Semantic Parsing: The Task, the State of the Art and the Future

ACL 2010 Tutorial

Uppsala, Sweden

Rohit J. Kate

Department of Computer Science

The University of Texas at Austin

USA

Yuk Wah Wong

Google Inc.

Pittsburgh, PA

USA

Acknowledgements

We thank Raymond Mooney, Ruifang Ge, David Chen, Wei Lu, Hwee Tou Ng and Luke Zettlemoyer for making their slides available to us.

Outline

1. Introduction to the task of semantic parsing
   a) Definition of the task
   b) Examples of application domains and meaning representation languages
   c) Distinctions from and relations to other NLP tasks
2. Semantic parsers
   a) Earlier hand-built systems
   b) Learning for semantic parsing
   c) Various forms of supervision
3. Semantic parsing beyond a sentence
   a) Learning language from perceptual contexts
   b) Using discourse contexts for semantic parsing
4. Research challenges and future directions
   a) Machine reading of documents: Connecting with knowledge representation
   b) Applying semantic parsing techniques to the Semantic Web
   c) Future research directions
5. Conclusions

Introduction to the Semantic Parsing Task

Semantic Parsing

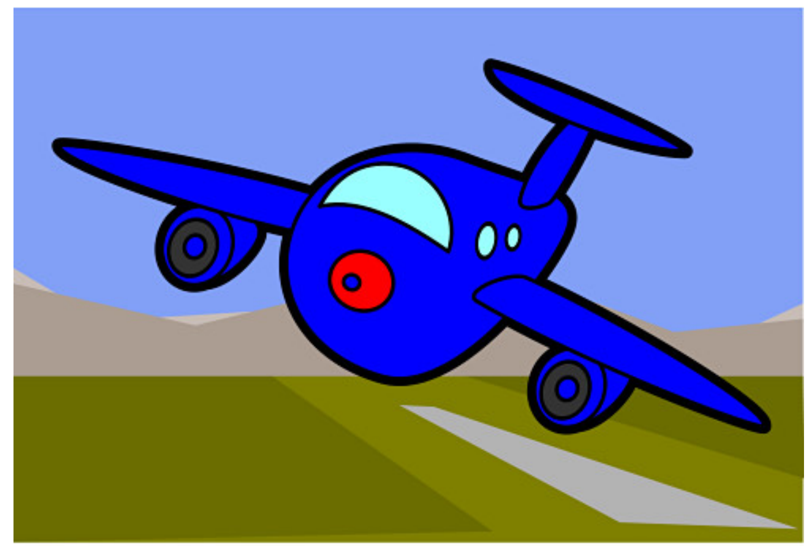
- "Semantic Parsing" is, ironically, a semantically ambiguous term
  - Semantic role labeling
  - Finding generic relations in text
  - Transforming a natural language sentence into its meaning representation

Semantic Parsing

- ***Semantic Parsing***: Transforming *natural language* (NL) sentences into computer executable complete *meaning representations* (MRs) for domain-specific applications
- Realistic semantic parsing currently entails domain dependence
- Example application domains
  - ATIS: Air Travel Information Service
  - CLang: Robocup Coach Language
  - Geoquery: A Database Query Application

ATIS: Air Travel Information Service

- Interface to an air travel database [Price, 1990]
- Widely-used benchmark for spoken language understanding



May I see all the flights from Cleveland to Dallas?

Semantic Parsing

Air-Transportation
Show: (Flight-Number)
Origin: (City "Cleveland")
Destination: (City "Dallas")

Query

NA 1439, TQ 23, …

CLang: RoboCup **C**oach **Lang**uage

- In RoboCup Coach competition teams compete to coach simulated players [*http://www.robocup.org*]
- The coaching instructions are given in a computer language called CLang [Chen et al. 2003]

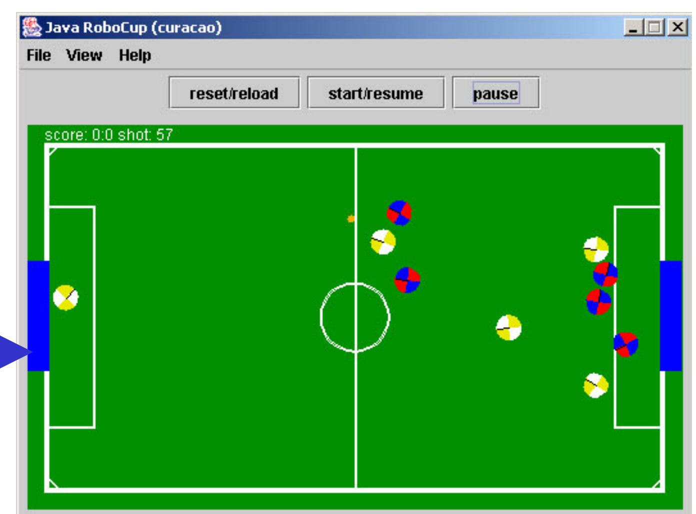
If the ball is in our goal area then player 1 should intercept it.

Semantic Parsing

(bpos (goal-area our) (do our {1} intercept))

CLang

Simulated soccer field



Geoquery: A Database Query Application

- Query application for U.S. geography database containing about 800 facts [Zelle & Mooney, 1996]

Which rivers run through the states bordering Texas?

Semantic Parsing

answer(traverse(next_to(stateid(‘texas’))))

Query

Answer

Arkansas, Canadian, Cimarron, Gila, Mississippi, Rio Grande …

What is the meaning of “meaning”?

- Representing the meaning of natural language is ultimately a difficult philosophical question
- Many attempts have been made to define generic formal semantics of natural language
  - Can they really be complete?
  - What can they do for us computationally?
  - Not so useful if the meaning of *Life* is defined as *Life’*
- Our meaning representation for semantic parsing ***does something useful*** for an application
- Procedural Semantics: The meaning of a sentence is a formal representation of a procedure that performs some action that is an appropriate response
  - Answering questions
  - Following commands

Meaning Representation Languages

- *Meaning representation language* (MRL) for an application is assumed to be present
- MRL is designed by the creators of the application to suit the application's needs independent of natural language
- CLang was designed by RoboCup community to send formal coaching instructions to simulated players
- Geoquery's MRL was based on the Prolog database

Engineering Motivation for Semantic Parsing

- Applications of domain-dependent semantic parsing
  - Natural language interfaces to computing systems
  - Communication with robots in natural language
  - Personalized software assistants
  - Question-answering systems
- Machine learning makes developing semantic parsers for specific applications more tractable
- Training corpora can be easily developed by tagging natural-language glosses with formal statements

Cognitive Science Motivation for Semantic Parsing

- Most natural-language learning methods require supervised training data that is not available to a child
  - No POS-tagged or treebank data
- Assuming a child can infer the likely meaning of an utterance from context, NL-MR pairs are more cognitively plausible training data

Distinctions from Other NLP Tasks: Deeper Semantic Analysis

- *Information extraction* involves shallow semantic analysis

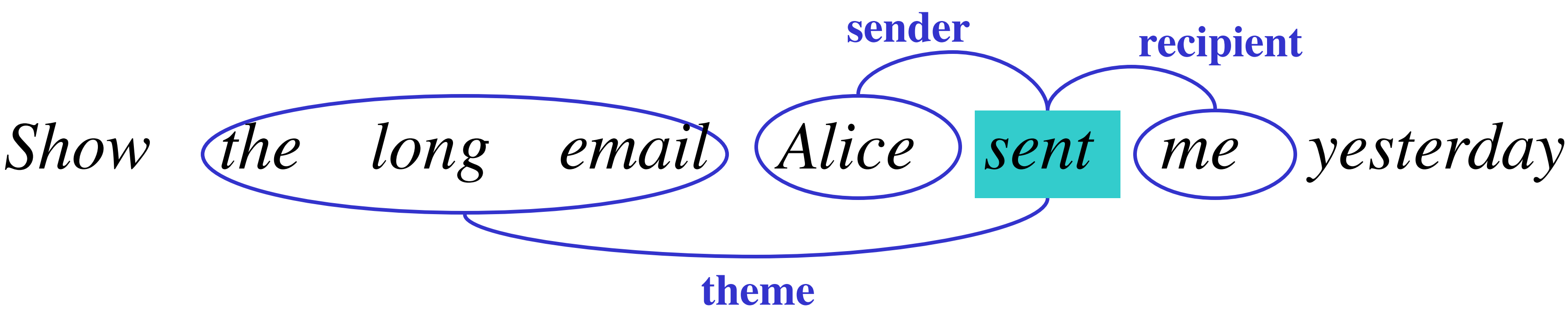
Show the long email Alice sent me yesterday

↓

Sender	Sent-to	Type	Time
Alice	*Me*	*Long*	*7/10/2010*

Distinctions from Other NLP Tasks: Deeper Semantic Analysis

- *Semantic role labeling* also involves shallow semantic analysis

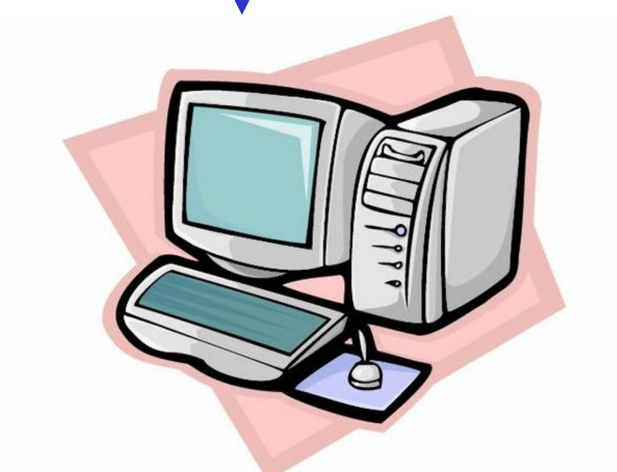


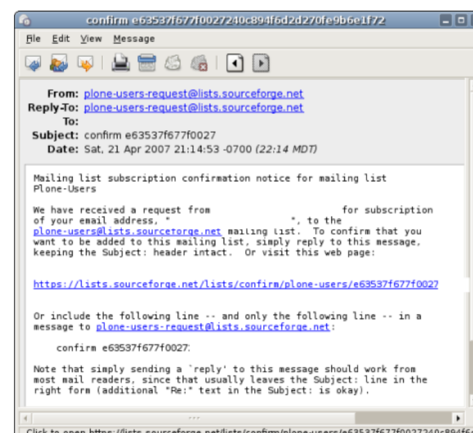
Distinctions from Other NLP Tasks: Deeper Semantic Analysis

- Semantic parsing involves deeper semantic analysis to understand the whole sentence for some application

Show the long email Alice sent me yesterday

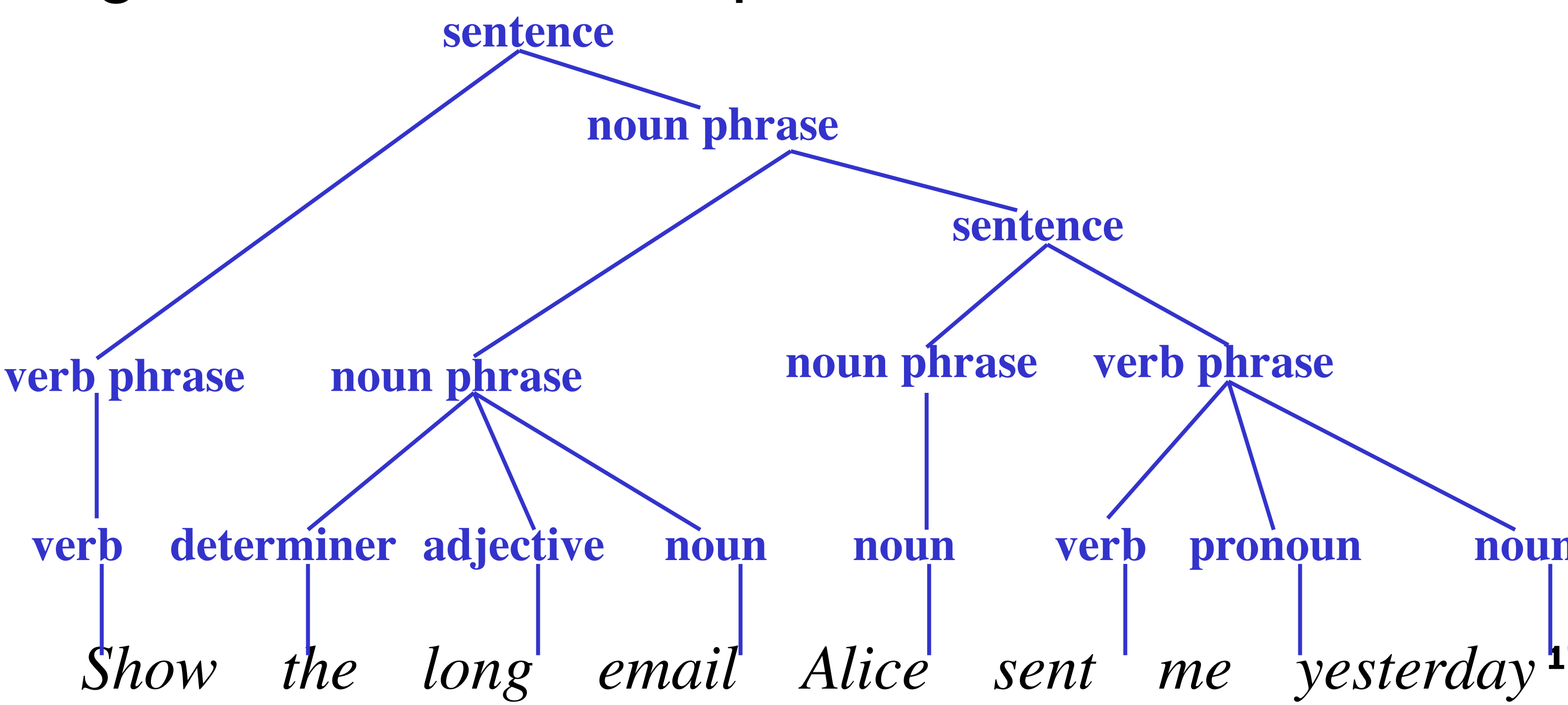
Semantic Parsing





Distinctions from Other NLP Tasks: Final Representation

- *Part-of-speech tagging*, *syntactic parsing*, *SRL* etc. generate some *intermediate linguistic representation*, typically for latter processing; in contrast, semantic parsing generates a final representation



Distinctions from Other NLP Tasks: Computer Readable Output

- The output of some NLP tasks, like *question-answering*, *summarization* and *machine translation*, are in NL and meant for *humans to read*
- Since humans are intelligent, there is some room for incomplete, ungrammatical or incorrect output in these tasks; credit is given for partially correct output
- In contrast, the output of semantic parsing is in formal language and is meant for *computers to read*; it is critical to get the exact output, strict evaluation with no partial credit

Distinctions from Other NLP Tasks

- Shallow semantic processing
  - Information extraction
  - Semantic role labeling
- Intermediate linguistic representations
  - Part-of-speech tagging
  - Syntactic parsing
  - Semantic role labeling
- Output meant for humans
  - Question answering
  - Summarization
  - Machine translation

Relations to Other NLP Tasks: Word Sense Disambiguation

- Semantic parsing includes performing word sense disambiguation

State? River?

Which rivers run through the states bordering Mississippi?

Semantic Parsing

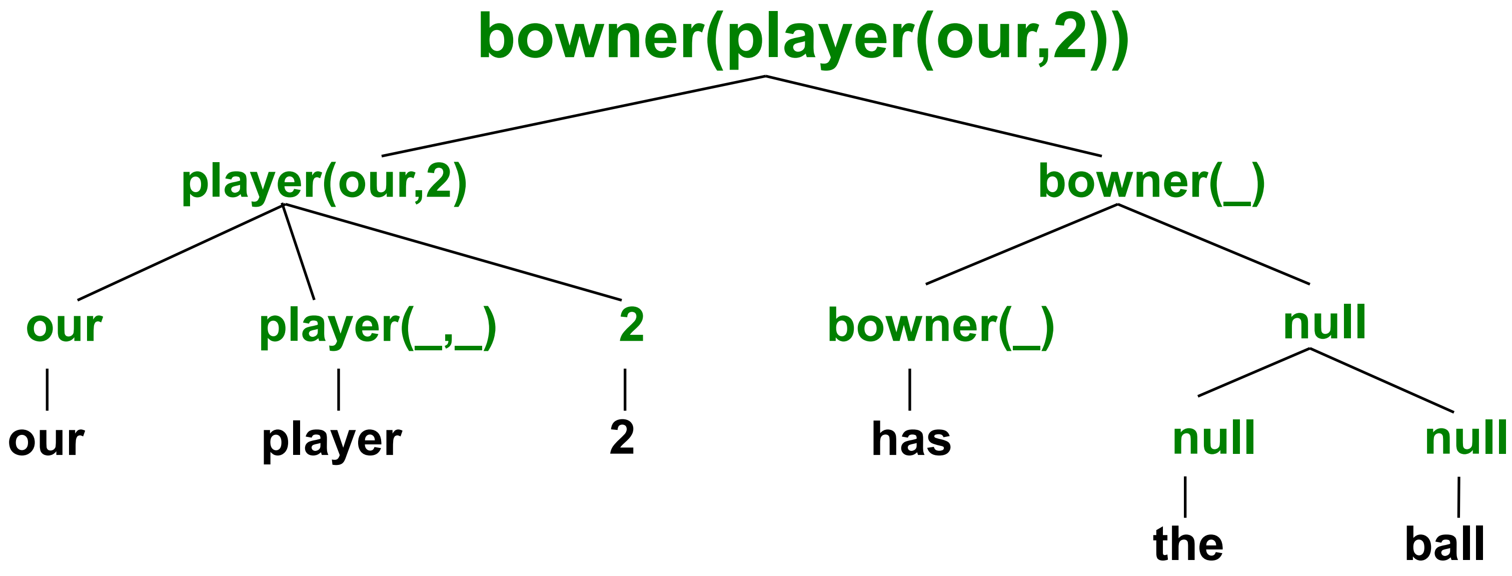
answer(traverse(next_to(stateid('mississippi'))))

Relations to Other NLP Tasks: Syntactic Parsing

- Semantic parsing inherently includes syntactic parsing but as dictated by the semantics

MR: **bowner(player(our,2))**

A semantic derivation:

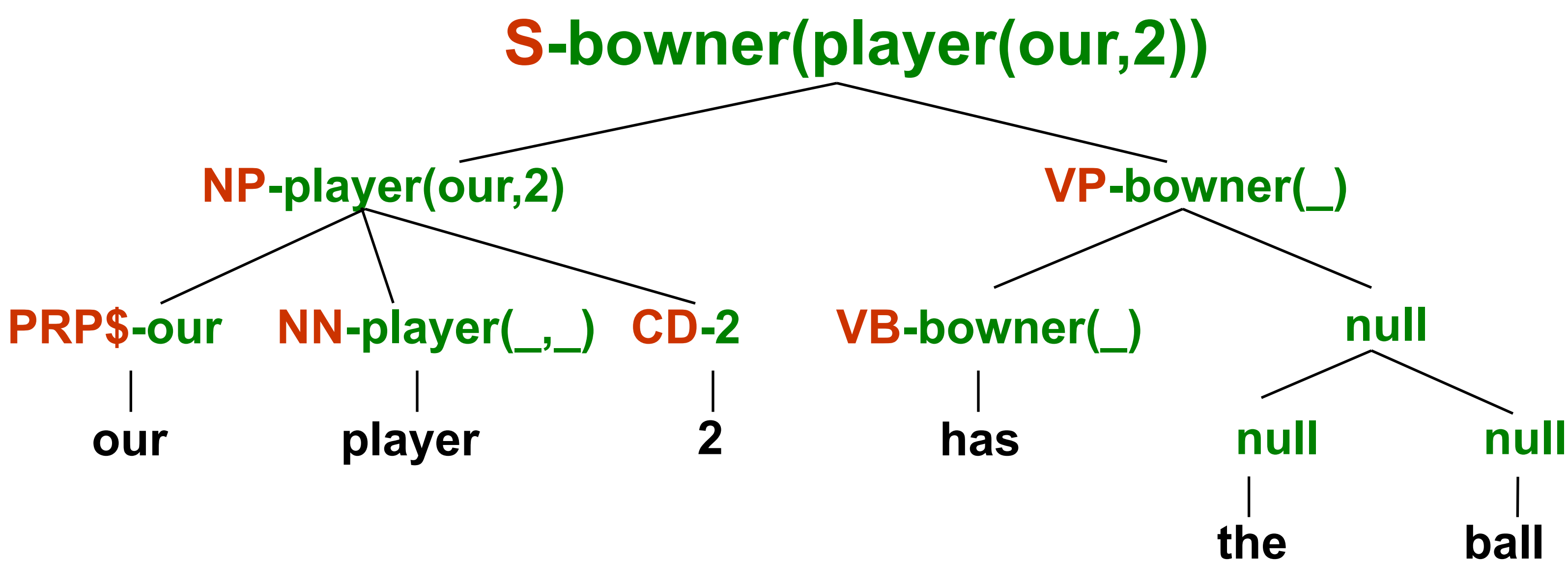


Relations to Other NLP Tasks: Syntactic Parsing

- Semantic parsing inherently includes syntactic parsing but as dictated by the semantics

MR: **bowner(player(our,2))**

A semantic derivation:



Relations to Other NLP Tasks: Machine Translation

- The MR could be looked upon as another NL [Papineni et al., 1997; Wong & Mooney, 2006]

Which rivers run through the states bordering Mississippi?

answer(traverse(next_to(stateid('mississippi'))))

Relations to Other NLP Tasks: Natural Language Generation

- Reversing a semantic parsing system becomes a natural language generation system [Jacobs, 1985; Wong & Mooney, 2007a]

Which rivers run through the states bordering Mississippi?

Semantic Parsing NL Generation

answer(traverse(next_to(stateid('mississippi'))))

Relations to Other NLP Tasks

- Tasks being performed within semantic parsing
  - Word sense disambiguation
  - Syntactic parsing as dictated by semantics
- Tasks closely related to semantic parsing
  - Machine translation
  - Natural language generation

References

Chen et al. (2003) Users manual: RoboCup soccer server manual for soccer server version 7.07 and later. Available at http://sourceforge.net/projects/sserver/

P. Jacobs (1985). PHRED: A generator for natural language interfaces. *Comp. Ling.*, 11(4):219-242.

K. Papineni, S. Roukos, T. Ward (1997). Feature-based language understanding. In *Proc. of EuroSpeech*, pp. 1435-1438. Rhodes, Greece.

P. Price (1990). Evaluation of spoken language systems: The ATIS domain. In *Proc. of the Third DARPA Speech and Natural Language Workshop*, pp. 91-95.

Y. W. Wong, R. Mooney (2006). Learning for semantic parsing with statistical machine translation. In *Proc. of HLT-NAACL*, pp. 439-446. New York, NY.

Y. W. Wong, R. Mooney (2007a). Generation by inverting a semantic parser that uses statistical machine translation. In *Proc. of NAACL-HLT*, pp. 172-179. Rochester, NY.

J. Zelle, R. Mooney (1996). Learning to parse database queries using inductive logic programming. In *Proc. of AAAI*, pp. 1050-1055. Portland, OR.

