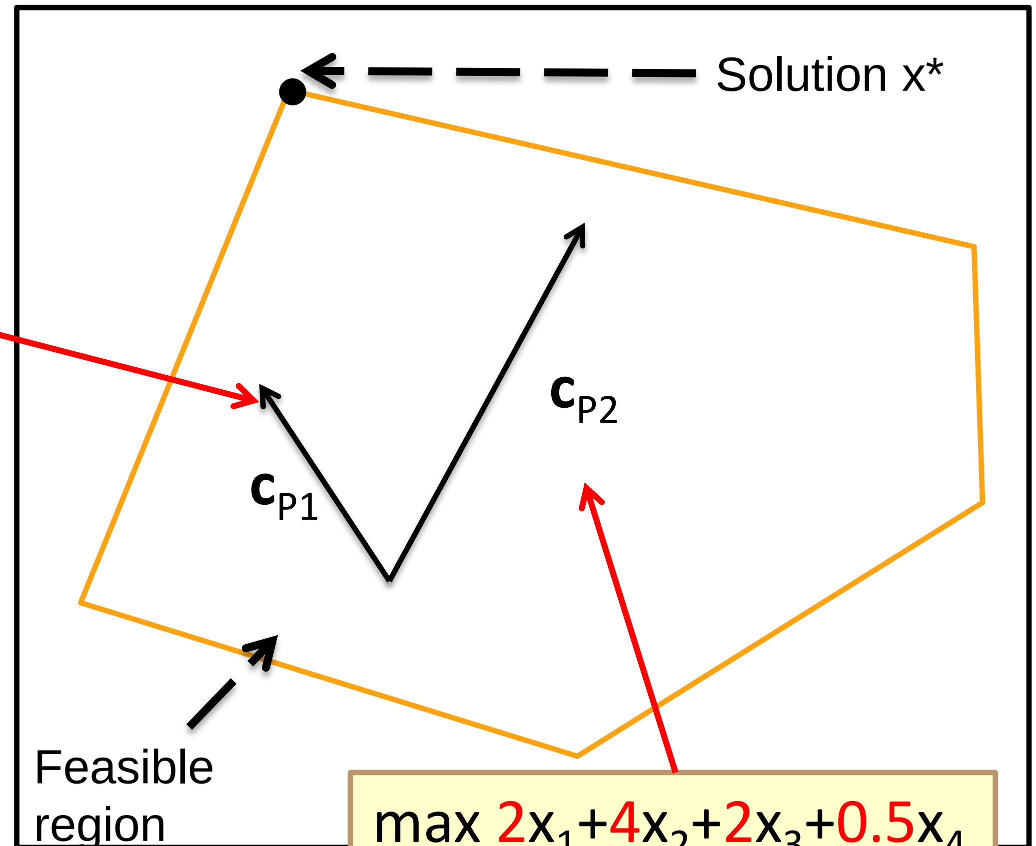
$$\max\ 2x_1+3x_2+2x_3+1x_4$$
$$x_1 + x_2 \le 1$$
$$x_3 + x_4 \le 1$$



$$\max\ 2x_1+4x_2+2x_3+0.5x_4$$
$$x_1 + x_2 \le 1$$
$$x_3 + x_4 \le 1$$

Theorem II (Geometric Interpretation)

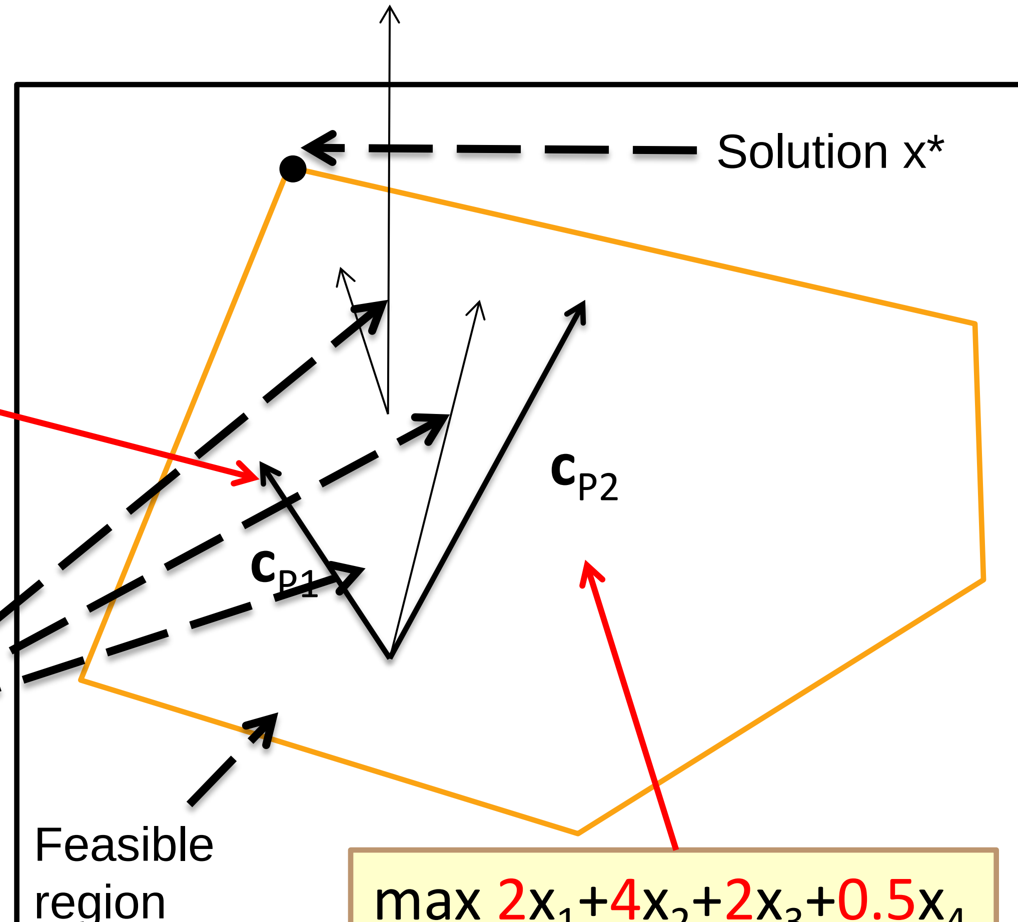
$$\max\ 2x_1+3x_2+2x_3+1x_4$$
$$x_1 + x_2 \le 1$$
$$x_3 + x_4 \le 1$$