Outline

1. Introduction to the task of semantic parsing
   a) Definition of the task
   b) Examples of application domains and meaning representation languages
   c) Distinctions from and relations to other NLP tasks
2. Semantic parsers
   a) Earlier hand-built systems
   b) Learning for semantic parsing
   c) Various forms of supervision
3. Semantic parsing beyond a sentence
   a) Learning language from perceptual contexts
   b) Using discourse contexts for semantic parsing
4. Research challenges and future directions
   a) Machine reading of documents: Connecting with knowledge representation
   b) Applying semantic parsing techniques to the Semantic Web
   c) Future research directions
5. Conclusions

Earlier Hand-Built Systems

Lunar (Woods et al., 1972)

- English as a query language for a 13,000-entry lunar geology database
  - As opposed to English-like formal languages
- Syntactic analysis followed by semantic interpretation
- Meaning representation: Non-standard logic with quantifiers modeled on English determiners
- System contains:
  - Grammar for a subset of English
  - Semantic interpretation rules
  - Dictionary of 3,500 words

Lunar (Woods et al., 1972)

How many breccias contain olivine

```
(FOR THE X12 /
   (SEQL
        (NUMBER X12 / (SEQ TYPECS) :
               (CONTAIN X12 (NPR* X14 / (QUOTE OLIV))
                      (QUOTE NIL))))) : T ;
   (PRINTOUT X12))
```

→ (5)

What are they

```
(FOR EVERY X12 / (SEQ TYPECS) :
   (CONTAIN X12 (NPR* X14 / (QUOTE OLIV))
        (QUOTE NIL)) ; (PRINTOUT X12))
```

→ S10019, S10059, S10065, S10067, S10073

Chat-80 (Warren & Pereira, 1982)

- Interface to a 1,590-entry world geography database
- Translating English into logical forms:
  - Syntactic analysis
  - Slot filling
  - Quantification scoping
- Query planning: Transforming logical forms into efficient Prolog programs
- Meaning representation: Prolog with standard quantifiers
- System contains:
  - Dictionary of 100 domain dependent words
  - Dictionary of 50 domain independent words
  - Grammar rules

Chat-80 (Warren & Pereira, 1982)

Which countries bordering the Mediterranean border Asian countries?

Logical form:

```
answer(C) <= country(C) & borders(C, mediterranean) &
   exists(C1, country(C1) & asian(C1) & borders(C, C1))
```

After query planning:

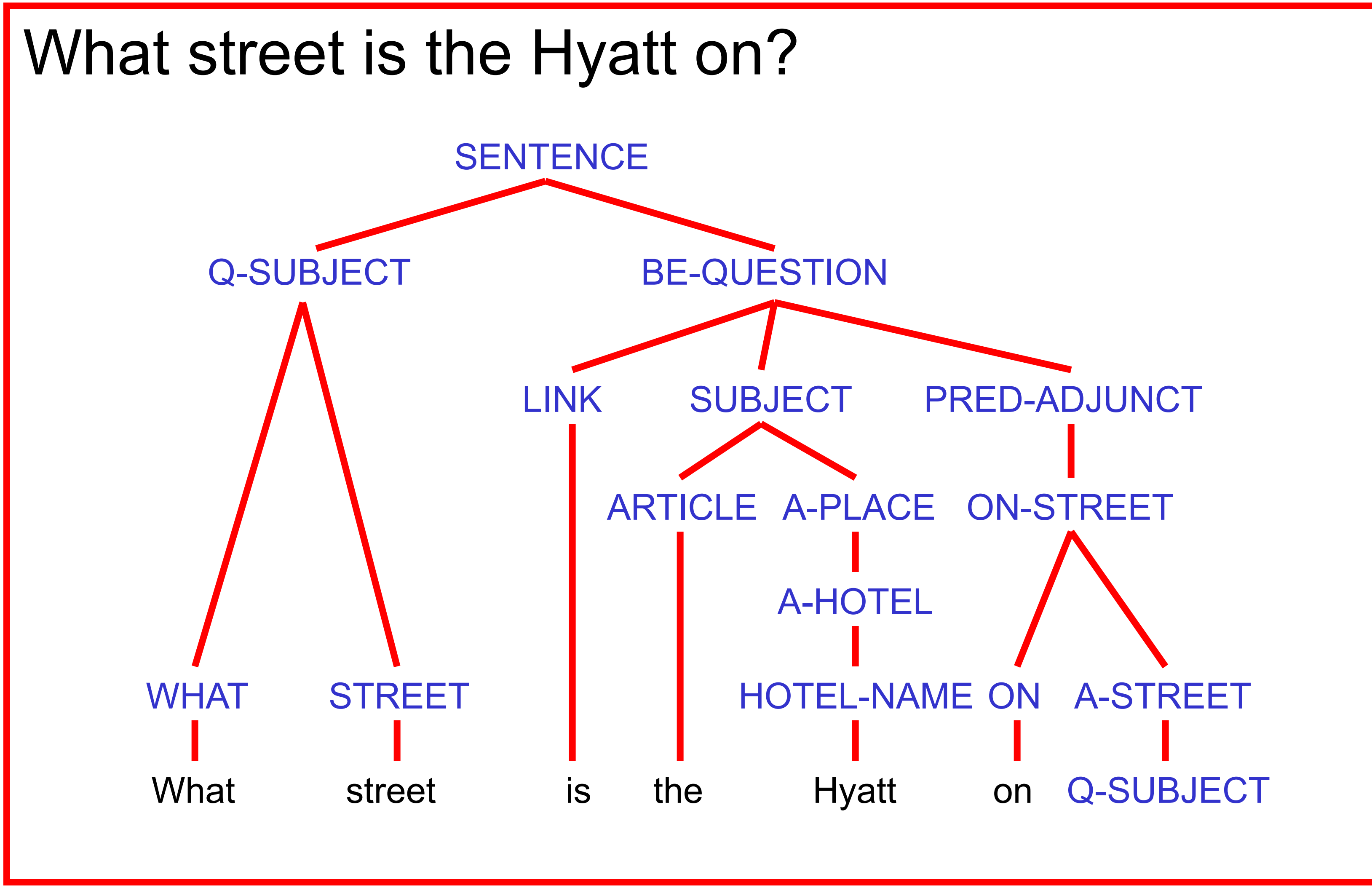
```
answer(C) <= borders(C, mediterranean) & {country(C)} &
   {borders(C, C1) & {asian(C1) & {country(C1)}}}
```

[Reads: Generate C bordering the mediterranean, then check that C is a country, and then check that it is possible to generate C1 bordering C, and then check that …]

Tina (Seneff, 1992)

- System contains:
  - Context-free grammar augmented with features that enforce syntactic and semantic constraints
  - Trained probabilities for transition networks
- Ported to multiple domains:
  - Resource management
  - City navigation
  - Air travel
- Porting to a new domain:
  - Parse new sentences one by one
  - Add context-free rules whenever a parse fails
  - Requires familiarity with grammar structure
  - Takes one person-month

Tina (Seneff, 1992)



References

J. Dowding, R. Moore, F. Andry, D. Moran (1994). Interleaving syntax and semantics in an efficient bottom-up parser. In *Proc. of ACL*, pp. 110-116. Las Cruces, NM.

S. Seneff (1992). TINA: A natural language system for spoken language applications. *Comp. Ling.*, 18(1):61-86.

W. Ward, S. Issar (1996). Recent improvement in the CMU spoken language understanding system. In *Proc. of the ARPA HLT Workshop*, pp. 213-216.

D. Warren, F. Pereira (1982). An efficient easily adaptable system for interpreting natural language queries. *American Journal of CL*, 8(3-4):110-122.

W. Woods, R. Kaplan, B. Nash-Webber (1972). The lunar sciences natural language information system: Final report. Tech. Rep. 2378, BBN Inc., Cambridge, MA.

Outline

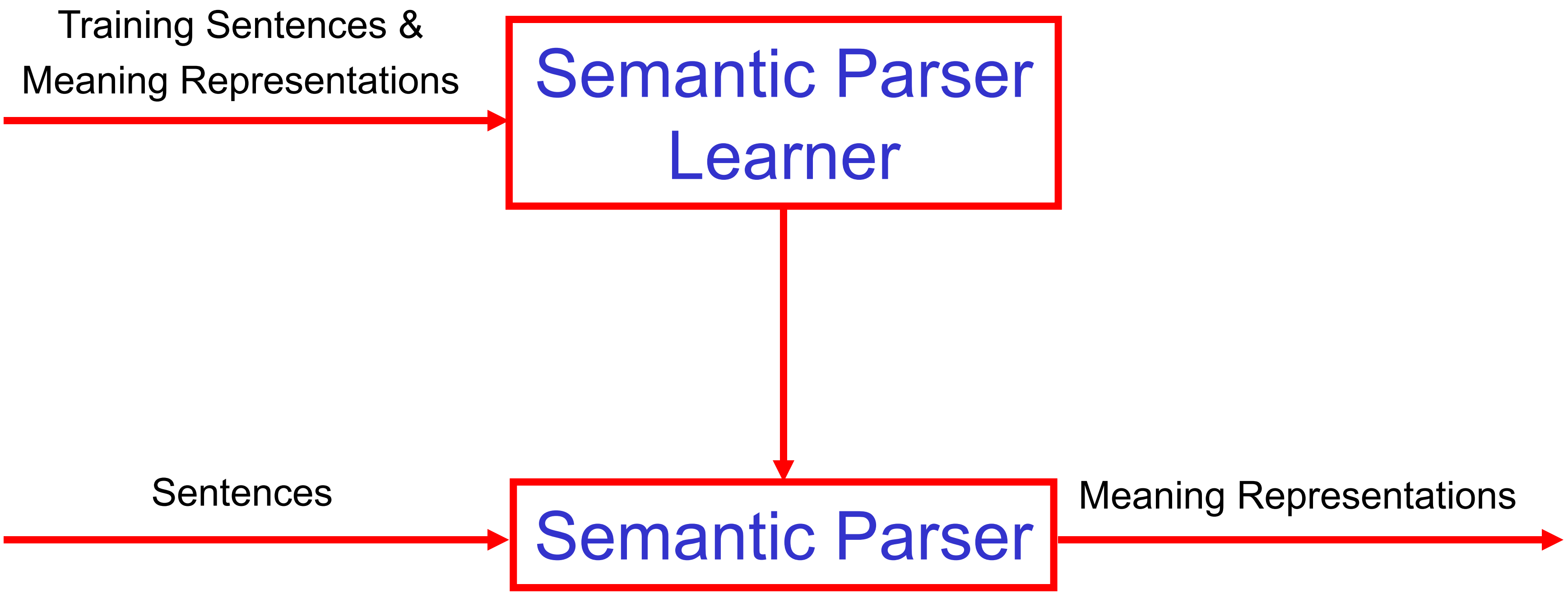
1. Introduction to the task of semantic parsing
    a) Definition of the task
    b) Examples of application domains and meaning representation languages
    c) Distinctions from and relations to other NLP tasks
2. Semantic parsers
    a) Earlier hand-built systems
    b) Learning for semantic parsing
        I. Semantic parsing learning task
        II. Early semantic parser learners
        III. Recent semantic parser learners
        IV. Exploiting syntax for semantic parsing
        V. Underlying commonalities and differences between semantic parsers
    c) Various forms of supervision
3. Semantic parsing beyond a sentence
    a) Learning language from perceptual contexts
    b) Using discourse contexts for semantic parsing
4. Research challenges and future directions
    a) Machine reading of documents: Connecting with knowledge representation
    b) Applying semantic parsing techniques to the Semantic Web
    c) Future research directions
5. Conclusions

Learning for Semantic Parsing

Motivations

- Manually programming robust semantic parsers is difficult
- It is easier to develop training corpora by associating natural-language sentences with meaning representations
- The increasing availability of training corpora, and the decreasing cost of computation, relative to engineering cost, favor the learning approach

Learning Semantic Parsers



Data Collection for ATIS

- Air travel planning scenarios (Hirschman, 1992)

> *You have 3 days for job hunting, and you have arranged job interviews in 2 different cities! Start from City-A and plan the flight and ground transportation itinerary to City-B and City-C, and back home to City-A.*

- Use of human wizards:
  - Subjects were led to believe they were talking to a fully automated systems
  - Human transcription and error correction behind the scene
- A group at SRI responsible for database reference answers
- Collected more than 10,000 utterances and 1,000 sessions for ATIS-3

Sample Session in ATIS

- may i see all the flights from cleveland to , dallas
- can you show me the flights that leave before noon , only
- could you sh- please show me the types of aircraft used on these flights

Air-Transportation

Show: (Aircraft)

Origin: (City "Cleveland")

Destination: (City "Dallas")

Departure-Time: (< 1200)

Chanel (Kuhn & De Mori, 1995)

- Consists of a set of decision trees
- Each tree builds part of a meaning representation
- Some trees decide whether a given attribute should be displayed in query results
- Some trees decide the semantic role of a given substring
  - Each correpsonds to a query constraint

Chanel (Kuhn & De Mori, 1995)

show me TIME flights from CITY1 to CITY2 and how much they cost

Tree 1:
Display aircraft_code?

Tree 23:
Display fare_id?

Tree 114:
Display booking_class?

CITY tree:
For each CITY: origin, dest, or stop?

TIME tree:
For each TIME: arrival or departure?

Display Attributes: {flight_id, fare_id}
Constraints: {from_airport = CITY1, to_airport = CITY2, departure_time = TIME}

Statistical Parsing (Miller et al., 1996)

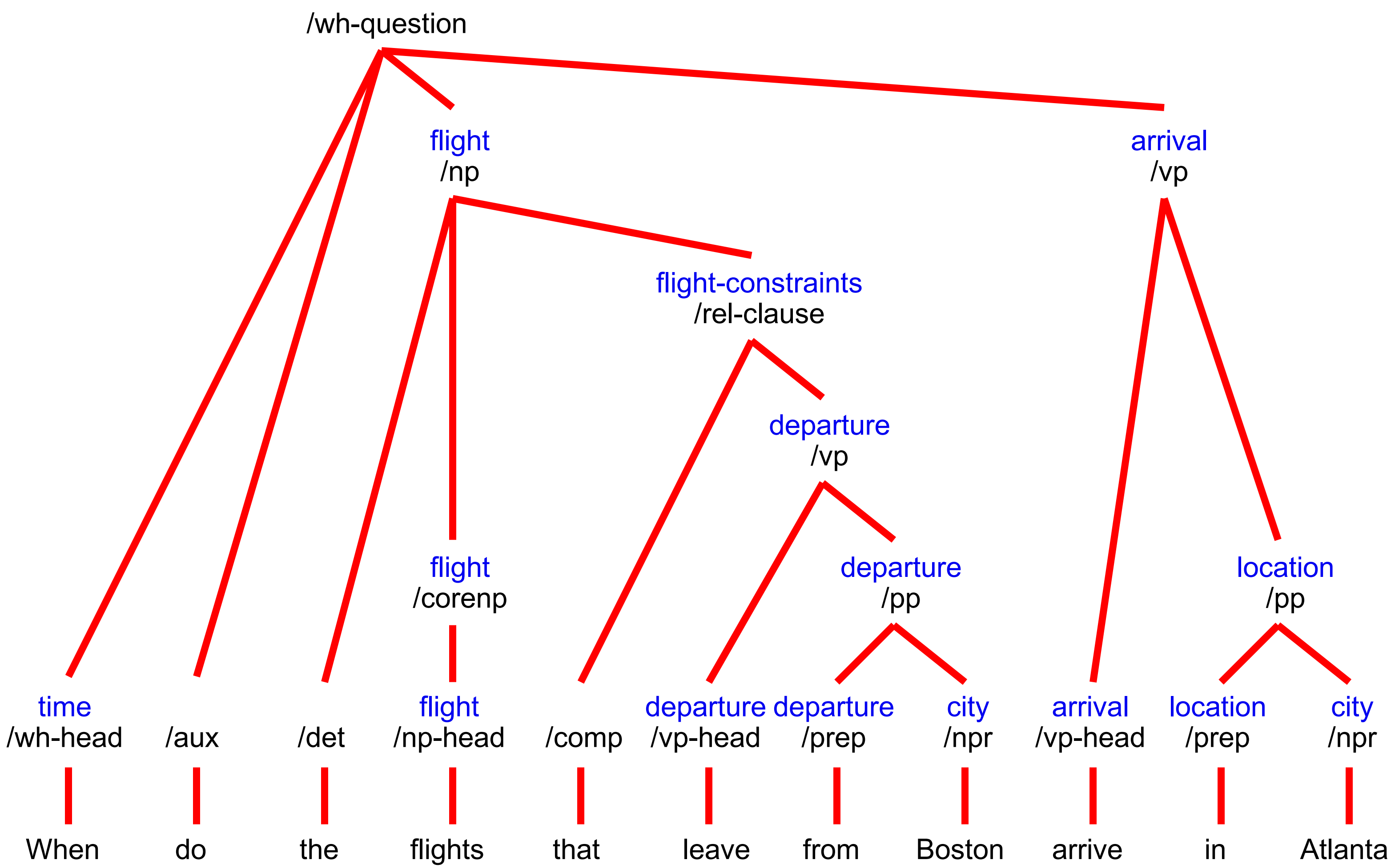
- Find most likely meaning M_0, given words W and history H

$$
\begin{aligned}
M_0 &= \arg\max_{M} P(M|W,H) \\
&= \arg\max_{M} \max_{M',T} P(M|H,M')P(M'|T)P(W|T)
\end{aligned}
$$

(M': pre-discourse meaning, T: parse tree)

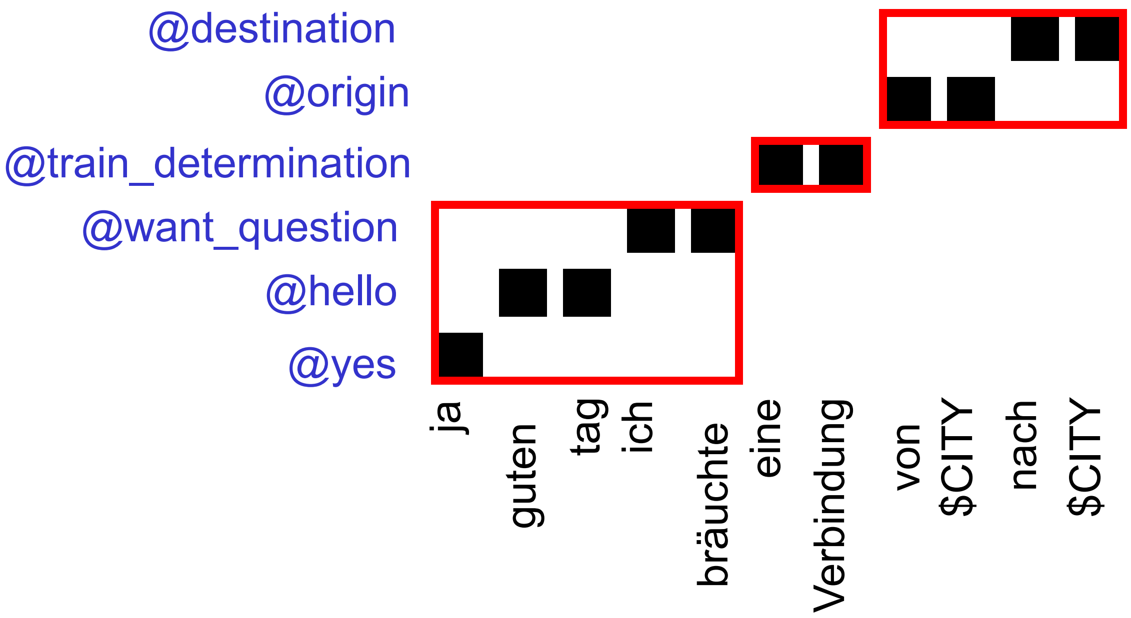
- Three successive stages: parsing, semantic interpretation, and discourse
- Parsing model similar to Seneff (1992)
  - Requires annotated parse trees for training

Statistical Parsing (Miller et al., 1996)



Machine Translation

- Translation from a natural-language source sentence to a formal-language target sentence
- Papineni et al. (1997), Macherey et al. (2001)



Other Approaches

- Inductive logic programming (Zelle & Mooney, 1996)
- Hierarchical translation (Ramaswamy & Kleindienst, 2000)
- Composite of HMM and CFG (Wang & Acero, 2003)
- Hidden vector state model (He & Young, 2006)
- Constraint satisfaction (Popescu, 2004)

Recent Approaches

- Different levels of supervision
  - Ranging from fully supervised to unsupervised
- Advances in machine learning
  - Structured learning
  - Kernel methods
- Grammar formalisms
  - Combinatory categorial grammars
  - Sychronous grammars
- Unified framework for handling various phenomena
  - Spontaneous speech
  - Discourse
  - Perceptual context
  - Generation

References

Y. He, S. Young (2006). Spoken language understanding using the hidden vector state model. *Speech Communication*, 48(3-4):262-275.

L. Hirschman (1992). Multi-site data collection for a spoken language corpus. In *Proc. of HLT Workshop on Speech and Natural Language*, pp. 7-14. Harriman, NY.

R. Kuhn, R. De Mori (1995). The application of semantic classification trees to natural language understanding. *IEEE Trans. on PAMI*, 17(5):449-460.

K. Macherey, F. Och, H. Ney (2001). Natural language understanding using statistical machine translation. In *Proc. of Eurospeech*, pp. 2205-2208. Aalborg, Denmark.

S. Miller, D. Stallard, R. Bobrow, R. Schwartz (1996). A fully statistical approach to natural language interfaces. In *Proc. of ACL*, pp. 55-61. Santa Cruz, CA.

K. Papineni, S. Roukos, T. Ward (1997). Feature-based language understanding. In *Proc. of Eurospeech*, pp. 1435-1438. Rhodes, Greece.

References

A. Popescu, A. Armanasu, O. Etzioni, D. Ko, A. Yates (2004). Modern natural language interfaces to databases: Composing statistical parsing with semantic tractability. In *Proc. of COLING*. Geneva, Switzerland.

G. Ramaswamy, J. Kleindienst (2000). Hierarchical feature-based translation for scalable natural language understanding. In *Proc. of ICSLP*.

Y. Wang, A. Acero (2003). Combination of CFG and n-gram modeling in semantic grammar learning. In *Proc. of Eurospeech*, pp. 2809-2812. Geneva, Switzerland.

J. Zelle, R. Mooney (1996). Learning to parse database queries using inductive logic programming. In *Proc. of AAAI*, pp. 1050-1055. Portland, OR.

Outline

1. Introduction to the task of semantic parsing
    a) Definition of the task
    b) Examples of application domains and meaning representation languages
    c) Distinctions from and relations to other NLP tasks
2. Semantic parsers
    a) Earlier hand-built systems
    b) Learning for semantic parsing
        I. Semantic parsing learning task
        II. Early semantic parser learners
        III. Recent semantic parser learners
        IV. Exploiting syntax for semantic parsing
        V. Underlying commonalities and differences between semantic parsers
    c) Various forms of supervision
3. Semantic parsing beyond a sentence
    a) Learning language from perceptual contexts
    b) Using discourse contexts for semantic parsing
4. Research challenges and future directions
    a) Machine reading of documents: Connecting with knowledge representation
    b) Applying semantic parsing techniques to the Semantic Web
    c) Future research directions
5. Conclusions

Outline

- Zettlemoyer & Collins (2005, 2007)
  - Structured learning with combinatory categorial grammars (CCG)
- Wong & Mooney (2006, 2007a, 2007b)
  - Syntax-based machine translation methods
- Kate & Mooney (2006), Kate (2008a)
  - SVM with kernels for robust semantic parsing
- Lu et al. (2008)
  - A generative model for semantic parsing
- Ge & Mooney (2005, 2009)
  - Exploiting syntax for semantic parsing

Semantic Parsing using CCG

Combinatory Categorial Grammar

- Highly structured lexical entries
- A few general parsing rules (Steedman, 2000; Steedman & Baldridge, 2005)
- Each lexical entry is a word paired with a category

Texas := NP

borders := (S \ NP) / NP

Mexico := NP

New Mexico := NP

Parsing Rules (Combinators)

- Describe how adjacent categories are combined
- Functional application:

$$A / B \quad B \quad \Rightarrow \quad A \qquad (>)$$

$$B \quad A \backslash B \quad \Rightarrow \quad A \qquad (<)$$

$$\begin{array}{ccc}
\text{Texas} & \text{borders} & \text{New Mexico} \\
\hline
NP & (S \backslash NP) / NP & NP \\
\cline{2-3}
 & \multicolumn{2}{c}{S \backslash NP \quad >} \\
\hline
\multicolumn{3}{c}{S \quad <}
\end{array}$$

CCG for Semantic Parsing

- Extend categories with semantic types

Texas := NP : texas
borders := (S \ NP) / NP : λx.λy.borders(y, x)

- Functional application with semantics:

$$A / B : f \qquad B : a \qquad \Rightarrow \qquad A : f(a) \qquad (>)$$

$$B : a \qquad A \backslash B : f \qquad \Rightarrow \qquad A : f(a) \qquad (<)$$

Sample CCG Derivation

Texas borders New Mexico

NP
texas

(S \ NP) / NP
λx.λy.borders(y, x)

NP
new_mexico

>

S \ NP
λy.borders(y, new_mexico)

<

S
borders(texas, new_mexico)

Another Sample CCG Derivation

Texas	borders	New Mexico
NP	(S \ NP) / NP	NP
texas	λx.λy.borders(y, x)	mexico

borders New Mexico >

S \ NP

λy.borders(y, mexico)

Texas borders New Mexico <

S

borders(texas, mexico)

Probabilistic CCG for Semantic Parsing

Zettlemoyer & Collins (2005)

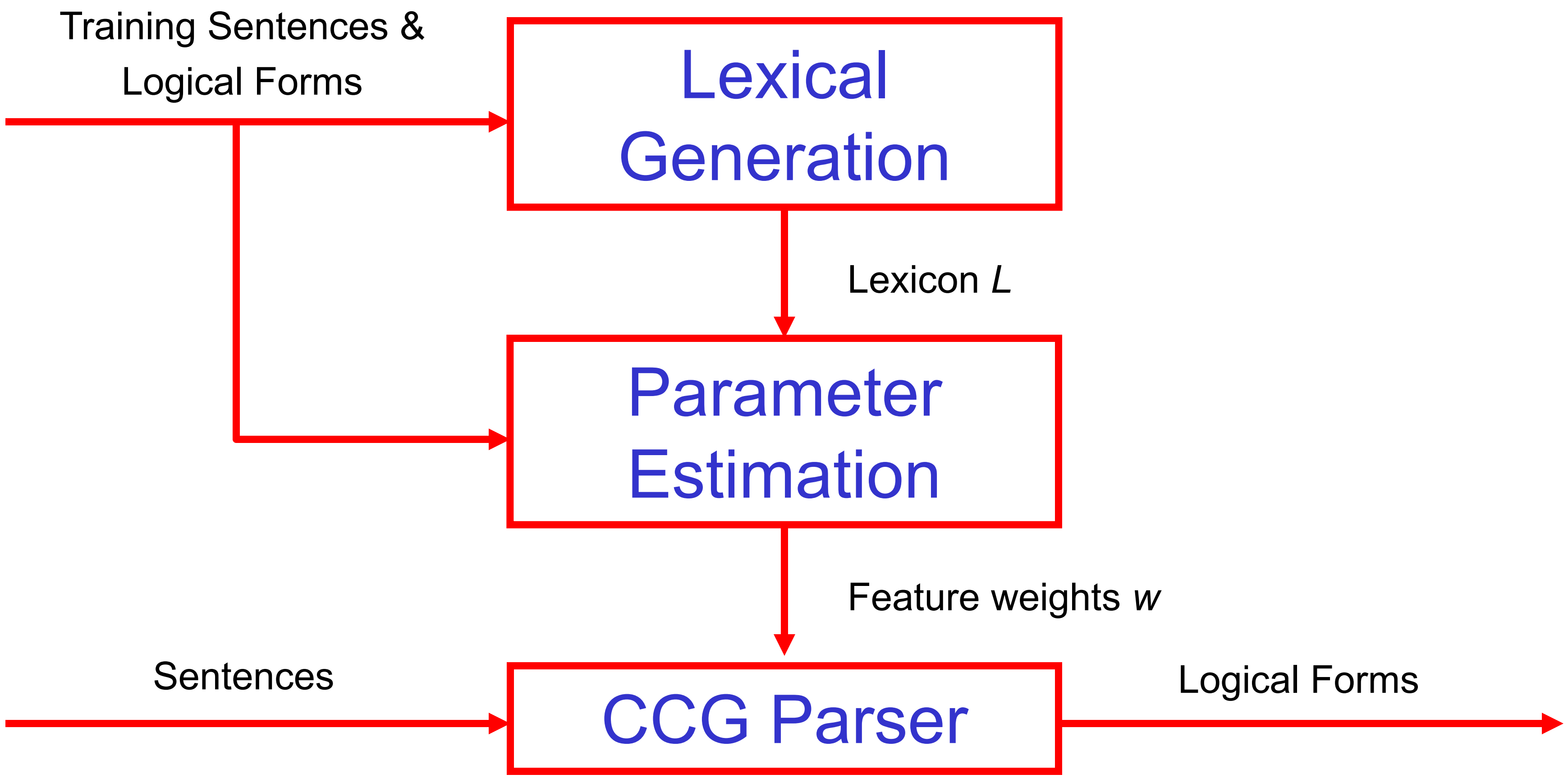
- L (lexicon) =

Texas := NP : texas
borders := (S \ NP) / NP : λx.λy.borders(y, x)
Mexico := NP : mexico
New Mexico := NP : new_mexico

- w (feature weights)

- Features:
  - $f_i(x, d)$: Number of times lexical item i is used in derivation d
- Log-linear model: $P_w(d \mid x) \propto \exp(w \cdot f(x, d))$
- Best derviation: $d^* = \text{argmax}_d\, w \cdot f(x, d)$
  - Consider all possible derivations d for the sentence x given the lexicon L

Learning Probabilistic CCG



Lexical Generation

- Input:

Texas borders New Mexico

borders(texas, new_mexico)

- Output lexicon:

Texas := NP : texas

borders := (S \ NP) / NP : λx.λy.borders(y, x)

New Mexico := NP : new_mexico

Lexical Generation

Input sentence:

Texas borders New Mexico

Output substrings:

Texas

borders

New

Mexico

Texas borders

borders New

New Mexico

Texas borders New

…

$\times$

Input logical form:

borders(texas, new_mexico)

Output categories:

NP : texas

NP : new _mexico

(S \ NP) / NP : $\lambda x.\lambda y.\text{borders}(y, x)$

(S \ NP) / NP : $\lambda x.\lambda y.\text{borders}(x, y)$

…

Category Rules

Input Trigger	Output Category
constant c	NP : c
arity one predicate p	N : $\lambda x.p(x)$
arity one predicate p	S \ NP : $\lambda x.p(x)$
arity two predicate p	(S \ NP) / NP : $\lambda x.\lambda y.p(y, x)$
arity two predicate p	(S \ NP) / NP : $\lambda x.\lambda y.p(x, y)$
arity one predicate p	N / N : $\lambda g.\lambda x.p(x) \wedge g(x)$
arity two predicate p and constant c	N / N : $\lambda g.\lambda x.p(x, c) \wedge g(x)$
arity two predicate p	(N \ N) / NP : $\lambda x.\lambda g.\lambda y.p(y, x) \wedge g(x)$
arity one function f	NP / N : $\lambda g.\text{argmax/min}(g(x), \lambda x.f(x))$
arity one function f	S / NP : $\lambda x.f(x)$

Parameter Estimation

- Maximum conditional likelihood

$$\prod_i \mathrm{P}_w(z_i|x_i) \quad = \quad \prod_i \sum_d \mathrm{P}_w(z_i, d|x_i)$$

- Derivations *d* are not annotated, treated as hidden variables
- Stochastic gradient ascent (LeCun et al., 1998)
- Keep only those lexical items that occur in the highest scoring derivations of training set

Results

- Test for correct logical forms
- Precision: # correct / total # parsed sentences
- Recall: # correct / total # sentences
- For Geoquery, 96% precision, 79% recall
- Low recall due to incomplete lexical generation:
  – Through which states does the Mississippi run?

Relaxed CCG for Spontaneous Speech

Zettlemoyer & Collins (2007)

- Learned CCG works well for grammatical sentences

Show me	the latest	flight	from Boston	to Prague	on Friday
S / NP	NP / N	N	N \ N	N \ N	N \ N

- Works less well for spontaneous speech

Boston	to Prague	the latest	on Friday
NP	N \ N	NP / N	N \ N

- Problems:
  - Flexible word order
  - Missing content words

Flexible Word Order

- Functional application:

$$A / B : f \qquad B : a \qquad \Rightarrow \qquad A : f(a) \qquad (>)$$

$$B : a \qquad A \backslash B : f \qquad \Rightarrow \qquad A : f(a) \qquad (<)$$

- Disharmonic application:

$$A \backslash B : f \qquad B : a \qquad \Rightarrow \qquad A : f(a) \qquad (>\approx)$$

$$B : a \qquad A / B : f \qquad \Rightarrow \qquad A : f(a) \qquad (<\approx)$$

$$\dfrac{\dfrac{\text{flights}}{\begin{array}{c} N \\ \lambda x.\text{flight}(x) \end{array}} \qquad \dfrac{\text{one way}}{\begin{array}{c} N / N \\ \lambda f.\lambda x.f(x) \wedge \text{one\_way}(x) \end{array}}}{\begin{array}{c} N \\ \lambda x.\text{flight}(x) \wedge \text{one\_way}(x) \end{array}} <\approx$$

Missing Content Words

- Role-hypothesizing type shifting:

$$NP : c \quad \Rightarrow \quad N \backslash N : \lambda f.\lambda x.f(x) \wedge p(x, c) \qquad (\mathbf{T}_R)$$

$$
\begin{array}{ccc}
\text{flights} & \text{Boston} & \text{to Prague} \\
\hline
N & NP & N \backslash N \\
\lambda x.flight(x) & BOS & \lambda f.\lambda x.f(x) \wedge to(x, PRG)
\end{array}
$$

$$
\begin{array}{c}
NP : BOS \\
\hline
N \backslash N \\
\lambda f.\lambda x.f(x) \wedge from(x, BOS)
\end{array}\ \mathbf{T}_R
$$

$$
\begin{array}{c}
\text{flights} \quad \text{Boston} \\
\hline
N \\
\lambda x.flight(x) \wedge from(x, BOS)
\end{array}\ <
$$

$$
\begin{array}{c}
\text{flights} \quad \text{Boston} \quad \text{to Prague} \\
\hline
N \\
\lambda x.flight(x) \wedge from(x, BOS) \wedge to(x, PRG)
\end{array}\ <
$$

Complete Derivation

Boston	to Prague	the latest	on Friday
NP	N \ N	NP / N	N \ N
BOS	$\lambda f.\lambda x.f(x) \wedge to(x, PRG)$	$\lambda f.argmax(\lambda x.f(x), \lambda x.time(x))$	$\lambda f.\lambda x.f(x) \wedge day(x, FRI)$

T_R

N \ N

$\lambda f.\lambda x.f(x) \wedge from(x, BOS)$

T_N

N

$\lambda x.day(x, FRI)$

$<B$

N \ N

$\lambda f.\lambda x.f(x) \wedge from(x, BOS) \wedge to(x, PRG)$

$<\approx B$

NP \ N

$\lambda f.argmax(\lambda x.f(x) \wedge from(x, BOS) \wedge to(x, PRG), \lambda x.time(x))$

$>\approx$

NP

$argmax(\lambda x.from(x, BOS) \wedge to(x, PRG) \wedge day(x, FRI), \lambda x.time(x))$

Parameter Estimation

- New parsing rules can significantly relax word order
  - Introduce features to count the number of times each new parsing rule is used in a derivation
- Error-driven, perceptron-style parameter updates

1. Let $d^* = \arg\max_d w \cdot f(x_i, d)$
2. If $z(d^*) = z_i$, go to the next example
3. Let $d' = \arg\max_{d \text{ s.t. } z(d)=z_i} w \cdot f(x_i, d)$
4. $w = w + f(x_i, d') - f(x_i, d^*)$

Results

- For ATIS, 91% precision, 82% recall
- For Geoquery, 95% precision, 83% recall
  - Up from 79% recall

References

Y. LeCun, L. Bottou, Y. Bengio, P. Haffner (1998). Gradient-based learning applied to document recognition. In *Proc. of the IEEE*, 86(11):2278-2324.

M. Steedman (2000). *The Syntactic Process*. MIT Press.

M. Steedman, J. Baldridge (2005). Combinatory categorial grammar. To appear in: R. Borsley, K. Borjars (eds.) *Non-Transformational Syntax*, Blackwell.

L. Zettlemoyer, M. Collins (2005). Learning to map sentences to logical form: Structured classification with probabilistic categorial grammars. In *Proc. of UAI*. Edinburgh, Scotland.

L. Zettlemoyer, M. Collins (2007). Online learning of relaxed CCG grammars for parsing to logical form. In *Proc. of EMNLP-CoNLL*, pp. 678-687. Prague, Czech Republic.

Semantic Parsing vs Machine Translation

WASP: A Machine Translation Approach to Semantic Parsing

Wong & Mooney (2006)

- Based on a semantic grammar of the natural language
- Uses machine translation techniques
  - Synchronous context-free grammars (SCFG)
  - Word alignments (Brown et al., 1993)

Synchronous Context-Free Grammar

- Developed by Aho & Ullman (1972) as a theory of compilers that combines syntax analysis and code generation in one phase
  - Formal to formal languages
- Used for syntax-based machine translation (Wu, 1997; Chiang 2005)
  - Natural to natural languages
- Generates a pair of strings in a derivation

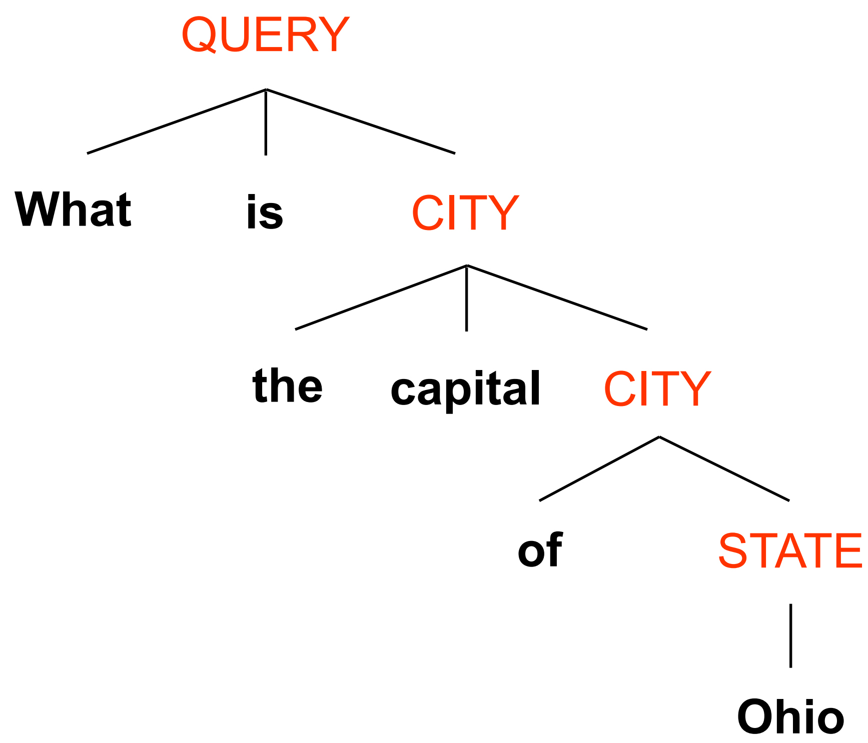
Context-Free Semantic Grammar

QUERY → **What is** CITY

CITY → **the capital** CITY

CITY → **of** STATE

STATE → **Ohio**



Sample SCFG Production

Natural language　　　　Formal language

QUERY → **What is** CITY / answer(CITY)

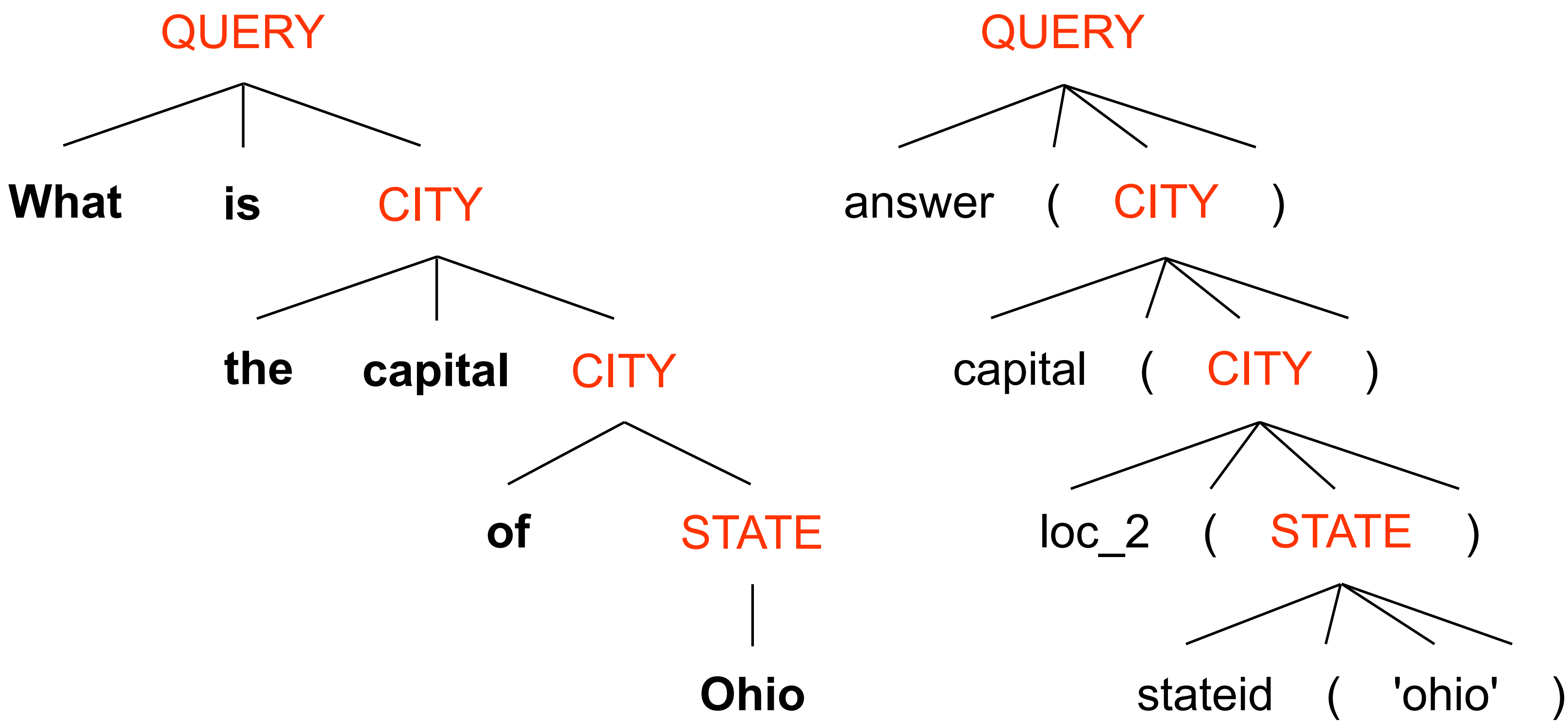
Sample SCFG Derivation

QUERY —————————— QUERY

Sample SCFG Derivation

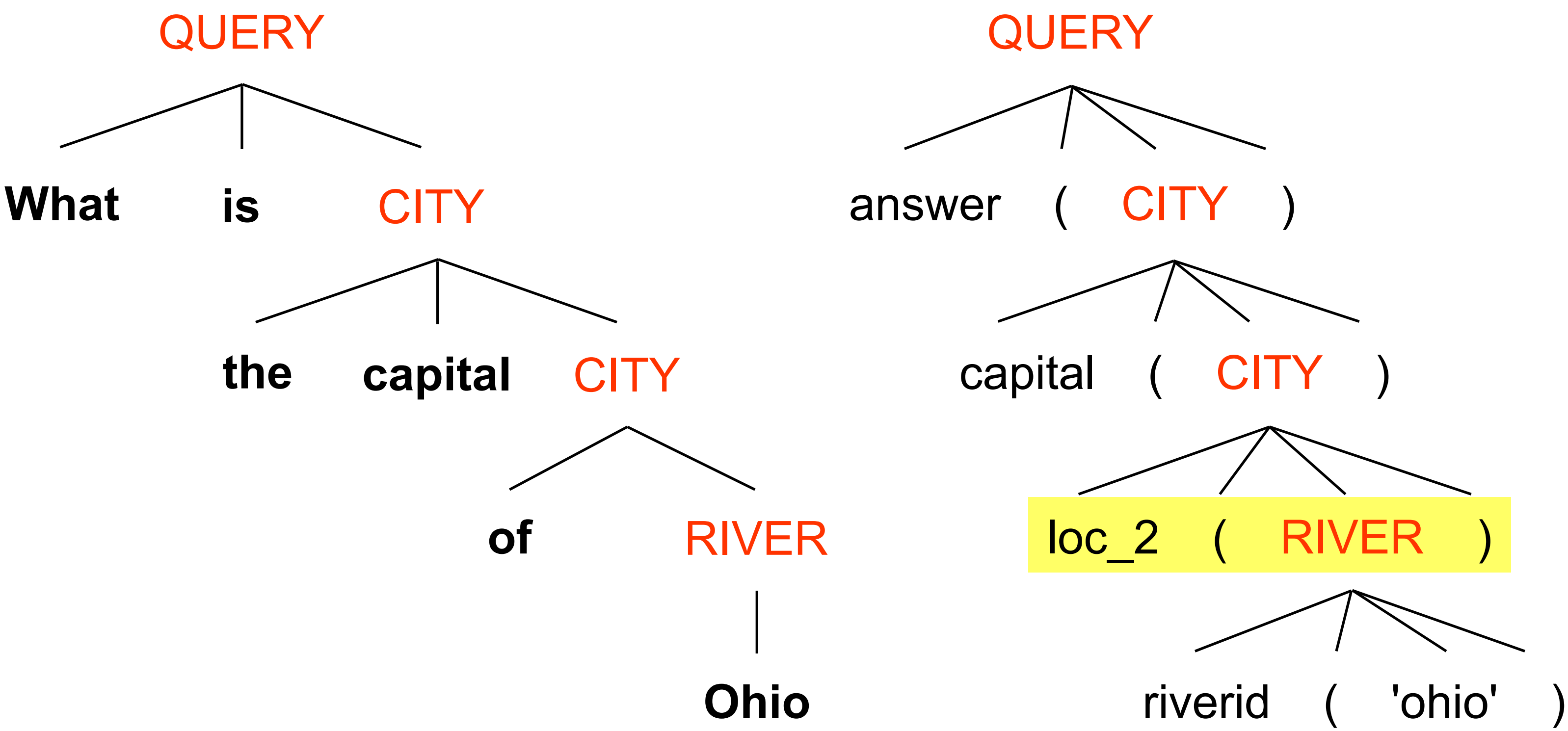
QUERY → **What is** CITY / answer(CITY)