$$\max\ 2x_1+4x_2+2x_3+0.5x_4$$
$$x_1 + x_2 \le 1$$
$$x_3 + x_4 \le 1$$

Theorem II (Geometric Interpretation)

$$\max \ 2x_1+3x_2+2x_3+1x_4$$
$$x_1 + x_2 \leq 1$$
$$x_3 + x_4 \leq 1$$



ILPs corresponding to all these objective vectors will share the same maximizer *for this feasible region*

$$\max \ 2x_1+4x_2+2x_3+0.5x_4$$
$$x_1 + x_2 \leq 1$$
$$x_3 + x_4 \leq 1$$

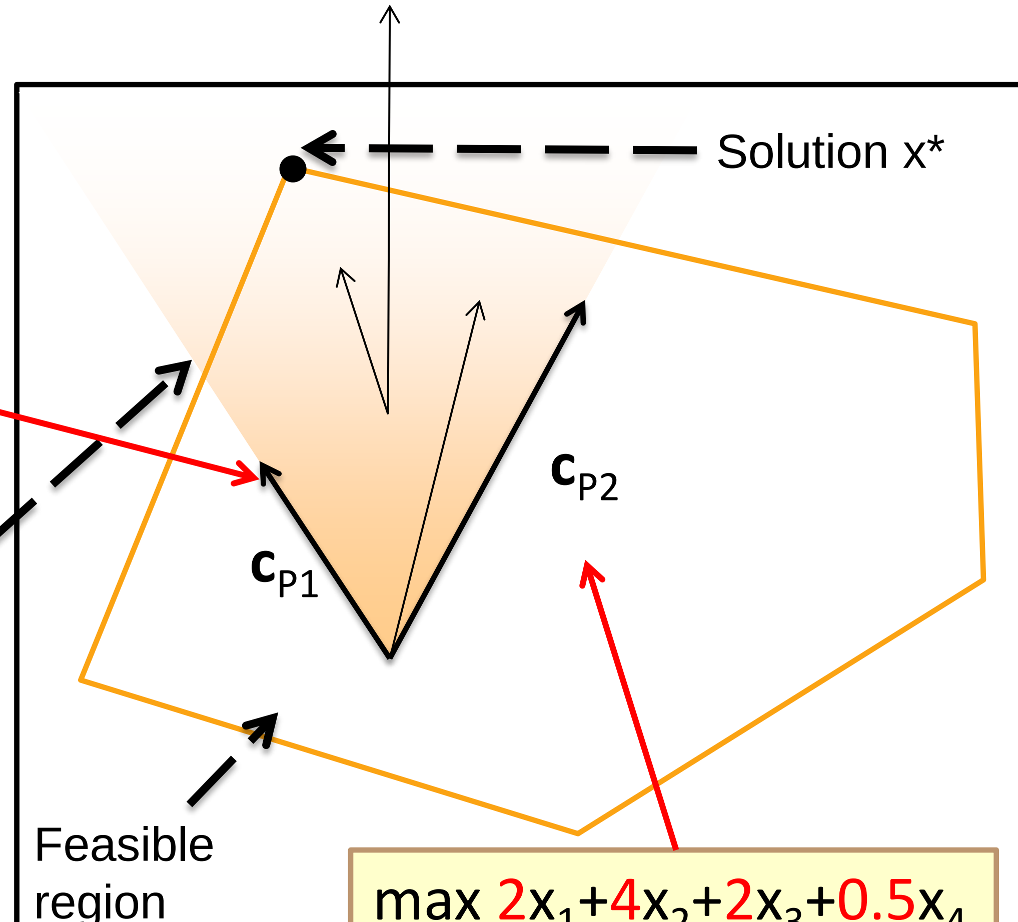
Theorem II (Geometric Interpretation)

$\max\ 2x_1 + 3x_2 + 2x_3 + 1x_4$

$x_1 + x_2 \leq 1$

$x_3 + x_4 \leq 1$

All ILPs in the ***cone*** will share the maximizer



$\max\ 2x_1 + 4x_2 + 2x_3 + 0.5x_4$

$x_1 + x_2 \leq 1$

$x_3 + x_4 \leq 1$

Amortized Inference Experiments

- Setup
  - Verb semantic role labeling; Entity and Relations
  - Speedup & Accuracy are measured over WSJ test set (Section 23) and Test of E & R
  - Baseline: solving ILPs using the Gurobi solver.

Amortized Inference Experiments

- Setup
  - Verb semantic role labeling; Entity and Relations
  - Speedup & Accuracy are measured over WSJ test set (Section 23) and Test of E & R
  - Baseline: solving ILPs using the Gurobi solver.

- For amortization
  - Cache 250,000 inference problems (objective, solution) from Gigaword
  - For each problem in test set either call the inference engine or re-use a solution from the cache, if our theorems hold.

Amortized Inference Experiments

- Setup
  - Verb semantic role labeling; Entity and Relations
  - Speedup & Accuracy are measured over WSJ test set (Section 23) and Test of E & R
  - Baseline: solving ILPs using the Gurobi solver.

- For amortization
  - Cache 250,000 inference problems (objective, solution) from Gigaword
  - For each problem in test set either call the inference engine or re-use a solution from the cache, if our theorems hold.

No training data is needed for this method.
Once you have a model, you can generate a large cache that will be then used to save you time at evaluation time.