Sample SCFG Derivation



What is the capital of Ohio answer(capital(loc_2(stateid('ohio'))))

Another Sample SCFG Derivation



What is the capital of Ohio answer(capital(loc_2(riverid('ohio'))))

Probabilistic SCFG for Semantic Parsing

- *S* (start symbol) = QUERY
- *L* (lexicon) =

QUERY → **What is** CITY / answer(CITY)

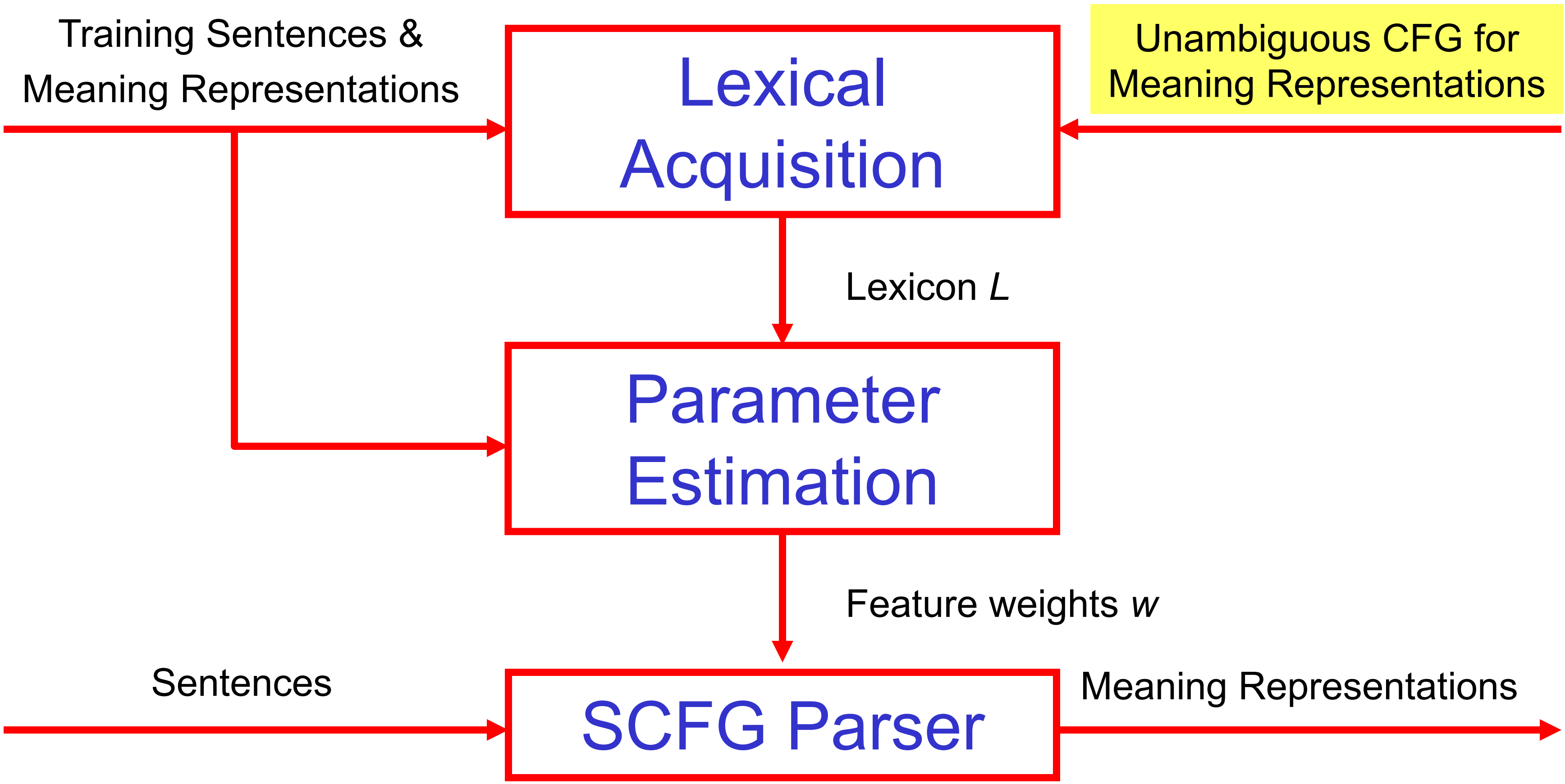
CITY → **the capital** CITY / capital(CITY)

CITY → **of** STATE / loc_2(STATE)

STATE → **Ohio** / stateid('ohio')

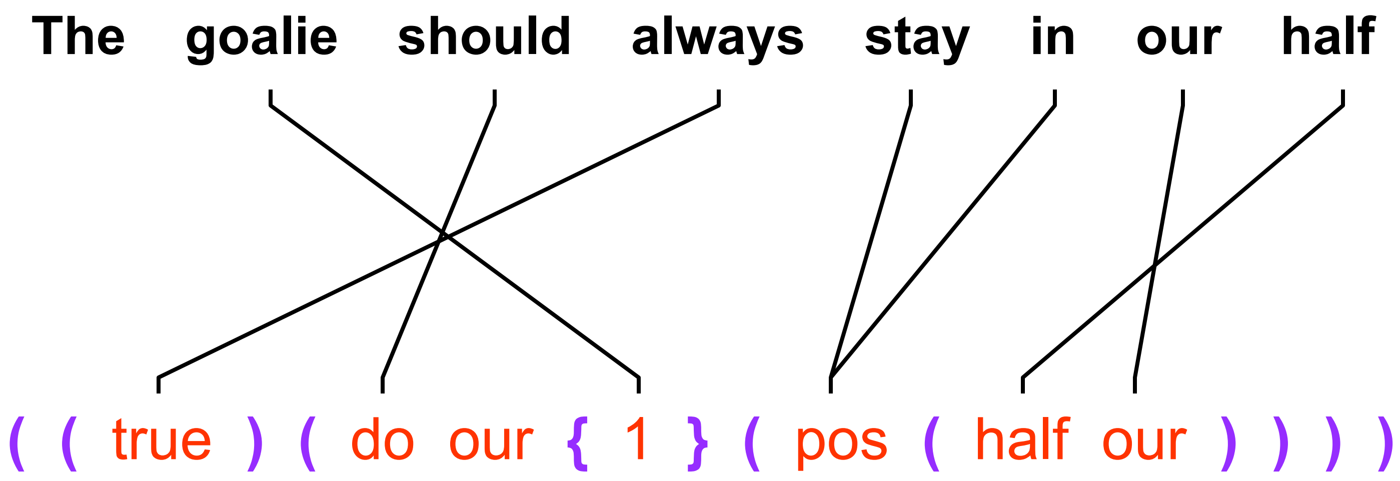
- *w* (feature weights)
- Features:
  - $f_i(x, d)$: Number of times production *i* is used in derivation *d*
- Log-linear model: $P_w(d \mid x) \propto \exp(w \,.\, f(x, d))$
- Best derviation: $d^* = \text{argmax}_d \; w \,.\, f(x, d)$

Learning Probabilistic SCFG

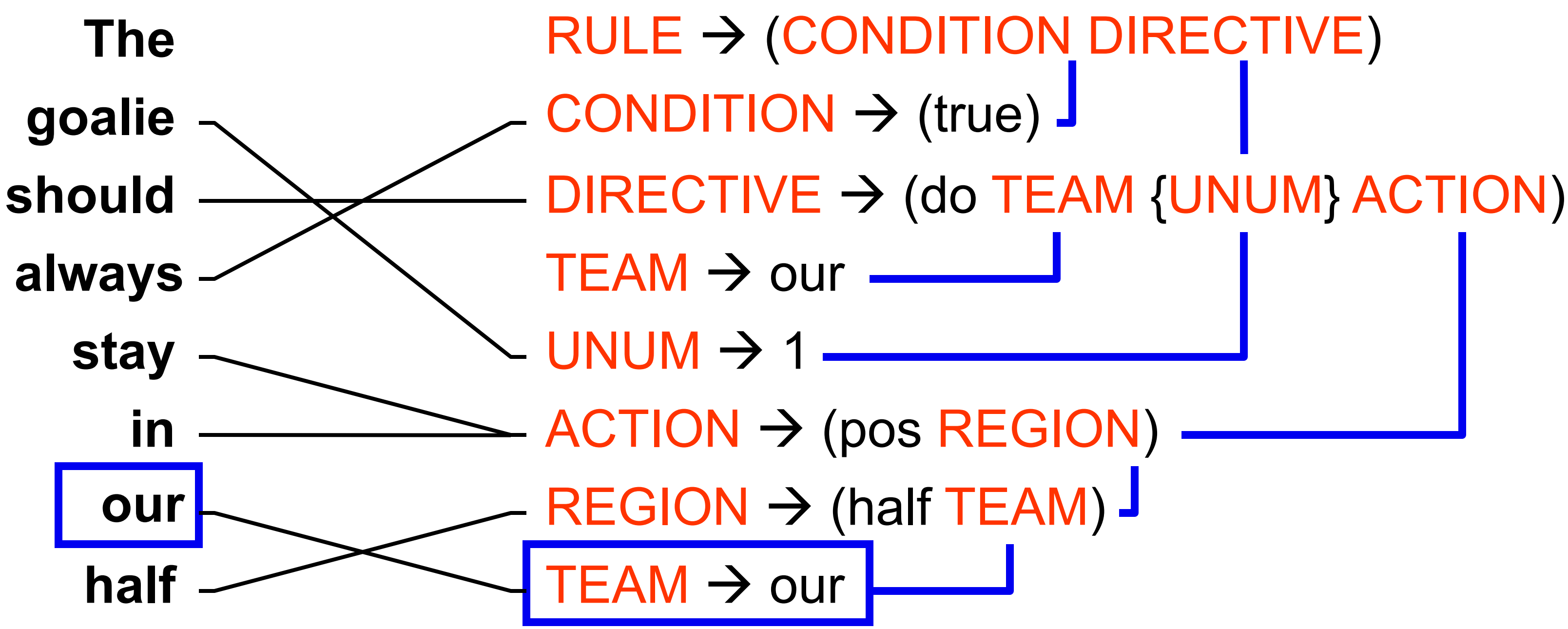


Lexical Acquisition

- SCFG productions are extracted from word alignments between training sentences and their meaning representations



Extracting SCFG Productions



Extracting SCFG Productions

The
goalie
should
always
stay
in
TEAM
half

RULE → (CONDITION DIRECTIVE)
CONDITION → (true)
DIRECTIVE → (do TEAM {UNUM} ACTION)
TEAM → our
UNUM → 1
ACTION → (pos REGION)
REGION → (half TEAM)

TEAM → **our** / our

Extracting SCFG Productions

The
goalie
should
always
stay
in
TEAM
half

RULE → (CONDITION DIRECTIVE)
CONDITION → (true)
DIRECTIVE → (do TEAM {UNUM} ACTION)
TEAM → our
UNUM → 1
ACTION → (pos REGION)
REGION → (half TEAM)

Extracting SCFG Productions

The
goalie
should
always
stay
in
REGION

RULE → (CONDITION DIRECTIVE)
CONDITION → (true)
DIRECTIVE → (do TEAM {UNUM} ACTION)
TEAM → our
UNUM → 1
ACTION → (pos REGION)

REGION → TEAM **half** / (half TEAM)

Output SCFG Productions

TEAM → **our** / our

REGION → TEAM **half** / (half TEAM)

ACTION → **stay in** REGION / (pos REGION)

UNUM → **goalie** / 1

RULE → **[the]** UNUM **should always** ACTION
/ ((true) (do our {UNUM} ACTION))

- Phrases can be non-contiguous

Handling Logical Forms with Variables

Wong & Mooney (2007b)

FORM → **state** / $\lambda x.\mathrm{state}(x)$

FORM → **by area** / $\lambda x.\lambda y.\mathrm{area}(x, y)$

FORM → **[the] smallest** FORM FORM
/ $\lambda x.\mathrm{smallest}(y, (\mathrm{FORM}(x), \mathrm{FORM}(x, y))$

QUERY → **what is** FORM
/ $\mathrm{answer}(x, \mathrm{FORM}(x))$

- Operators for variable binding

Generation by Inverting WASP

Wong & Mooney (2007a)

- Mapping a meaning representation to natural language
- Can be seen as inverse of semantic parsing
- Same synchronous grammar is used for both semantic parsing and generation

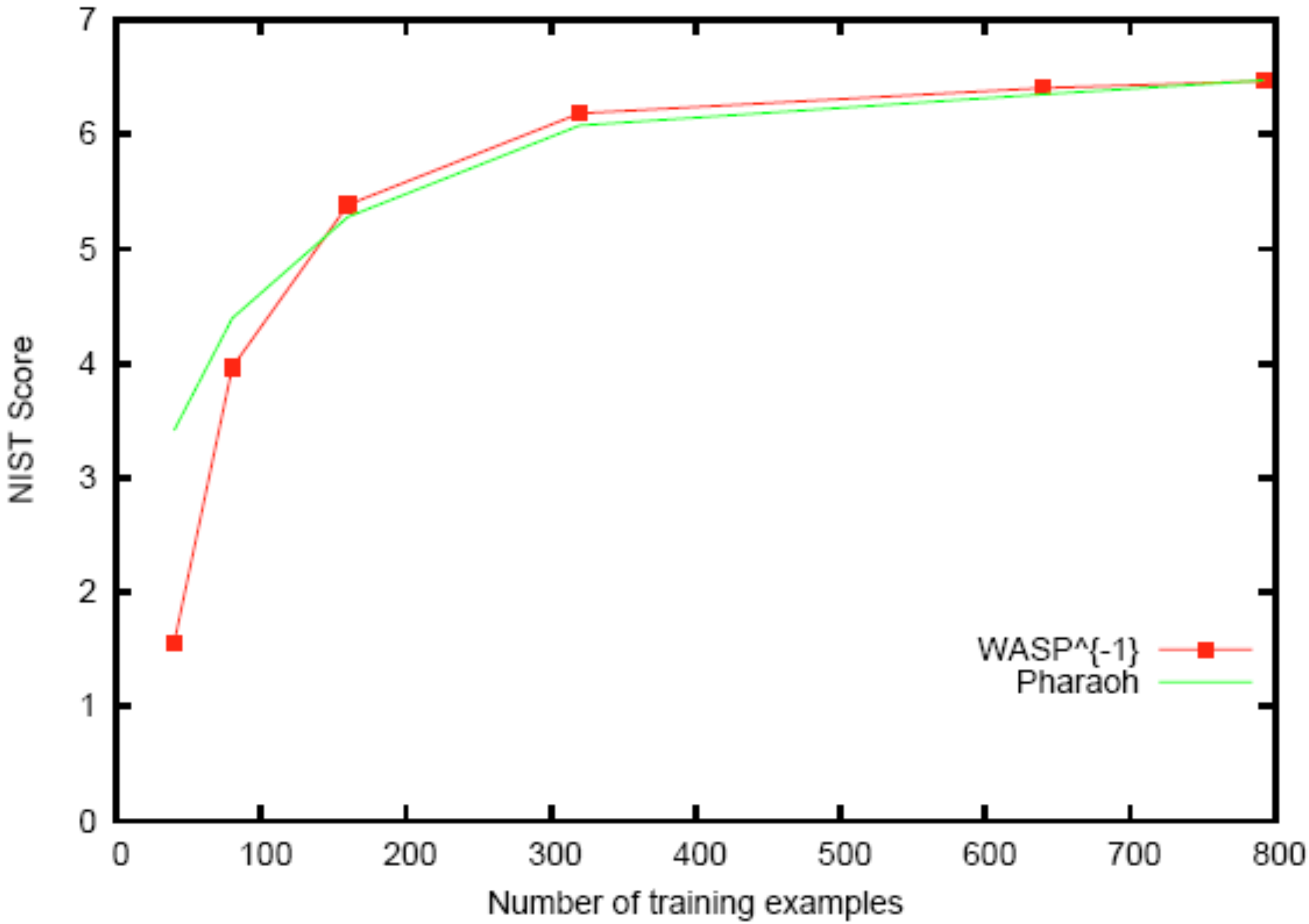
Output — Input

QUERY → **What is** CITY / answer(CITY)

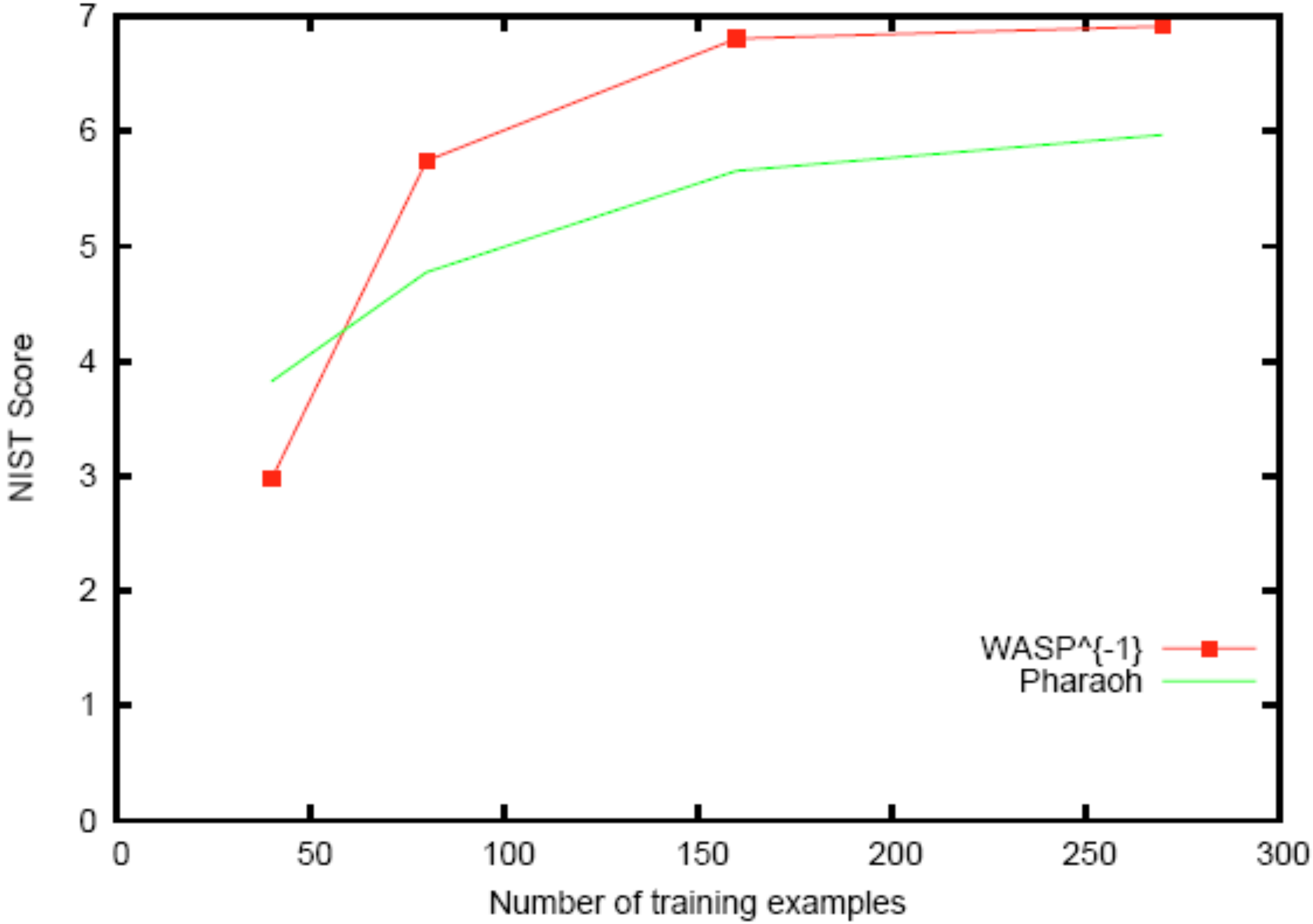
Generation by Inverting WASP

- Same procedure for lexical acquisition
- Chart generator is very similar to chart parser, but treats meaning representations as input
- Input can be logical forms with variables
- Log-linear probabilistic model inspired by Pharaoh (Koehn et al., 2003), a phrase-based MT system
  - Uses a bigram language model for target language
- Resulting system is called WASP^{-1}

NIST Scores for Geoquery



NIST Score for RoboCup



References

A. Aho, J. Ullman (1972). *The Theory of Parsing, Translation, and Compiling*. Prentice Hall.

P. Brown, V. Della Pietra, S. Della Pietra, R. Mercer (1993). The mathematics of statistical machine translation: Parameter estimation. *Comp. Ling.*, 19(2):263-312.

D. Chiang (2005). A hierarchical phrase-based model for statistical machine translation. In *Proc. of ACL*, pp. 263-270. Ann Arbor, MI.

P. Koehn, F. Och, D. Marcu (2003). Statistical phrase-based translation. In *Proc. of HLT-NAACL*. Edmonton, Canada.

Y. W. Wong, R. Mooney (2006). Learning for semantic parsing with statistical machine translation. In *Proc. of HLT-NAACL*, pp. 439-446. New York, NY.

References

Y. W. Wong, R. Mooney (2007a). Generation by inverting a semantic parser that uses statistical machine translation. In *Proc. of NAACL-HLT*, pp. 172-179. Rochester, NY.

Y. W. Wong, R. Mooney (2007b). Learning synchronous grammars for semantic parsing with lambda calculus. In *Proc. of ACL*, pp. 960-967. Prague, Czech Republic.

D. Wu (1997). Stochastic inversion transduction grammars and bilingual parsing of parallel corpora. *Comp. Ling.*, 23(3):377-403.

Semantic Parsing Using Kernels

KRISP: **K**ernel-based **R**obust **I**nterpretation for **S**emantic **P**arsing

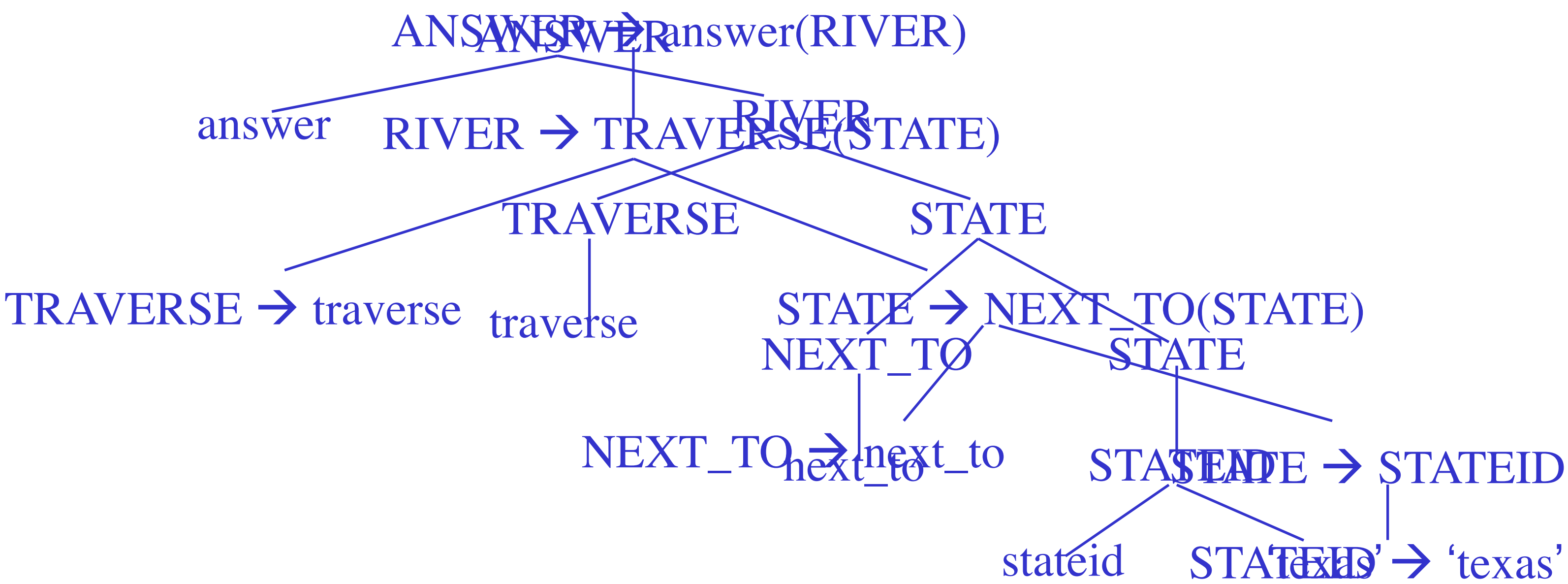
Kate & Mooney (2006), Kate (2008a)

- Learns semantic parser from NL sentences paired with their respective MRs given MRL grammar
- Productions of MRL are treated like semantic concepts
- A string classifier is trained for each production to estimate the probability of an NL string representing its semantic concept
- These classifiers are used to compositionally build MRs of the sentences

Meaning Representation Language

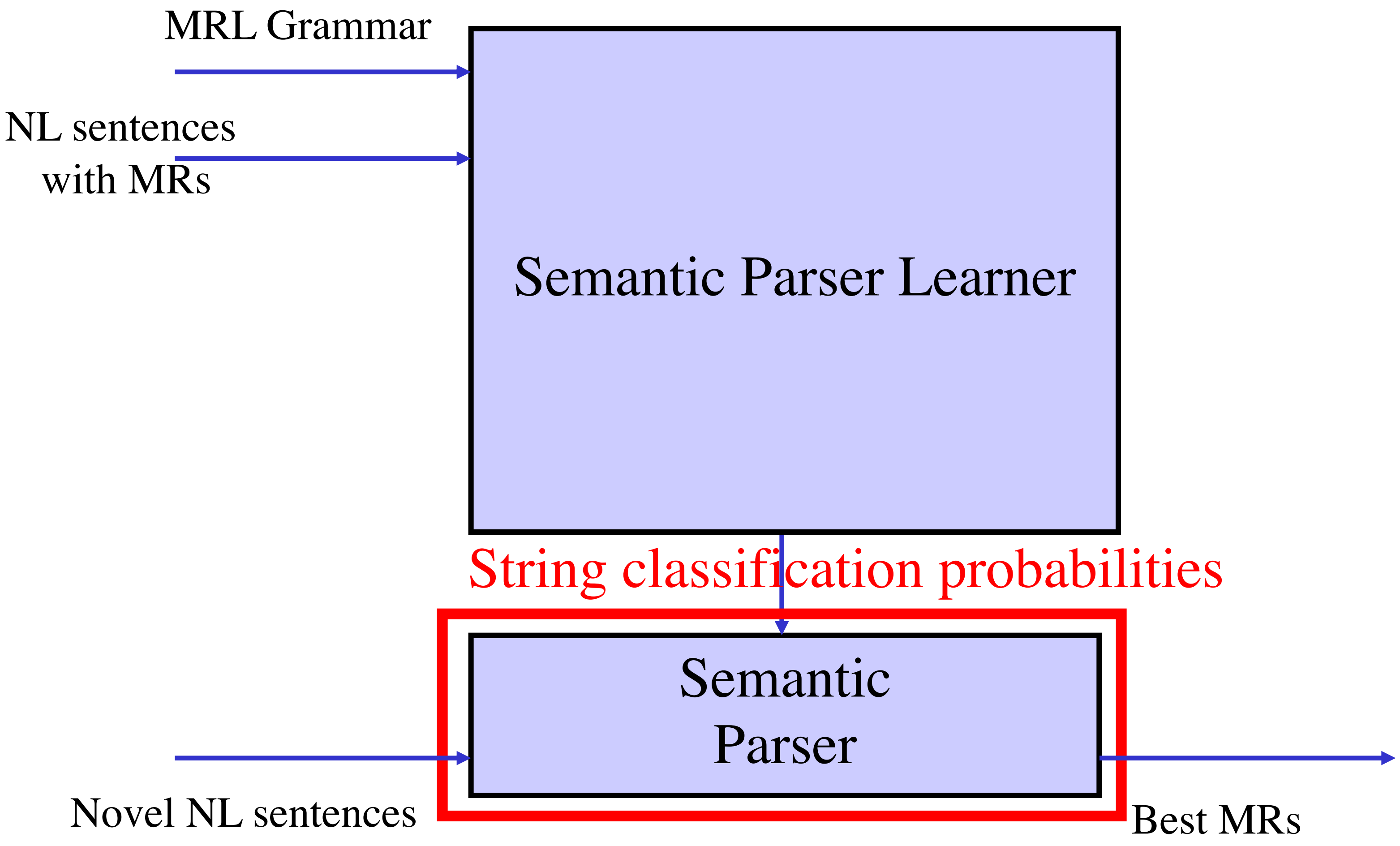
MR: answer(traverse(next_to(stateid('texas'))))

Parse tree of MR:



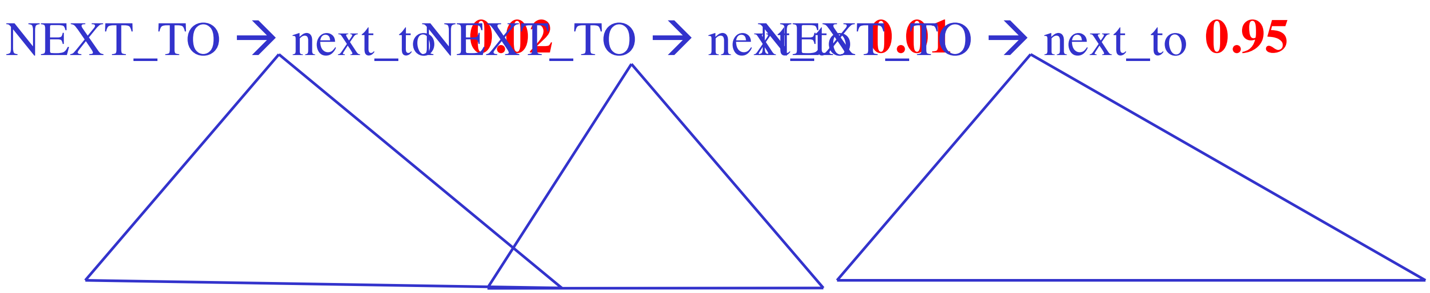
Productions: ANSWER → answer(RIVER)　　RIVER → TRAVERSE(STATE)
STATE → NEXT_TO(STATE)　　TRAVERSE → traverse
NEXT_TO → next_to　　STATEID → 'texas'

Overview of KRISP



Semantic Parsing by KRISP

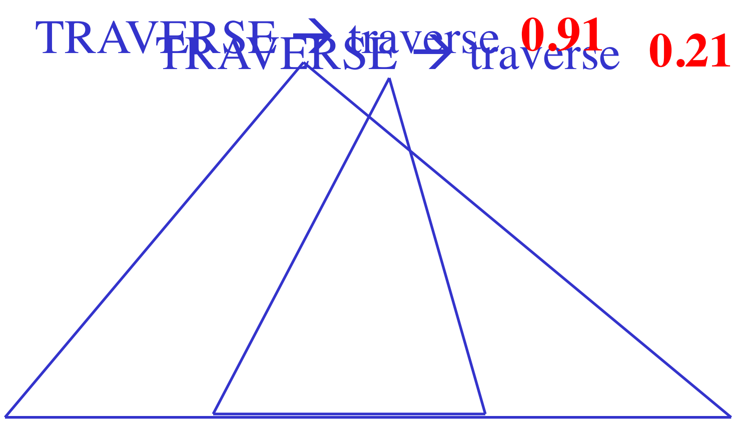
- String classifier for each production gives the probability that a substring represents the semantic concept of the production



Which rivers run through the states bordering Texas?

Semantic Parsing by KRISP

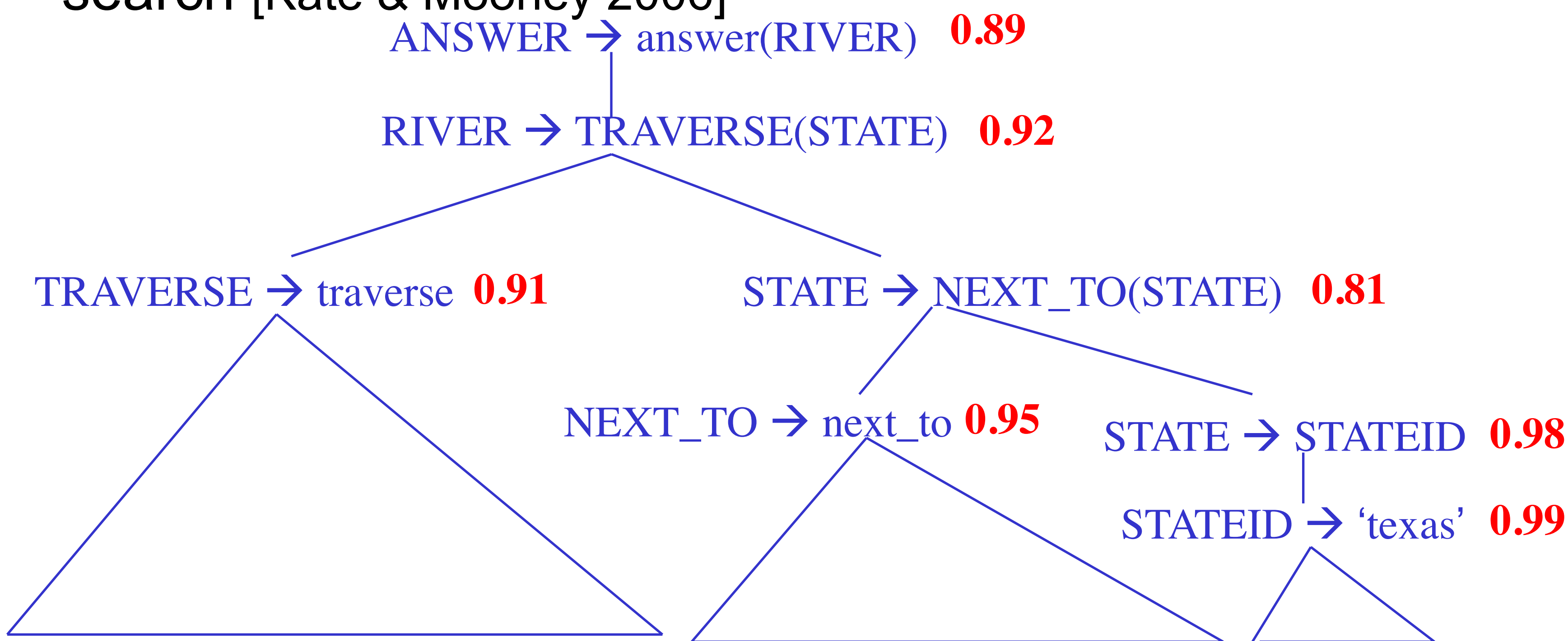
- String classifier for each production gives the probability that a substring represents the semantic concept of the production



Which rivers run through the states bordering Texas?

Semantic Parsing by KRISP

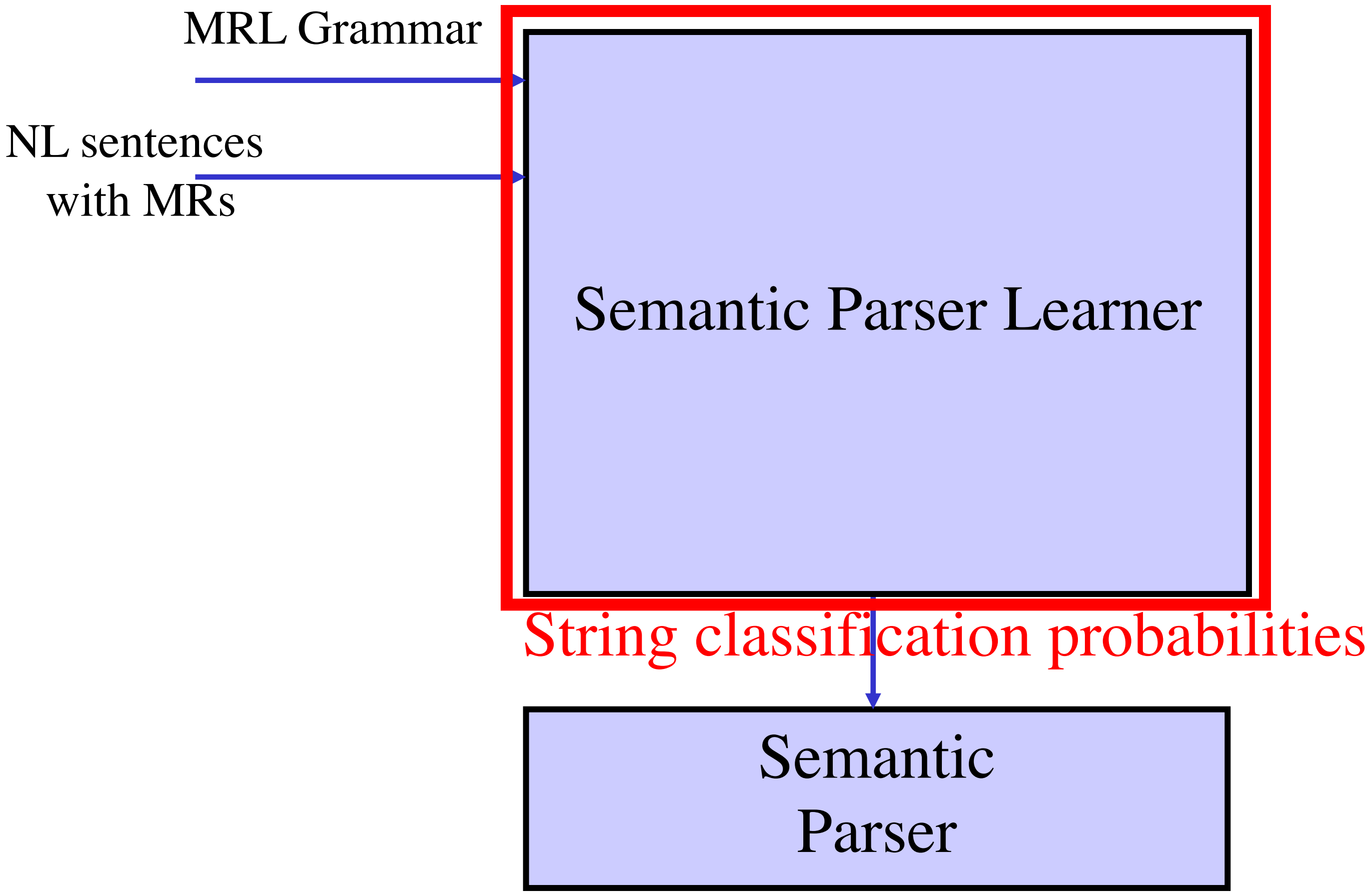
- Semantic parsing reduces to finding the *most probable derivation* of the sentence
- Efficient dymamic programming algorithm with beam search [Kate & Mooney 2006]



Which rivers run through the states bordering Texas?

Probability of the derivation is the product of the probabilities at the nodes.

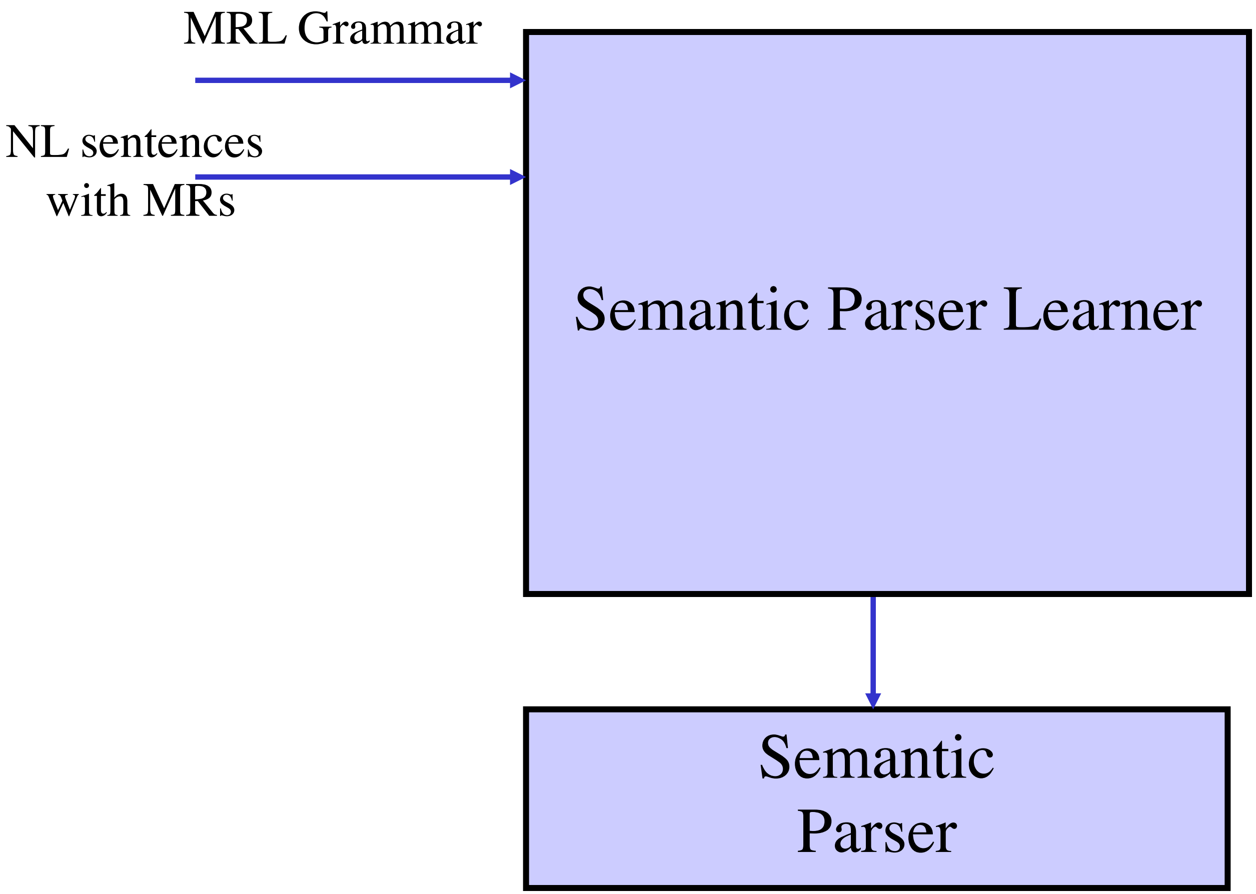
Overview of KRISP



KRISP's Training Algorithm

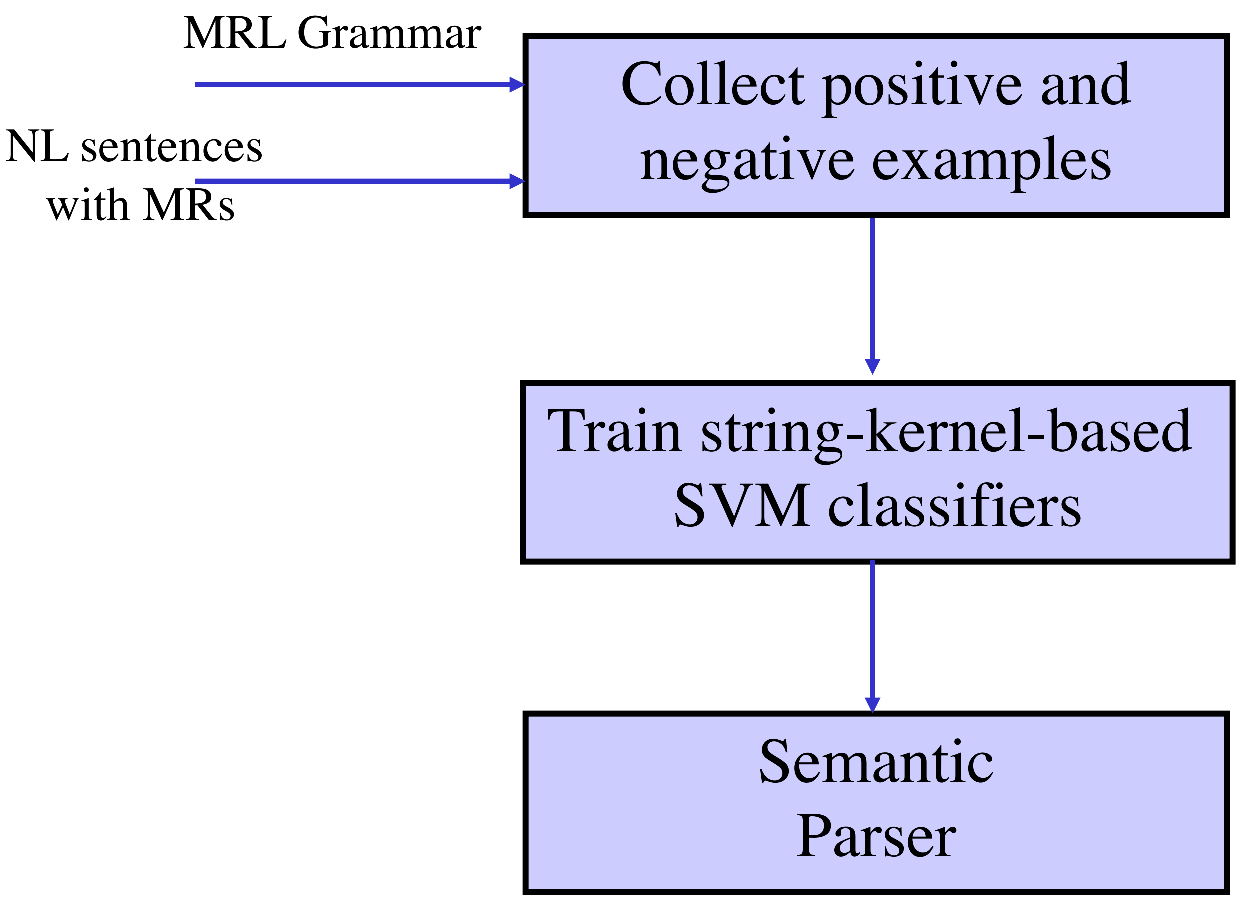
- Takes NL sentences paired with their respective MRs as input
- Obtains MR parses using MRL grammar
- Induces the semantic parser and refines it in iterations
- In the first iteration, for every production:
  - Call those sentences ***positives*** whose MR parses use that production
  - Call the remaining sentences ***negatives***
  - Trains *Support Vector Machine* (SVM) classifier [Cristianini 2000 & Shawe-Taylor] using string-subsequence kernel

Overview of KRISP



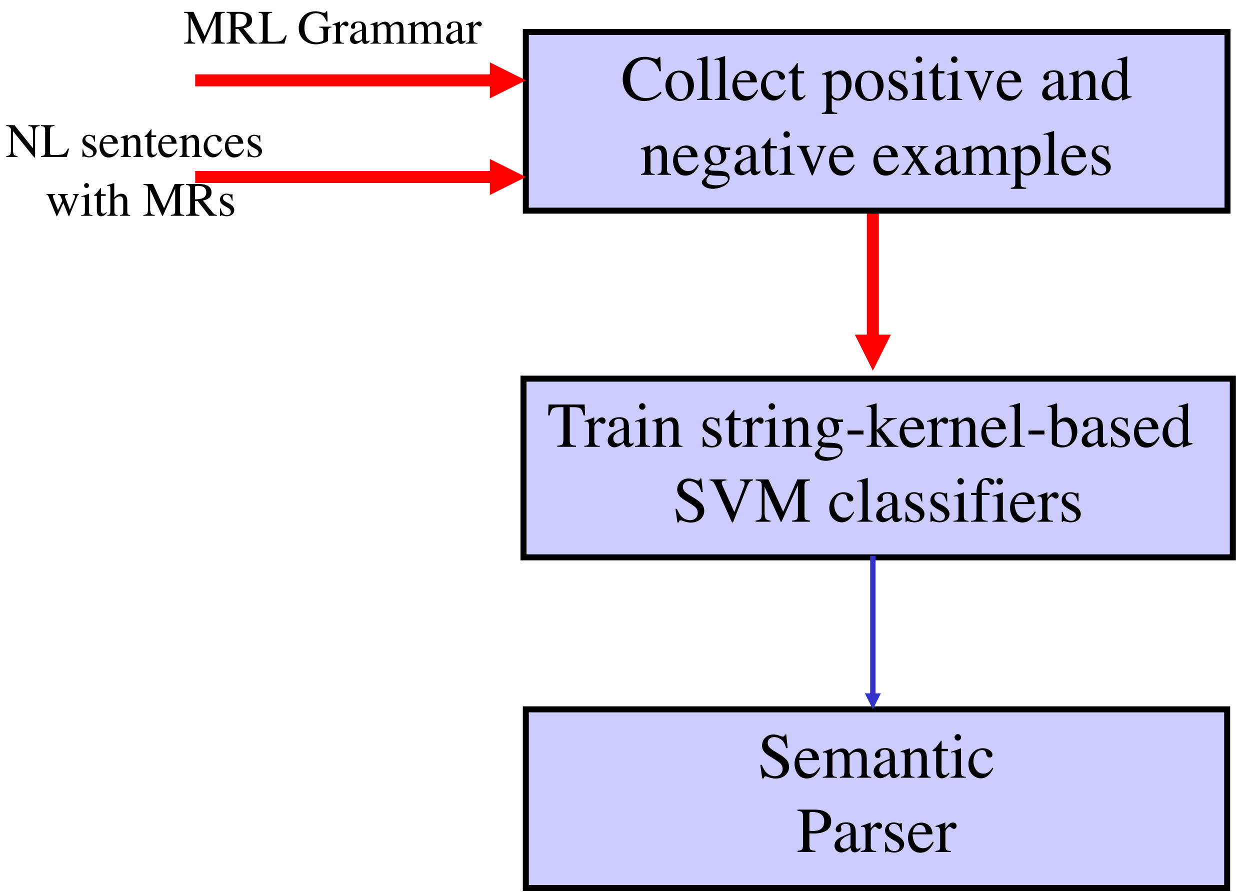
Overview of KRISP

Training



Overview of KRISP

Training



KRISP's Training Algorithm contd.