Speedup & Accuracy

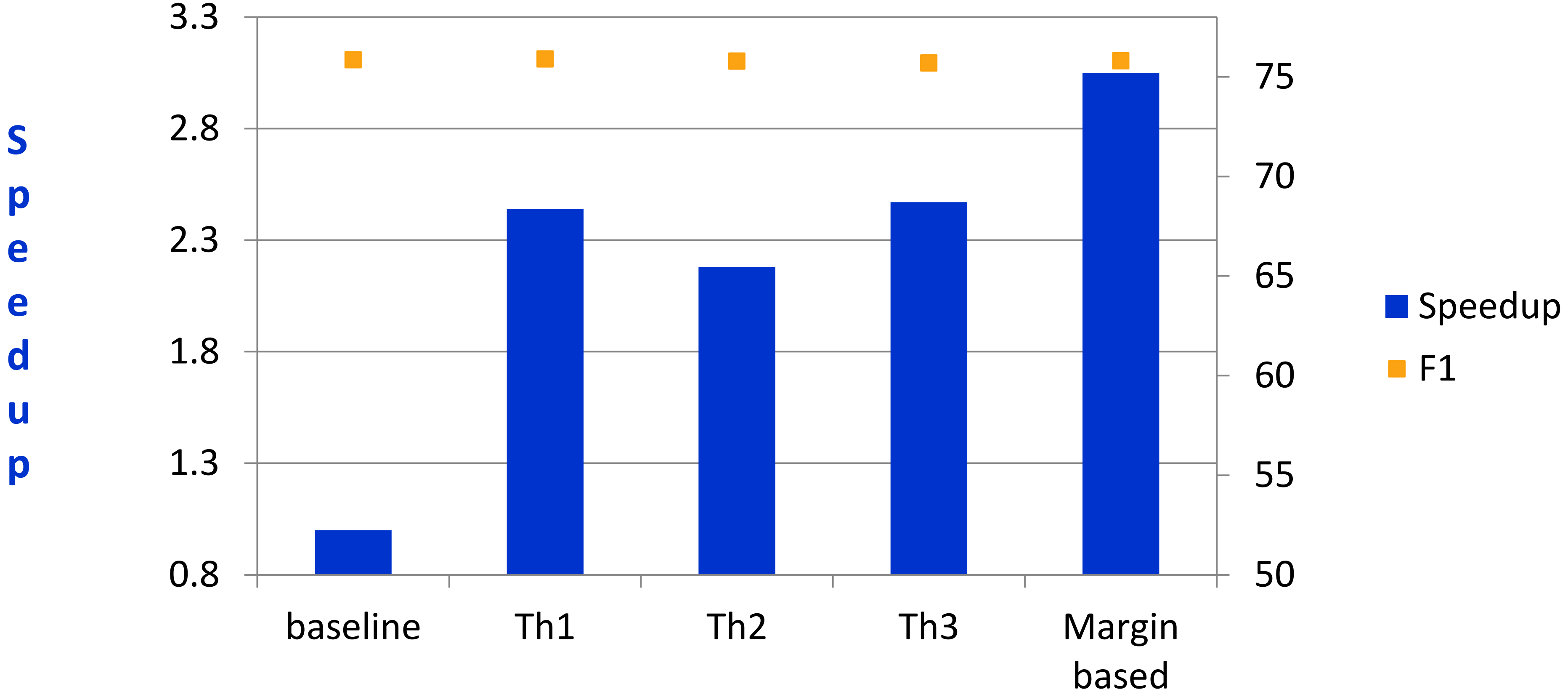
$$\text{Speedup} = \frac{\text{number of inference calls without amortization}}{\text{number of inference calls with amortization}}$$

**S
p
e
e
d
u
p**

Amortization schemes [EMNLP'12, ACL'13]

Speedup & Accuracy

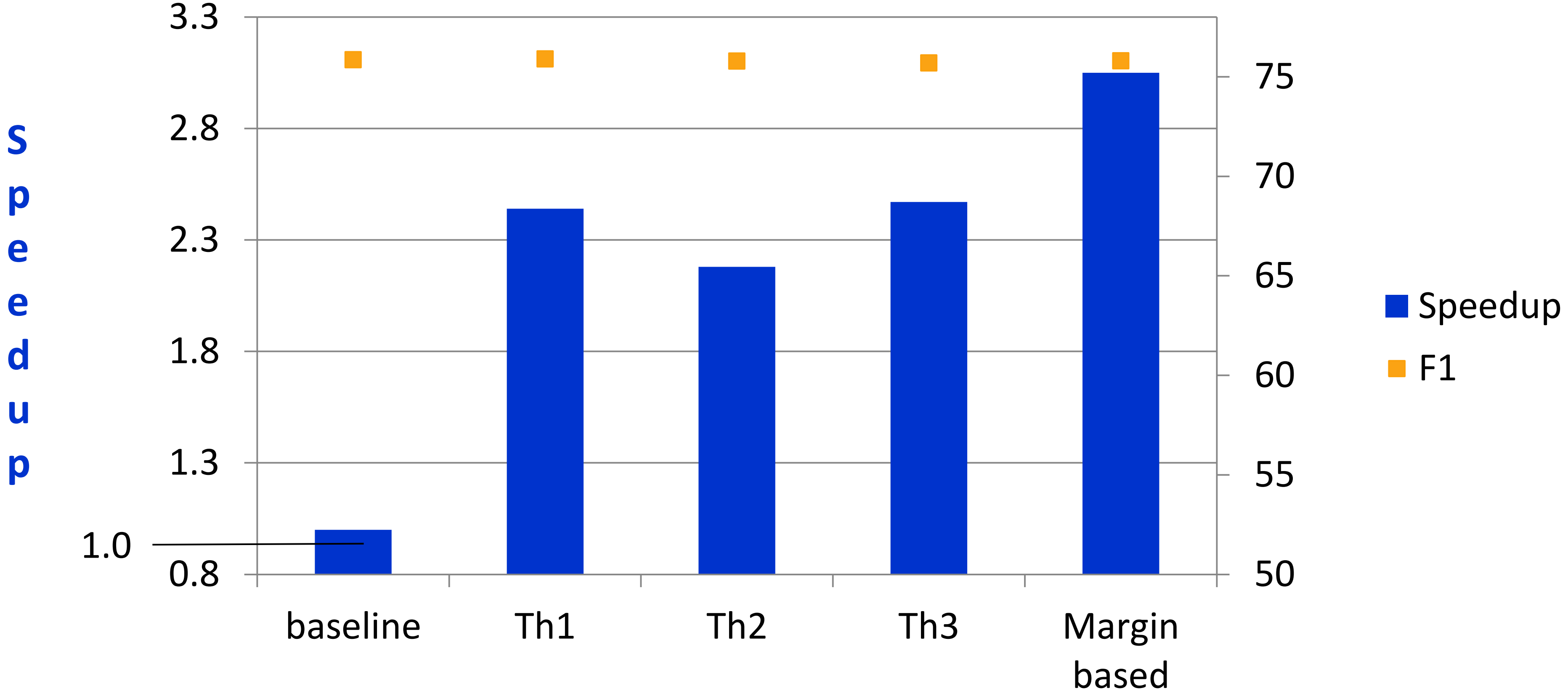
$$\text{Speedup} = \frac{\text{number of inference calls without amortization}}{\text{number of inference calls with amortization}}$$



Amortization schemes [EMNLP'12, ACL'13]