First Iteration

STATE → NEXT_TO(STATE)

Positives

- which rivers run through the states bordering texas?
- what is the most populated state bordering oklahoma ?
- what is the largest city in states that border california ?

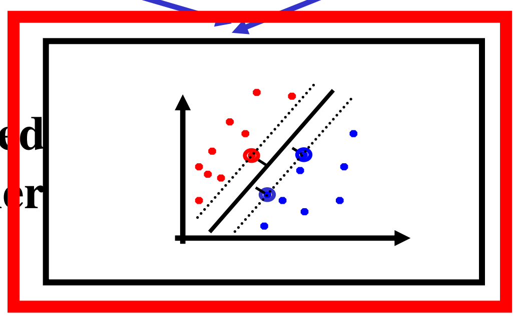
…

Negatives

- what state has the highest population ?
- what states does the delaware river run through ?
- which states have cities named austin ?
- what is the lowest point of the state with the largest area ?

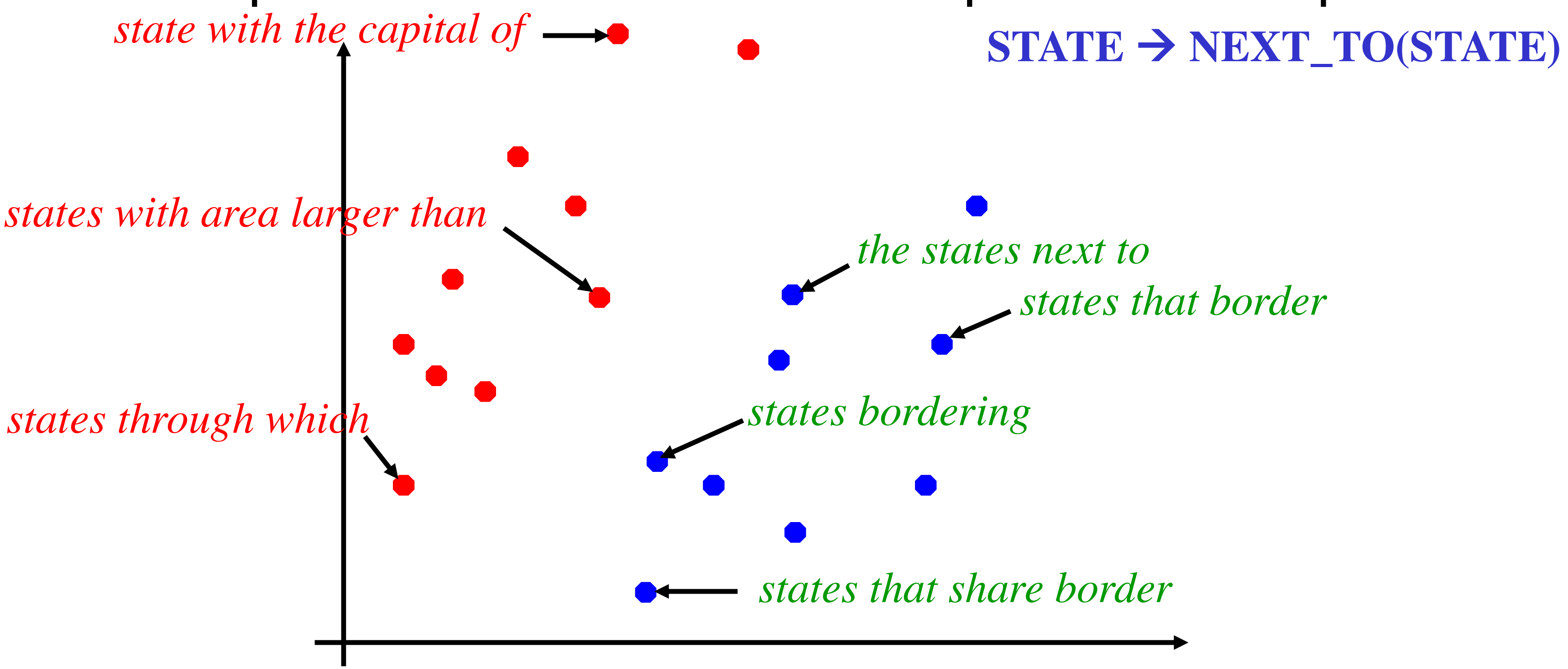
…

String-kernel-based SVM classifier



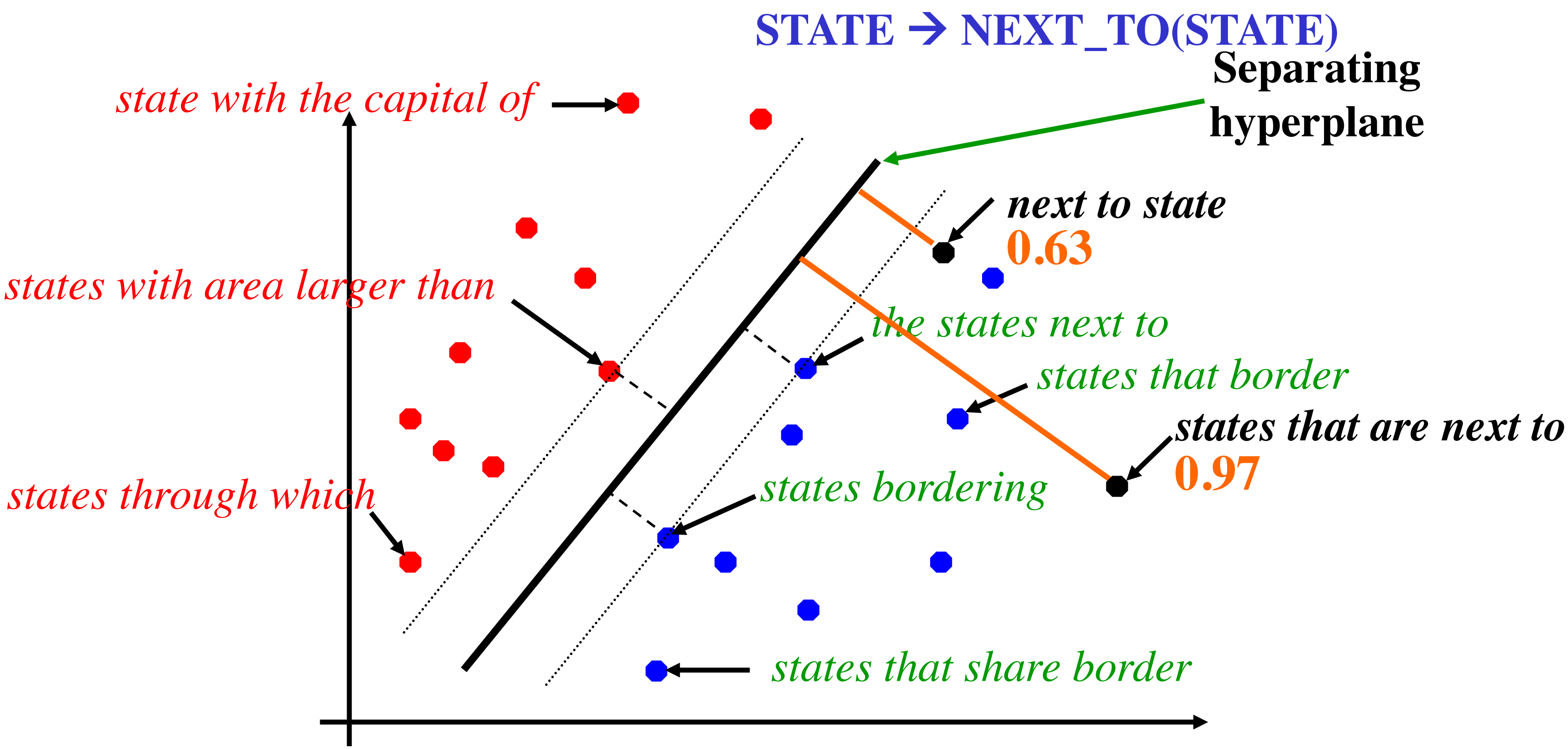
String Subsequence Kernel

- Define kernel between two strings as the number of common subsequences between them [Lodhi et al., 2002]
- The examples are implicitly mapped to the feature space of all subsequences and the kernel computes the dot products



Support Vector Machines

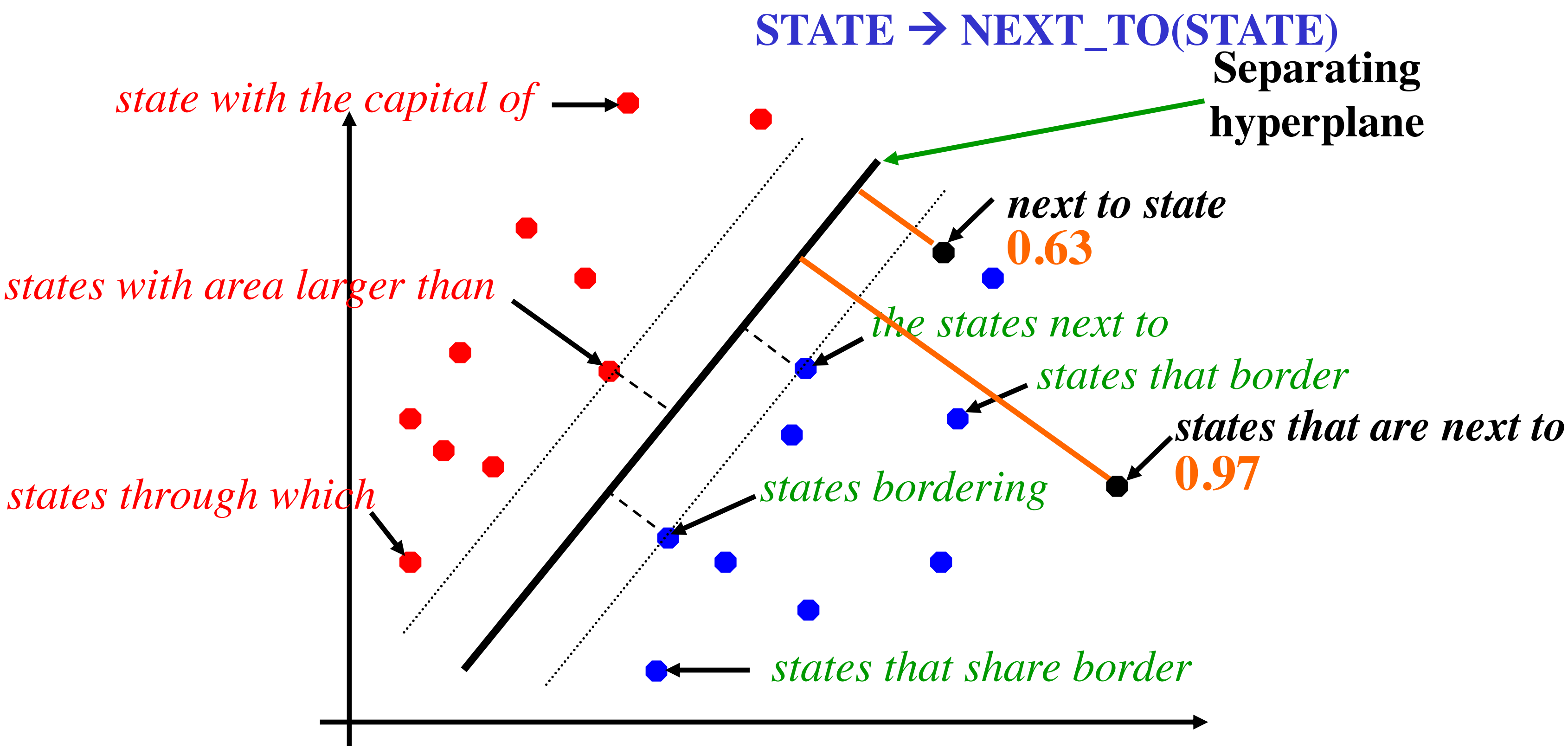
- SVMs find a separating hyperplane such that the margin is maximized



Probability estimate of an example belonging to a class can be obtained using its distance from the hyperplane [Platt, 1999]

Support Vector Machines

- SVMs find a separating hyperplane such that the margin is maximized



SVMs with string subsequence kernel softly capture different ways of expressing the semantic concept.

KRISP's Training Algorithm contd.

First Iteration

STATE → NEXT_TO(STATE)

Positives

- which rivers run through the states bordering texas?
- what is the most populated state bordering oklahoma ?
- what is the largest city in states that border california ?

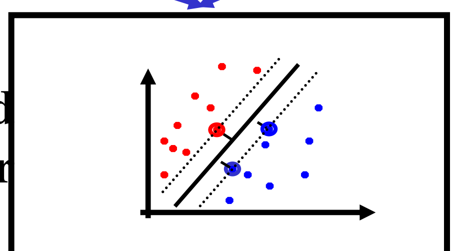
…

Negatives

- what state has the highest population ?
- what states does the delaware river run through ?
- which states have cities named austin ?
- what is the lowest point of the state with the largest area ?