Speedup & Accuracy

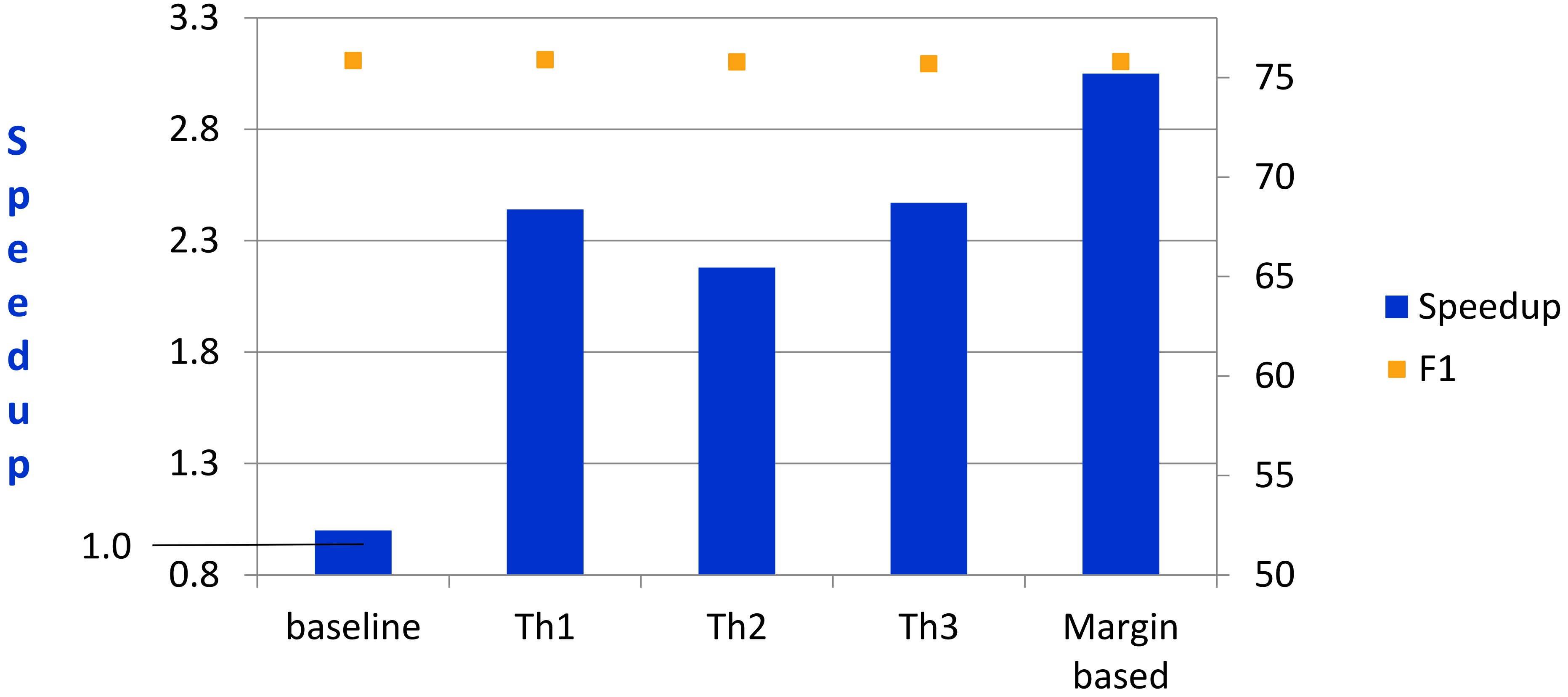
$$\text{Speedup} = \frac{\text{number of inference calls without amortization}}{\text{number of inference calls with amortization}}$$



Amortization schemes [EMNLP'12, ACL'13]

Speedup & Accuracy

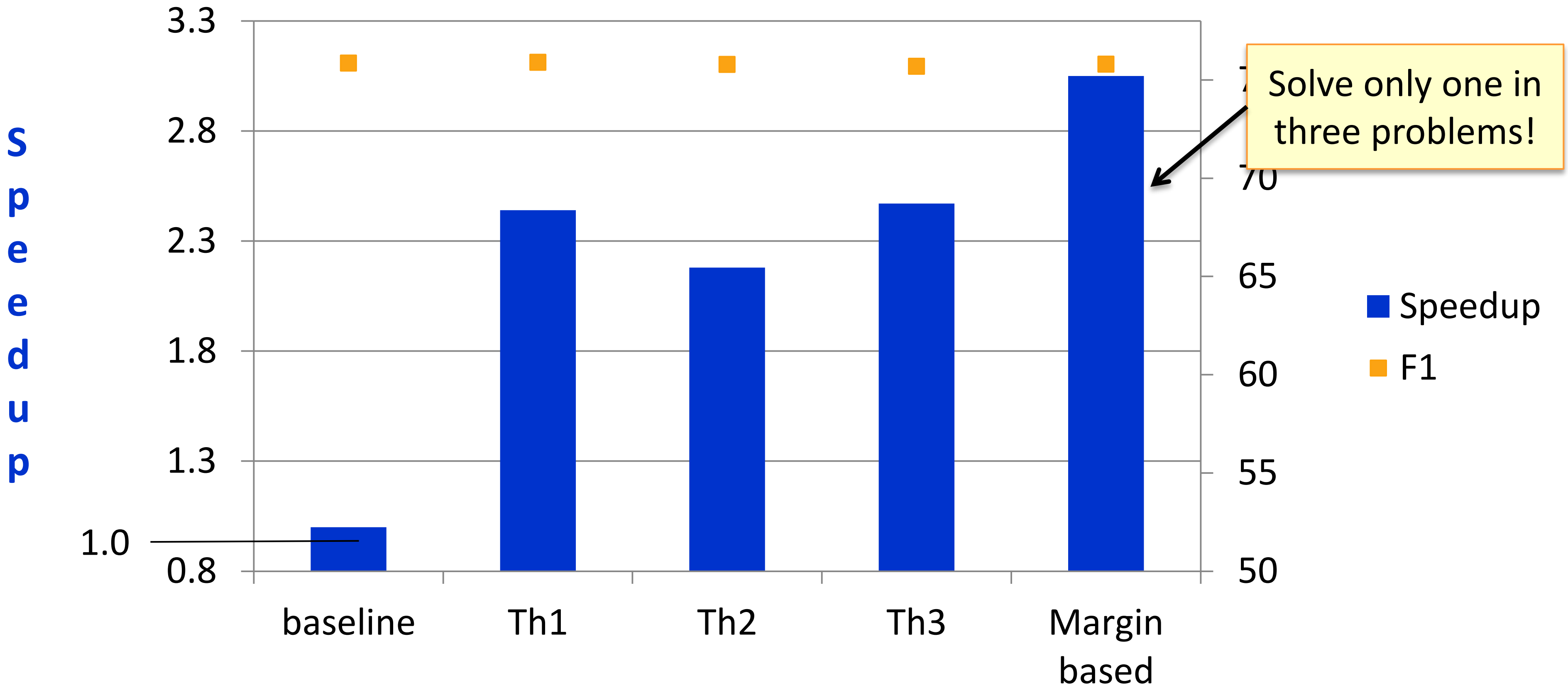
$$\text{Speedup} = \frac{\text{number of inference calls without amortization}}{\text{number of inference calls with amortization}}$$



Amortization schemes [EMNLP'12, ACL'13]

Speedup & Accuracy

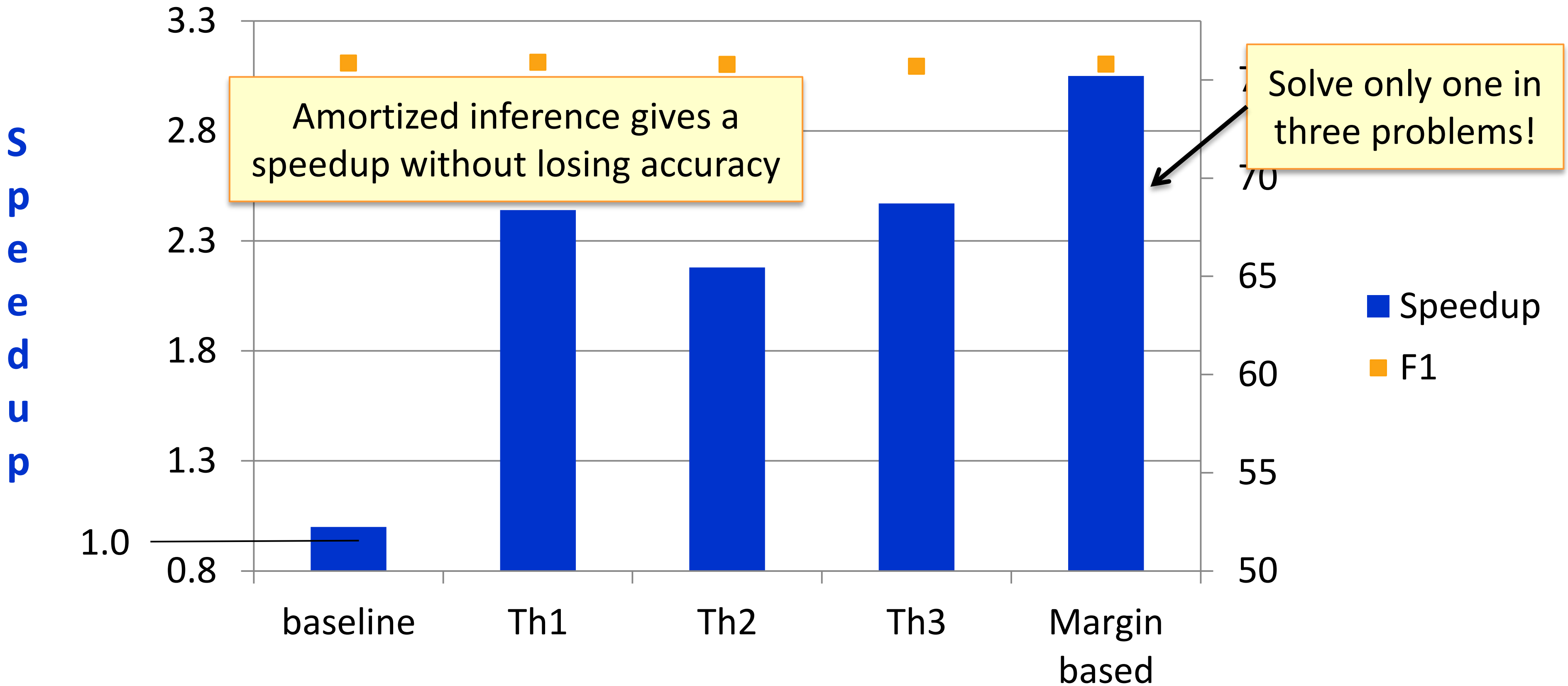
$$\text{Speedup} = \frac{\text{number of inference calls without amortization}}{\text{number of inference calls with amortization}}$$



Amortization schemes [EMNLP'12, ACL'13]

Speedup & Accuracy

$$\text{Speedup} = \frac{\text{number of inference calls without amortization}}{\text{number of inference calls with amortization}}$$



Amortization schemes [EMNLP'12, ACL'13]

Speedup & Accuracy

$$\text{Speedup} = \frac{\text{number of inference calls without amortization}}{\text{number of inference calls with amortization}}$$