…

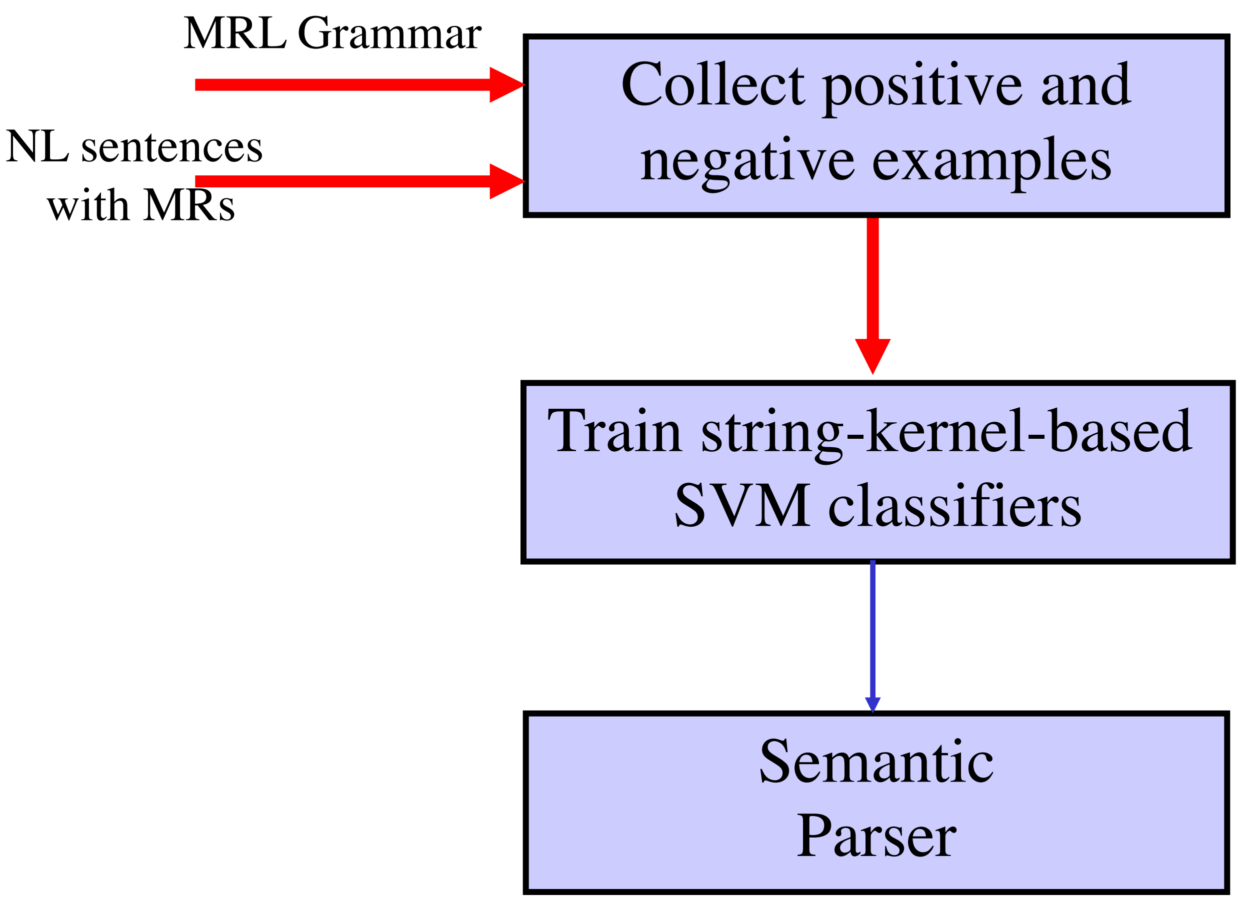
String-kernel-based SVM classifier



String classification probabilities

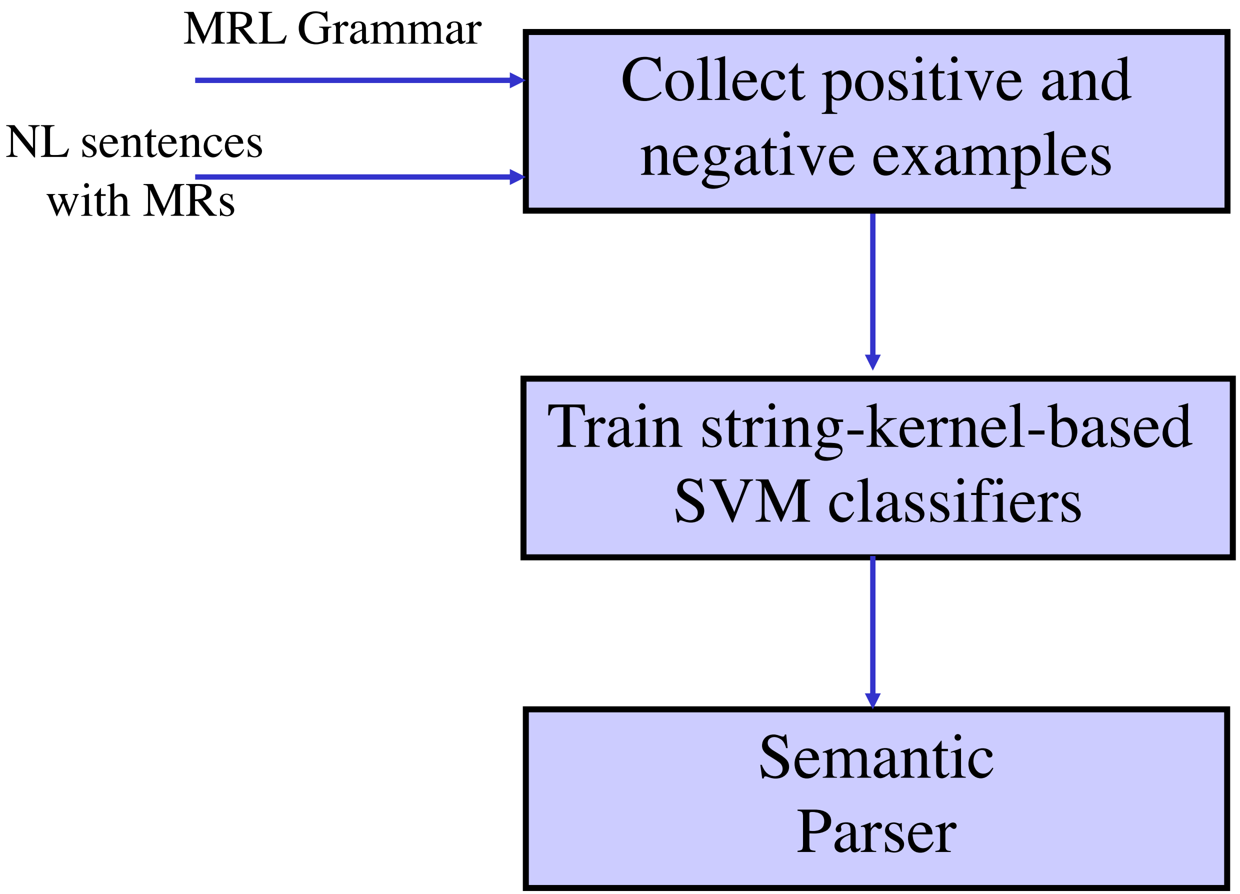
Overview of KRISP

Training



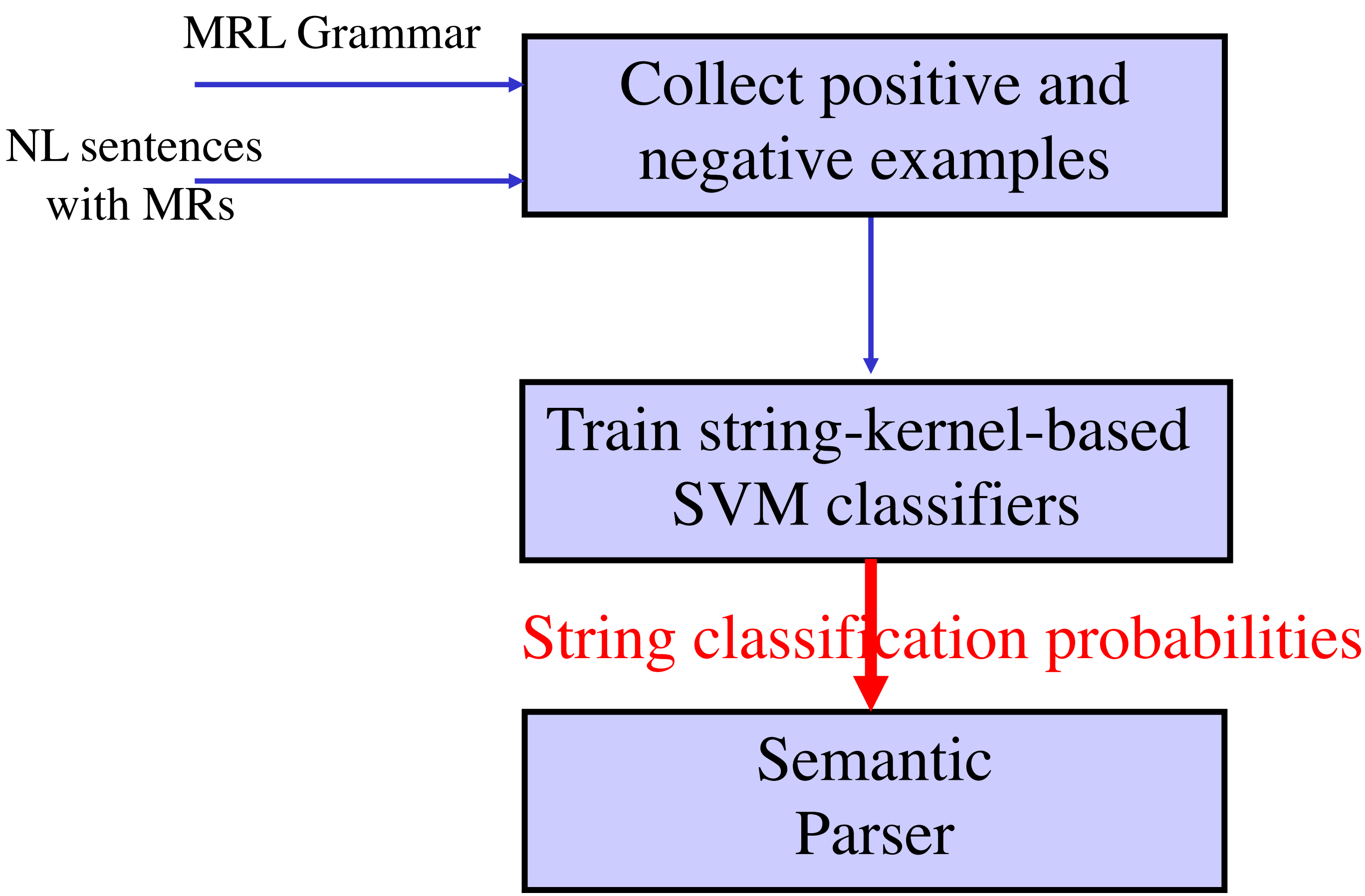
Overview of KRISP

Training



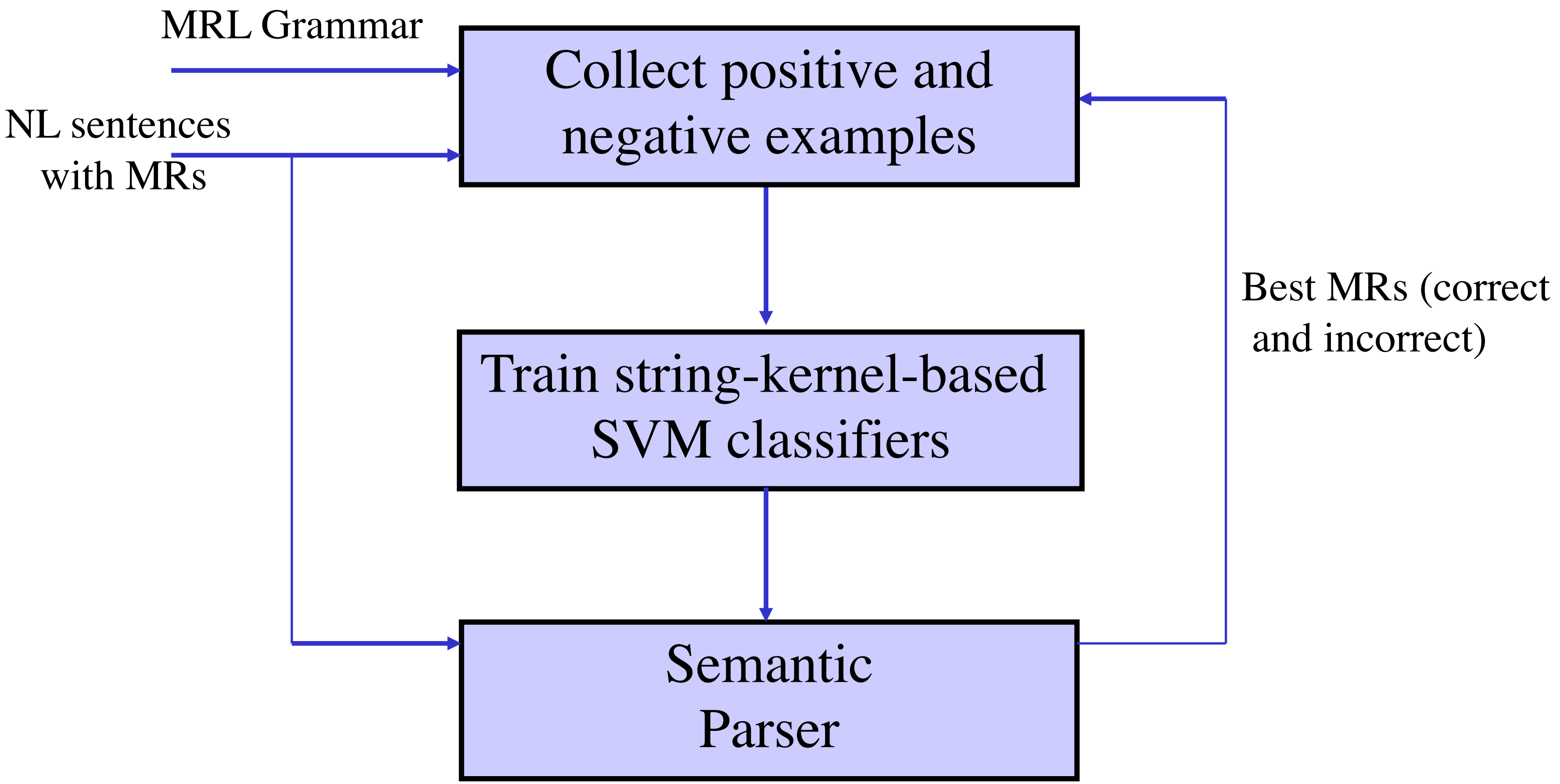
Overview of KRISP

Training



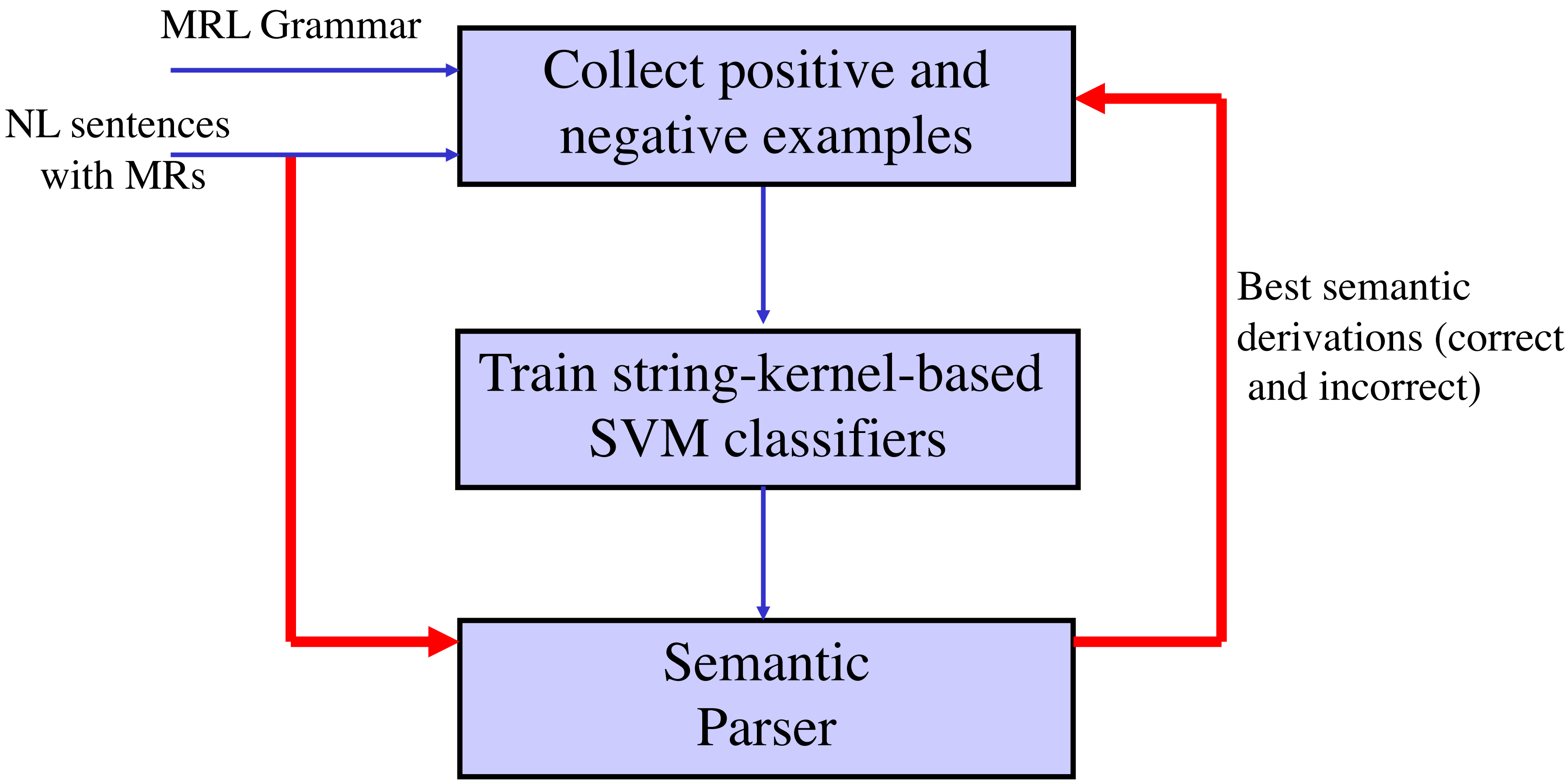
Overview of KRISP

Training



Overview of KRISP

Training

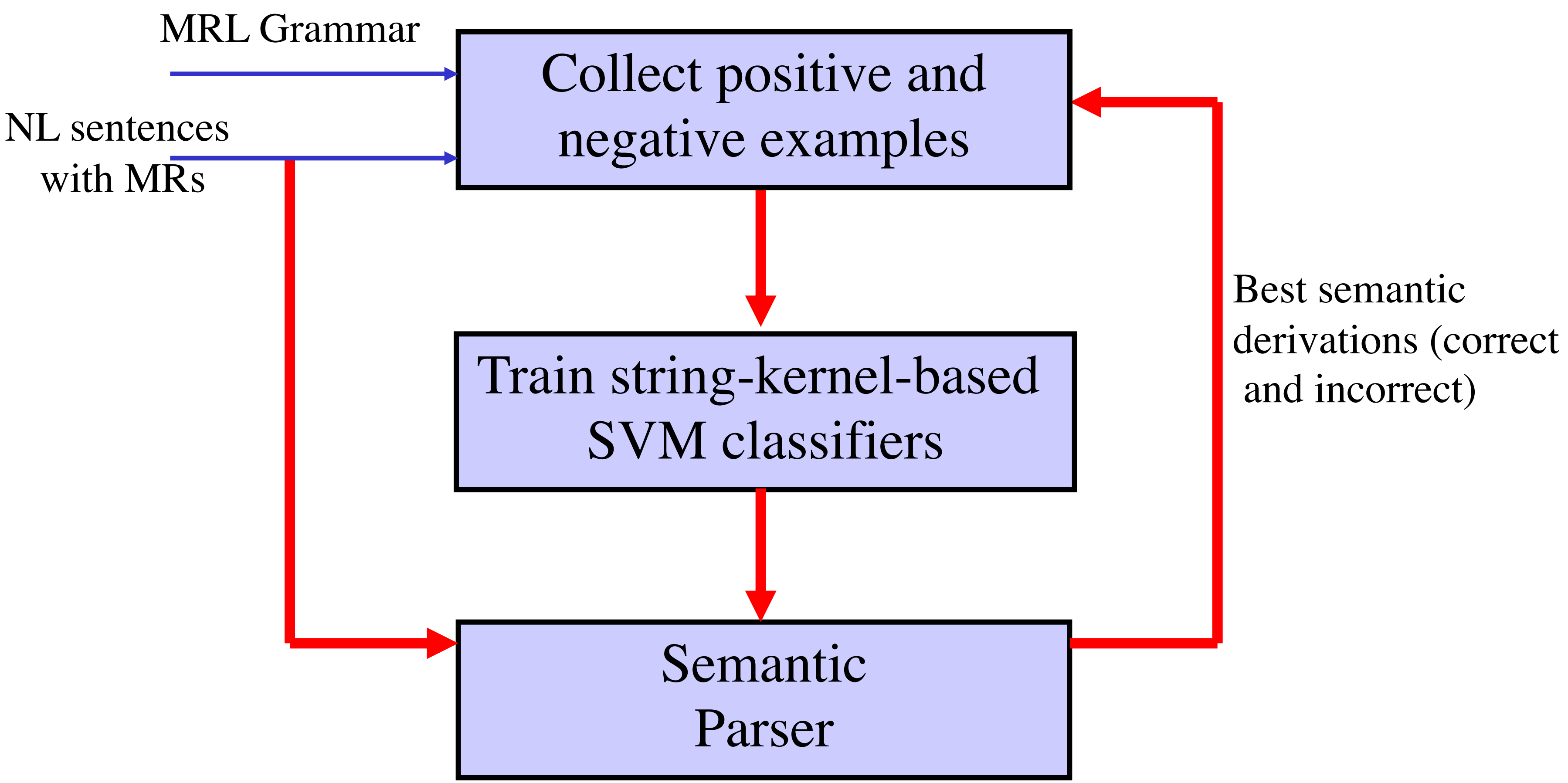


KRISP's Training Algorithm contd.

- Using these classifiers, it tries to parse the sentences in the training data
- Some of these derivations will give the correct MR, called ***correct derivations***, some will give incorrect MRs, called ***incorrect derivations***
- For the next iteration, collect positive examples from correct derivations and negative examples from incorrect derivations

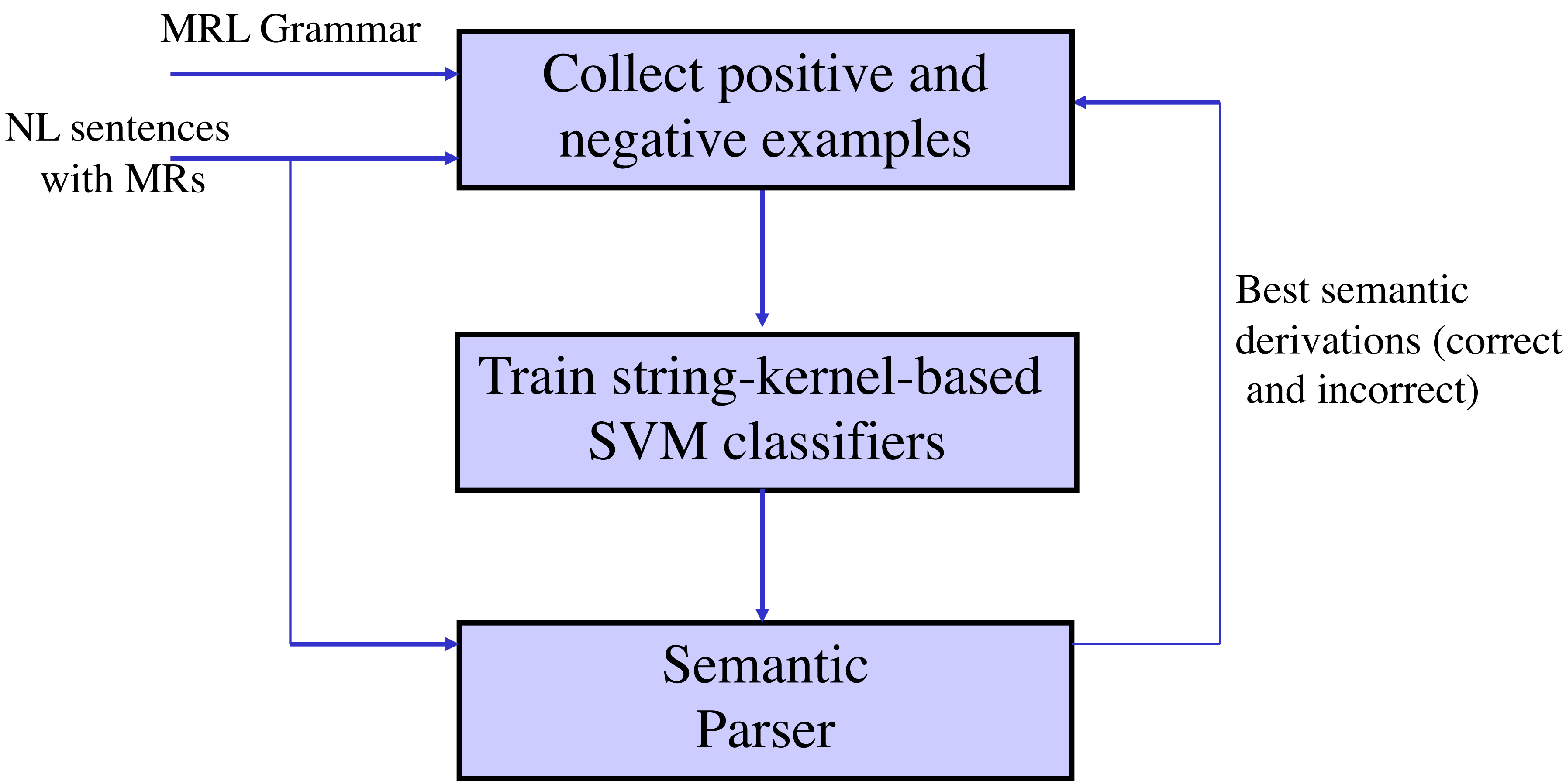
Overview of KRISP

Training

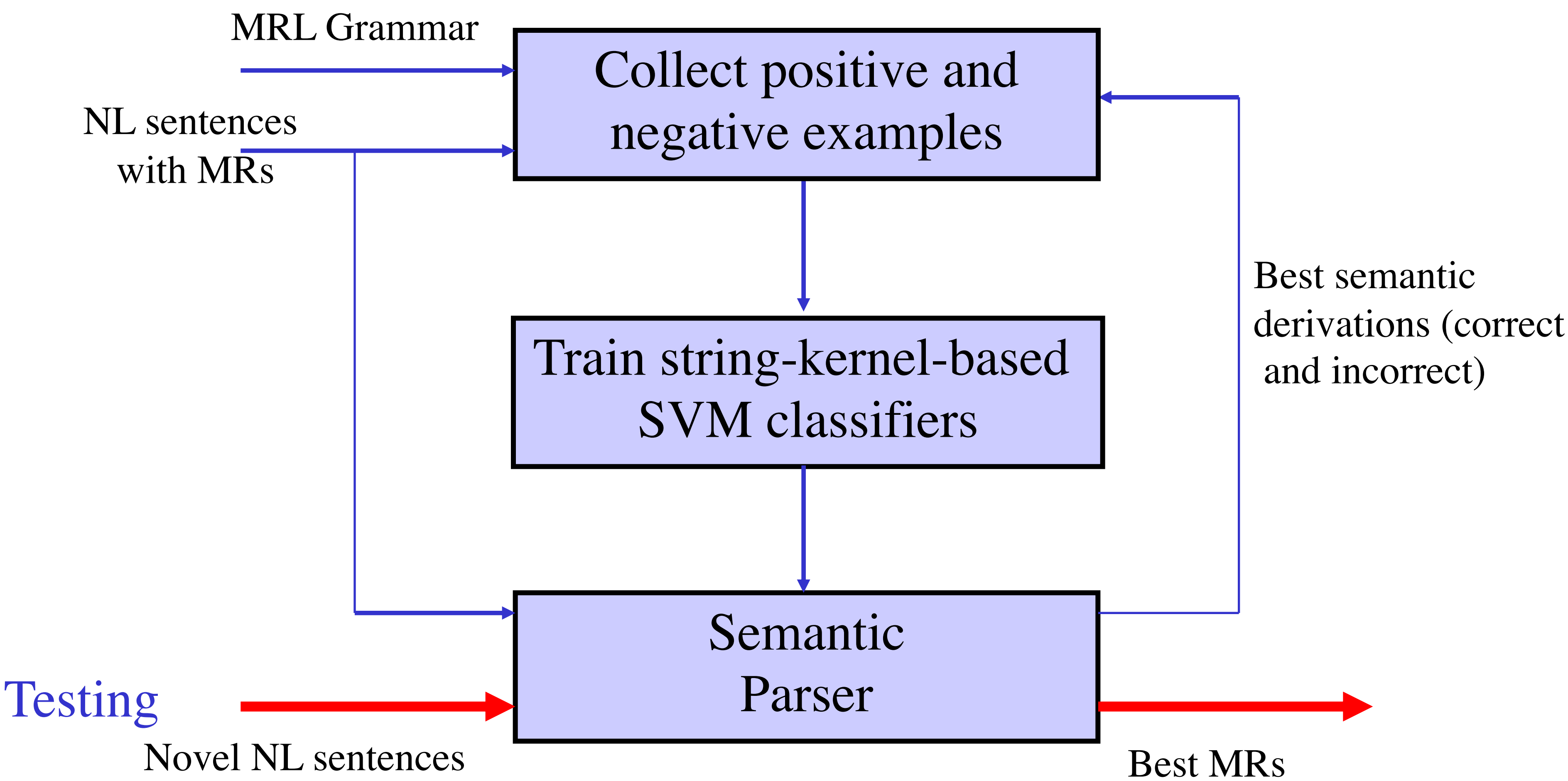


Overview of KRISP

Training



Overview of KRISP



A Dependency-based Word Subsequence Kernel

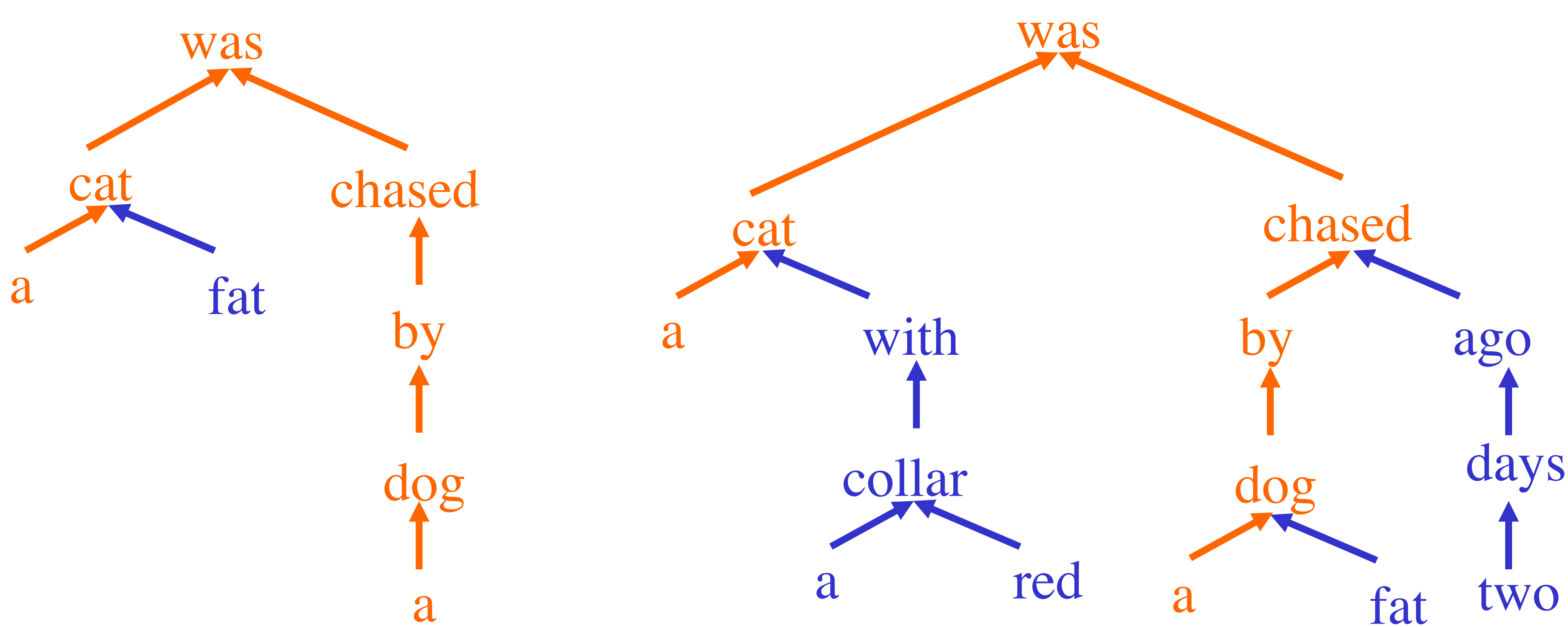
Kate (2008a)

- A word subsequence kernel can count linguistically meaningless subsequences
  - *A fat cat was chased by a dog.*
  - *A cat with a red collar was chased two days ago by a fat dog*
- A new kernel that counts only the linguistically meaningful subsequences

A Dependency-based Word Subsequence Kernel

Kate (2008a)

- Count the number of common paths in the dependency trees; efficient algorithm to do it



- Outperforms word subsequence kernel on semantic parsing

References

Huma Lodhi, Craig Saunders, John Shawe-Taylor, Nello Cristianini, and Chris Watkins (2002). Text classification using string kernels. *Journal of Machine Learning Research*, 2:419--444.