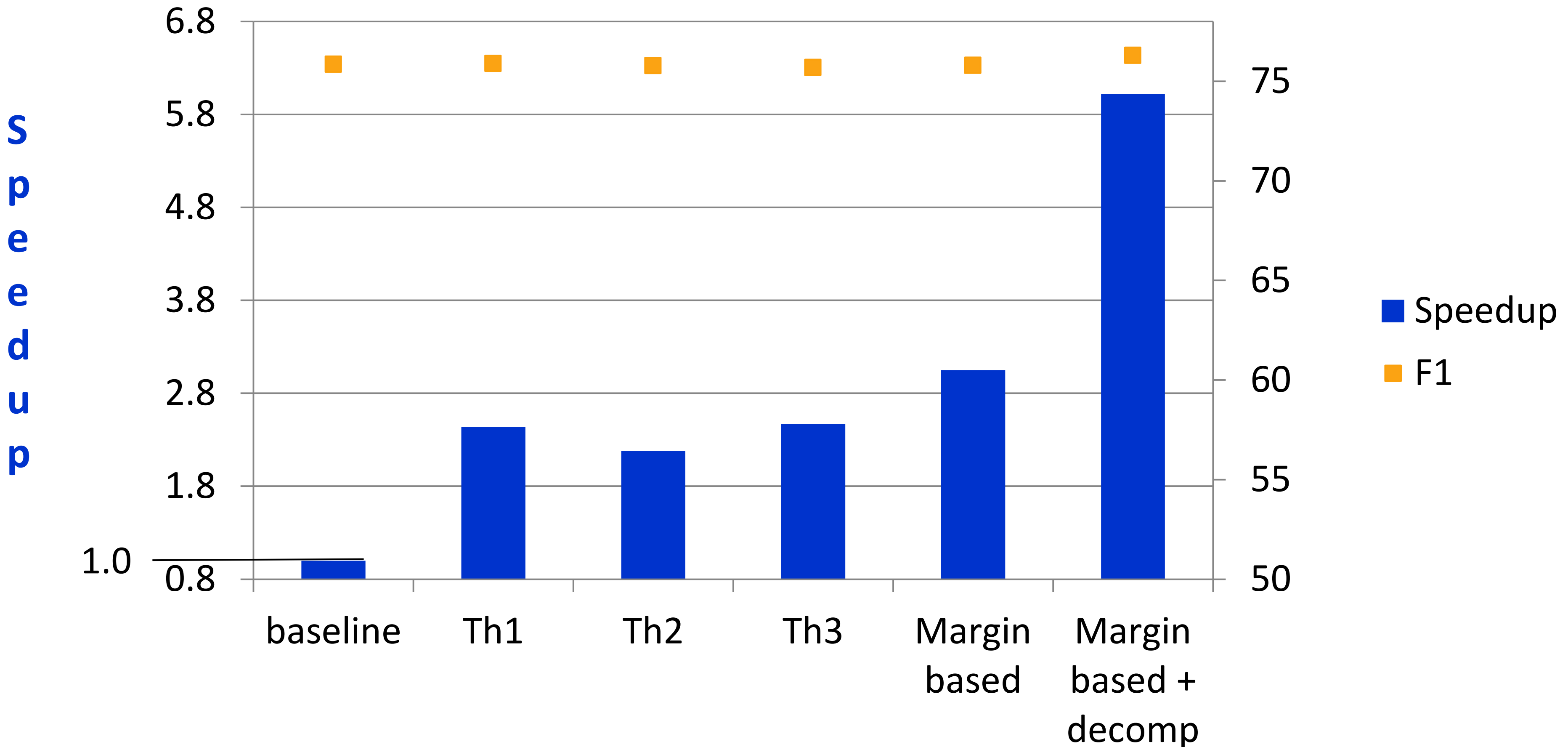
S
p
e
e
d
u
p

Amortization schemes [EMNLP'12, ACL'13]

Speedup & Accuracy

By decomposing the objective function, building on the fact that "smaller structures" are more redundant, it is possible to get even better results.

$$\text{Speedup} = \frac{\text{number of inference calls without amortization}}{\text{number of inference calls with amortization}}$$

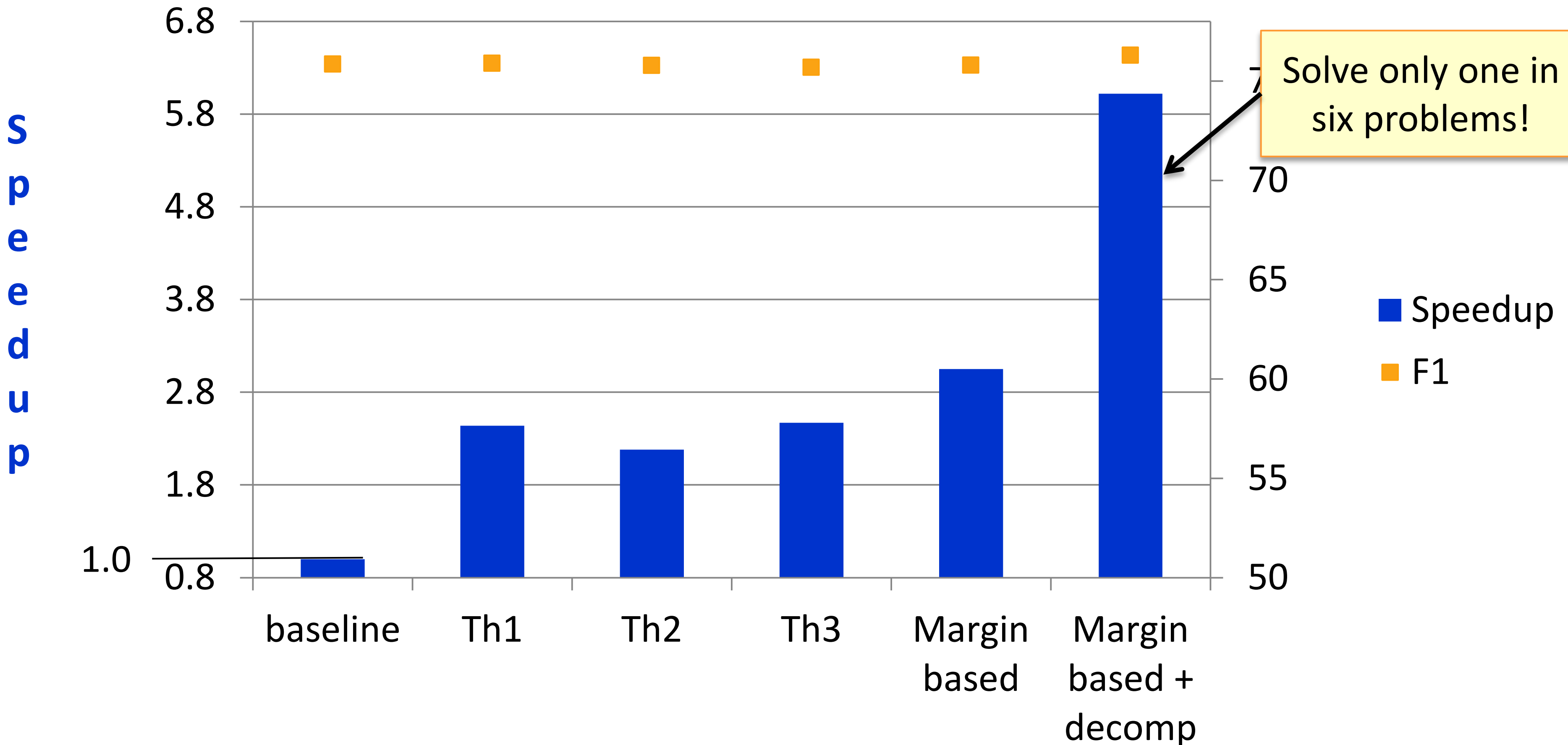


Amortization schemes [EMNLP'12, ACL'13]

Speedup & Accuracy

By decomposing the objective function, building on the fact that "smaller structures" are more redundant, it is possible to get even better results.

$$\text{Speedup} = \frac{\text{number of inference calls without amortization}}{\text{number of inference calls with amortization}}$$

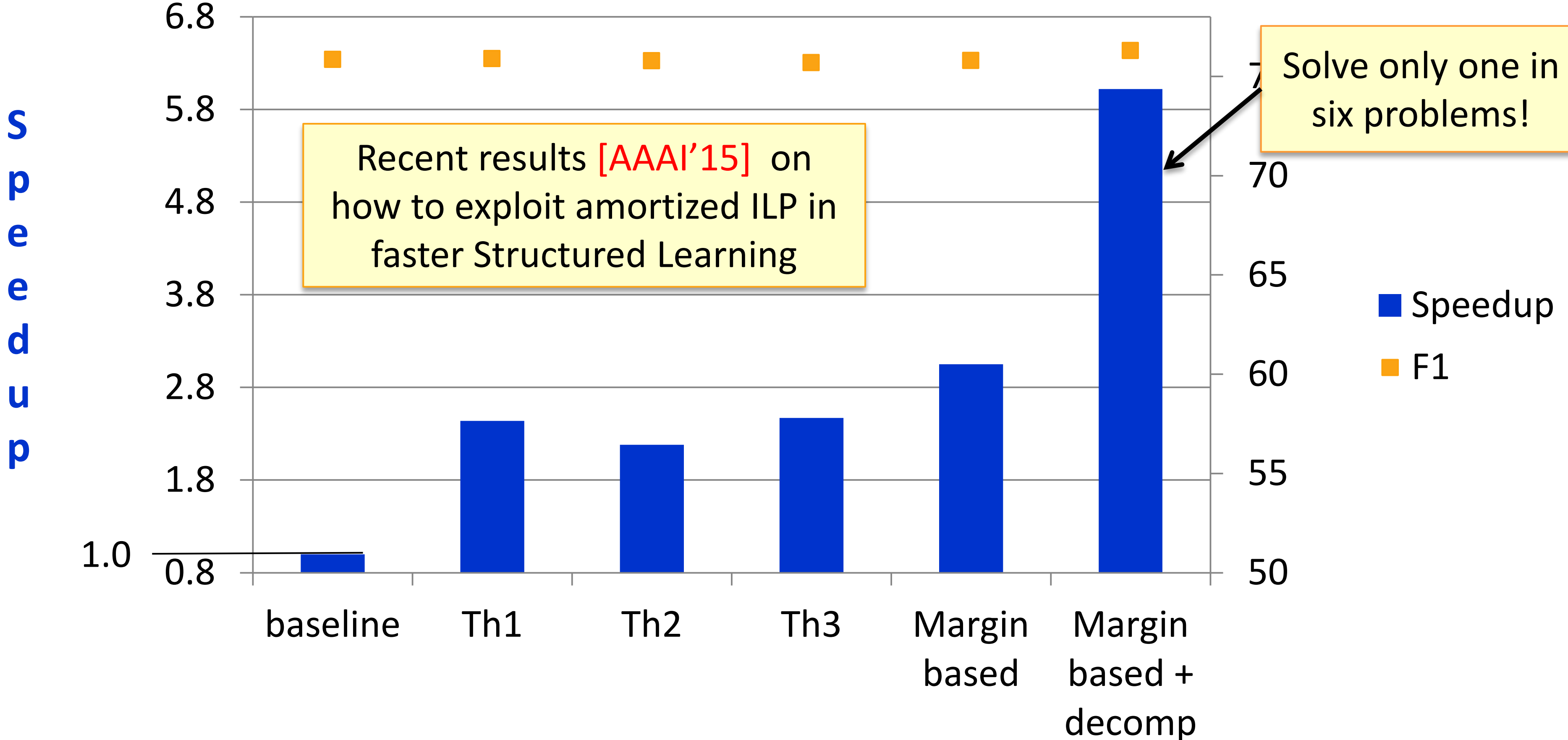


Amortization schemes [EMNLP'12, ACL'13]

Speedup & Accuracy

The results show that, indeed, the inference formulation provides a new level of abstraction that can be exploited to re-use solutions

$$\text{Speedup} = \frac{\text{number of inference calls without amortization}}{\text{number of inference calls with amortization}}$$

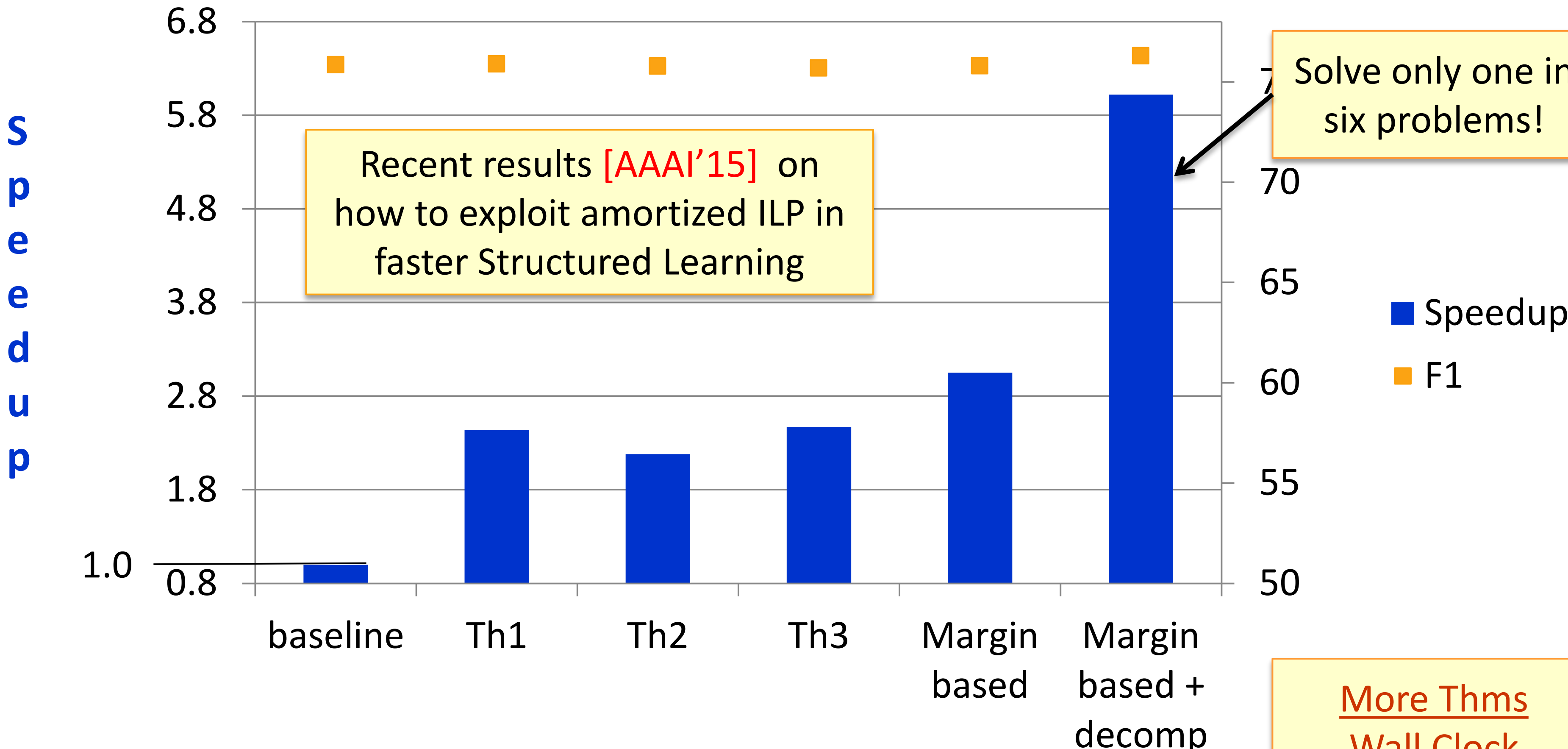


Amortization schemes [EMNLP'12, ACL'13]

Speedup & Accuracy

The results show that, indeed, the inference formulation provides a new level of abstraction that can be exploited to re-use solutions

$$\text{Speedup} = \frac{\text{number of inference calls without amortization}}{\text{number of inference calls with amortization}}$$



Amortization schemes [EMNLP'12, ACL'13]

More Thms
Wall Clock

Before Concluding

- Why is automating natural language understanding difficult?

Before Concluding

- Why is automating natural language understanding difficult?

  - When Tina pressed Joan to the floor she was hurt.
  - When Tina pressed Joan to the floor she was punished.
  - When Tina pressed charges against Joan she was jailed.

Before Concluding

- Why is automating natural language understanding difficult?
  - When Tina pressed Joan to the floor she was hurt.
  - When Tina pressed Joan to the floor she was punished.
  - When Tina pressed charges against Joan she was jailed.

Before Concluding

- Why is automating natural language understanding difficult?

  - When Tina pressed Joan to the floor she was hurt.
  - When Tina pressed Joan to the floor she was punished.
  - When Tina pressed charges against Joan she was jailed.

- Requires, among other things, thinking about the structure of the sentence – who does what to whom

Before Concluding

- Why is automating natural language understanding difficult?

  - When Tina pressed Joan to the floor she was hurt.
  - When Tina pressed Joan to the floor she was punished.
  - When Tina pressed charges against Joan she was jailed.

- Requires, among other things, thinking about the structure of the sentence – who does what to whom
  - (See our work in CoNLL'15 and NAACL'15 for interesting progress on this)

Before Concluding

- Why is automating natural language understanding difficult?

  - When Tina pressed Joan to the floor she was hurt.
  - When Tina pressed Joan to the floor she was punished.
  - When Tina pressed charges against Joan she was jailed.

- Requires, among other things, thinking about the structure of the sentence – who does what to whom
  - (See our work in CoNLL'15 and NAACL'15 for interesting progress on this)

- John had 6 books; he wanted to give it to two of his friends. How many will each one get?

Before Concluding

- Why is automating natural language understanding difficult?
  - When Tina pressed Joan to the floor she was hurt.
  - When Tina pressed Joan to the floor she was punished.
  - When Tina pressed charges against Joan she was jailed.
- Requires, among other things, thinking about the structure of the sentence – who does what to whom
  - (See our work in CoNLL'15 and NAACL'15 for interesting progress on this)
- John had 6 books; he wanted to give it to two of his friends. How many will each one get?
  - (See out EMNLP'15 work for interesting progress on Math word problems)

Before Concluding

- Why is automating natural language understanding difficult?
  - When Tina pressed Joan to the floor she was hurt.
  - When Tina pressed Joan to the floor she was punished.
  - When Tina pressed charges against Joan she was jailed.
- Requires, among other things, thinking about the structure of the sentence – who does what to whom
  - (See our work in CoNLL'15 and NAACL'15 for interesting progress on this)

share it with

- John had 6 books; he wanted to give it to two of his friends. How many will each one get?
  - (See out EMNLP'15 work for interesting progress on Math word problems)

Before Concluding

- Why is automating natural language understanding difficult?

  - When Tina pressed Joan to the floor she was hurt.
  - When Tina pressed Joan to the floor she was punished.
  - When Tina pressed charges against Joan she was jailed.

- Requires, among other things, thinking about the structure of the sentence – who does what to whom
  - (See our work in CoNLL'15 and NAACL'15 for interesting progress on this)

share it with

- John had 6 books; he wanted to give it to two of his friends. How many will each one get?
  - (See out EMNLP'15 work for interesting progress on Math word problems)

Conclusion

- NL interpretation is an inference problem best modelled as a knowledge constrained optimization problem over multiple statistically learned models.
- A lot more work is needed; there is a lot more that we try to do…
- Some recent samples from a research program that attempts to address some of the key scientific and engineering challenges between us and understanding natural language

Conclusion

- NL interpretation is an inference problem best modelled as a knowledge constrained optimization problem over multiple statistically learned models.
- A lot more work is needed; there is a lot more that we try to do…
- Some recent samples from a research program that attempts to address some of the key scientific and engineering challenges between us and understanding natural language
- **Knowledge – Reasoning – Learning Paradigms – Scaling up**
  - Thinking about Learning, Inference and knowledge via a Constrained Optimization framework that supports “best assignment” Inference.

Conclusion

- NL interpretation is an inference problem best modelled as a knowledge constrained optimization problem over multiple statistically learned models.
- A lot more work is needed; there is a lot more that we try to do…
- Some recent samples from a research program that attempts to address some of the key scientific and engineering challenges between us and understanding natural language
- **Knowledge – Reasoning – Learning Paradigms – Scaling up**
  - Thinking about Learning, Inference and knowledge via a Constrained Optimization framework that supports "best assignment" Inference.

Check out our tools, demos, LBJava, CCM tutorial,…

Conclusion

Thank You!

- NL interpretation is an inference problem best modelled as a knowledge constrained optimization problem over multiple statistically learned models.
- A lot more work is needed; there is a lot more that we try to do…
- Some recent samples from a research program that attempts to address some of the key scientific and engineering challenges between us and understanding natural language
- **Knowledge – Reasoning – Learning Paradigms – Scaling up**
  - Thinking about Learning, Inference and knowledge via a Constrained Optimization framework that supports "best assignment" Inference.

Check out our tools, demos, LBJava, CCM tutorial,…