John C. Platt (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in Large Margin Classifiers*, pages 185--208. MIT Press.

Rohit J. Kate and Raymond J. Mooney (2006). Using string-kernels for learning semantic parsers. In *Proc. of COLING/ACL-2006*, pp. 913-920, Sydney, Australia, July 2006.

Rohit J. Kate (2008a). A dependency-based word subsequence kernel. In *Proc. of EMNLP-2008*, pp. 400-409, Waikiki, Honolulu, Hawaii, October 2008.

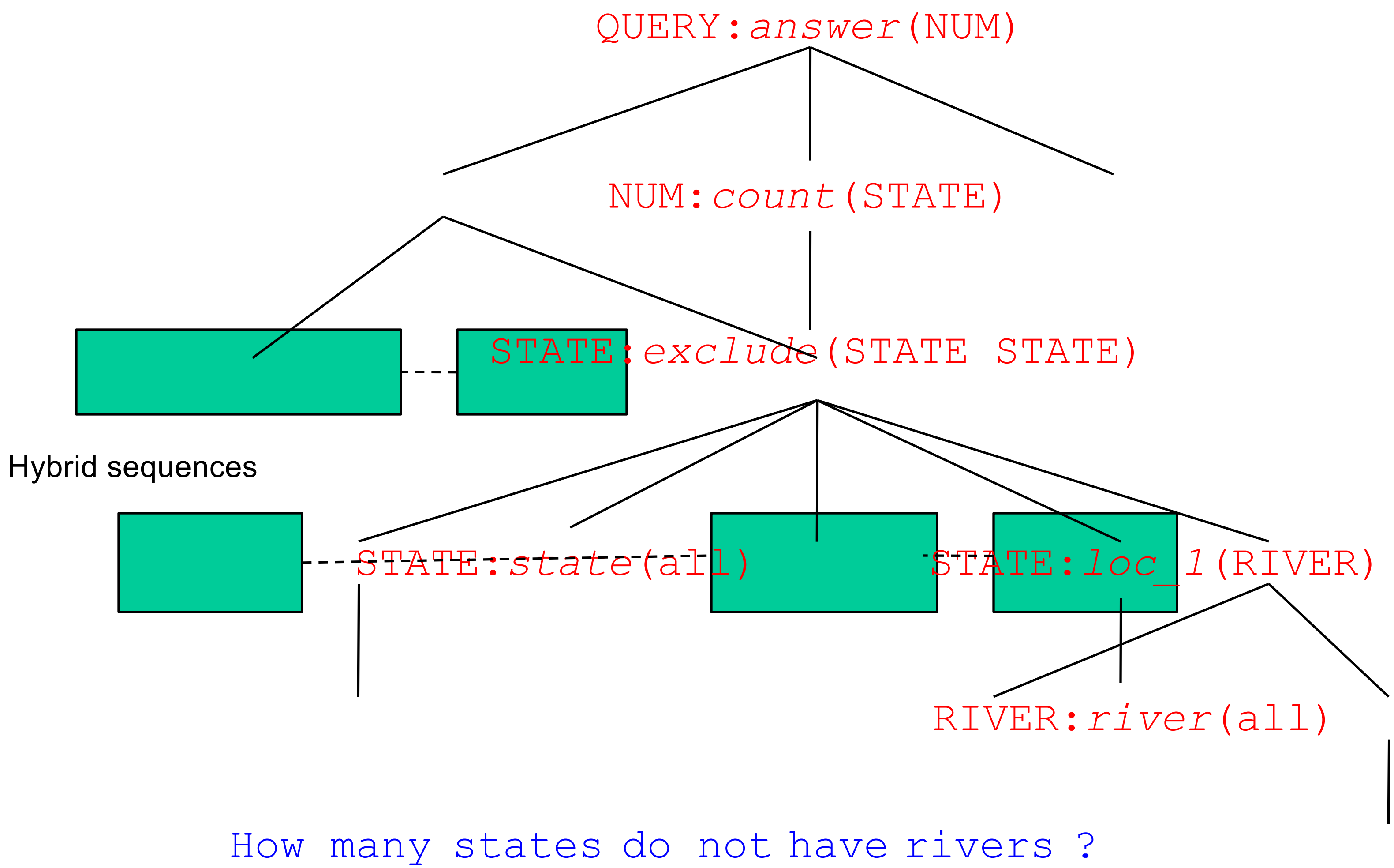
Rohit J. Kate (2008b). Transforming meaning representation grammars to improve semantic parsing. In *Proc. Of CoNLL-2008*, pp. 33-40, Manchester, UK, August 2008.

A Generative Model for Semantic Parsing

Hybrid Tree

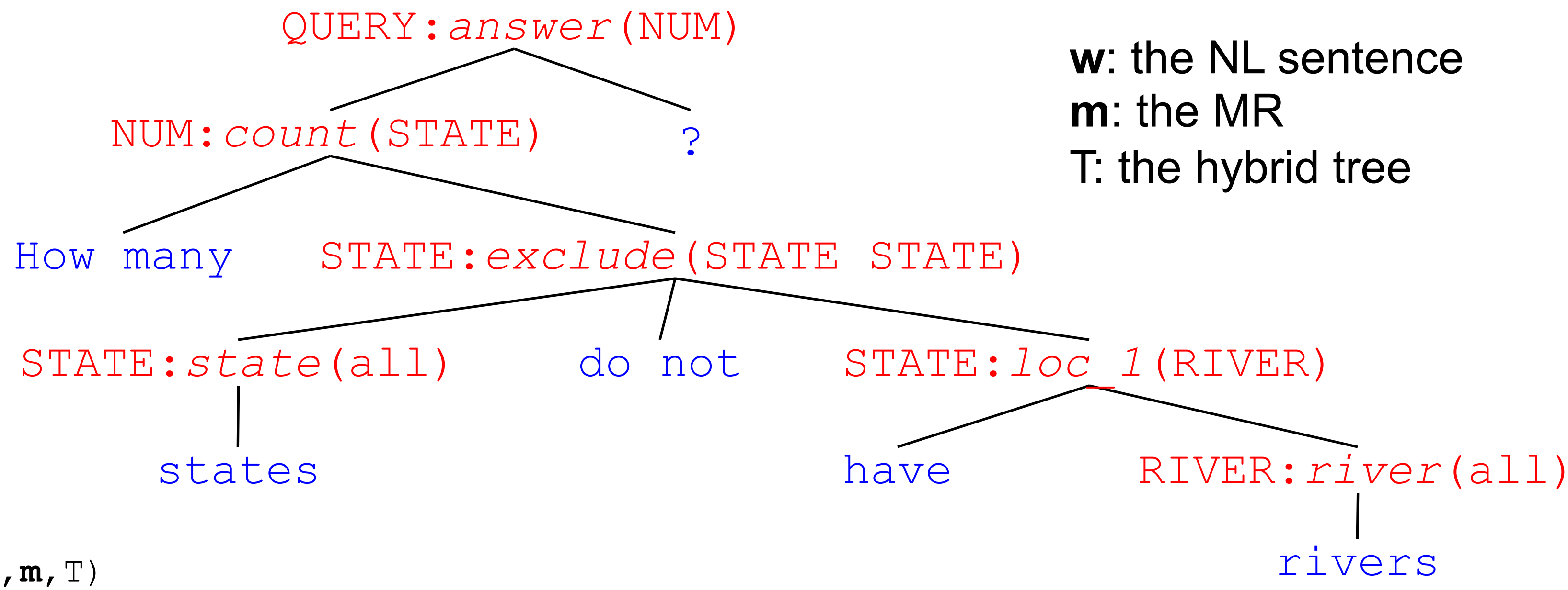
Lu et al. (2008)

NL-MR Pair



Model Parameters

w: the NL sentence
m: the MR
T: the hybrid tree

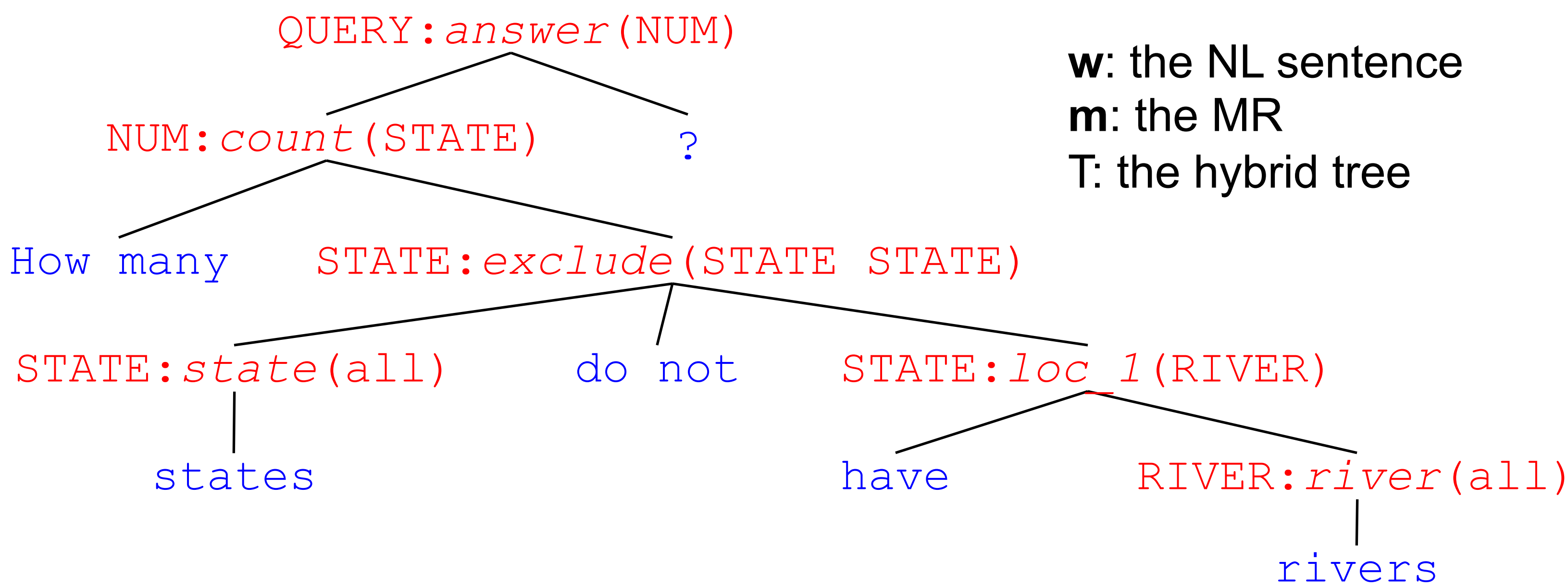


P(**w**,**m**,T)
=P(QUERY:*answer*(NUM)|-,arg=1)
*P(NUM ?|QUERY:*answer*(NUM))
*P(NUM:*count*(STATE)|QUERY:*answer*(NUM),arg=1)
*P(How many STATE|NUM:*count*(STATE))
*P(STATE:*exclude*(STATE STATE)|NUM:*count*(STATE),arg=1)
*P(STATE$_1$ do not STATE$_2$|STATE:*exclude*(STATE STATE))
*P(STATE:*state*(all)|STATE:*exclude*(STATE STATE),arg=1)
*P(states|STATE:*state*(all))
*P(STATE:*loc_1*(RIVER)|STATE:*exclude*(STATE STATE),arg=2)
*P(have RIVER|STATE:*loc_1*(RIVER))
*P(RIVER:*river*(all)|STATE:*loc_1*(RIVER),arg=1)
*P(rivers|RIVER:*river*(all))

MR Model Parameters

ρ(m'|m,arg=k)

Model Parameters



w: the NL sentence
m: the MR
T: the hybrid tree

P(How many STATE|NUM:count(STATE))
= P(m→wY|NUM:count(STATE))
* P(How|NUM:count(STATE),BEGIN)
* P(many|NUM:count(STATE),How)
* P(STATE|NUM:count(STATE),many)
* P(END|NUM:count(STATE),STATE)

Pattern Parameters

$\Phi(r|m)$

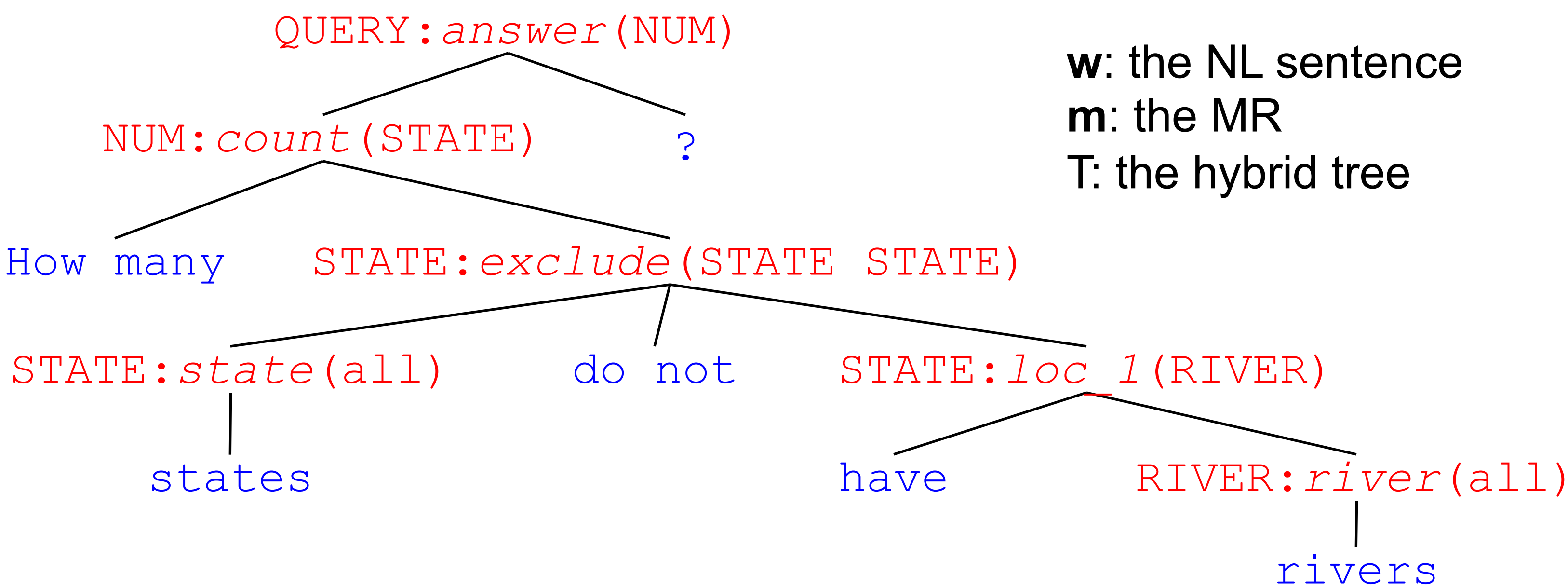
Hybrid Patterns

#RHS	Hybrid Pattern	# Patterns
0	M → w	1
1	M → [w] Y [w]	4
2	M → [w] Y [w] Z [w]	8
	M → [w] Z [w] Y [w]	8

- M is an MR production, w is a word sequence
- Y and Z are respectively the first and second child MR production

Note: [] denotes optional

Model Parameters



w: the NL sentence
m: the MR
T: the hybrid tree

P(How many STATE|NUM:*count*(STATE))
= P(m→wY|NUM:*count*(STATE))
* P(How|NUM:*count*(STATE),BEGIN)
* P(many|NUM:*count*(STATE),How)
* P(STATE|NUM:*count*(STATE),many)
* P(END|NUM:*count*(STATE),STATE)

Emission Parameters

θ(t|m,Λ)

Model Parameters

- MR model parameters

  $\Sigma_{m_i} \rho(m_i|m_j,arg=k) = 1$

  They model the meaning representation

- Emission parameters

  $\Sigma_t \Theta(t|m_j,\Lambda) = 1$

  They model the emission of words and semantic categories of MR productions. Λ is the context.

- Pattern parameters

  $\Sigma_r \Phi(r|m_j) = 1$

  They model the selection of hybrid patterns

Parameter Estimation

- MR model parameters are easy to estimate
- Learning the emission parameters and pattern parameters is challenging
- Inside-outside algorithm with EM
  - Naïve implementation: O(n^6m)
  - n: number of words in an NL sentence
  - m: number of MR productions in an MR
- Improved efficient algorithm
  - Two-layer dynamic programming
  - Improved time complexity: O(n^3m)

Reranking

- Weakness of the generative model
  - Lacks the ability to model long range dependencies
- Reranking with the averaged perceptron
  - Output space
    - Hybrid trees from exact top-k (k=50) decoding algorithm for each training/testing instance's NL sentence
  - Single correct reference
    - Output of Viterbi algorithm for each training instance
  - Long-range features

Reference

Wei Lu, Hwee Tou Ng, Wee Sun Lee and Luke S. Zettlemoyer (2008). A generative model for parsing natural language to meaning representations. In *Proc. of EMNLP-2008*, Waikiki, Honolulu, Hawaii, October 2008.

Outline

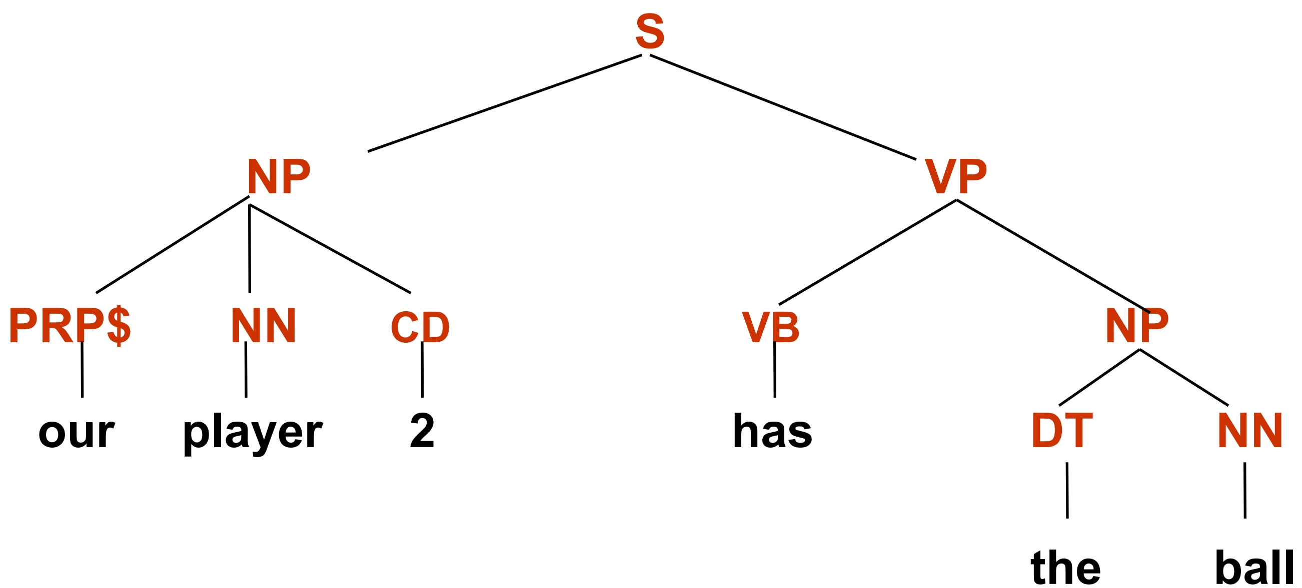
1. Introduction to the task of semantic parsing
   a) Definition of the task
   b) Examples of application domains and meaning representation languages
   c) Distinctions from and relations to other NLP tasks
2. Semantic parsers
   a) Earlier hand-built systems
   b) Learning for semantic parsing
      I. Semantic parsing learning task
      II. Early semantic parser learners
      III. Recent semantic parser learners
      IV. Exploiting syntax for semantic parsing
      V. Underlying commonalities and differences between semantic parsers
   c) Various forms of supervision
3. Semantic parsing beyond a sentence
   a) Learning language from perceptual contexts
   b) Using discourse contexts for semantic parsing
4. Research challenges and future directions
   a) Machine reading of documents: Connecting with knowledge representation
   b) Applying semantic parsing techniques to the Semantic Web
   c) Future research directions
5. Conclusions

Exploiting Syntax for Semantic Parsing

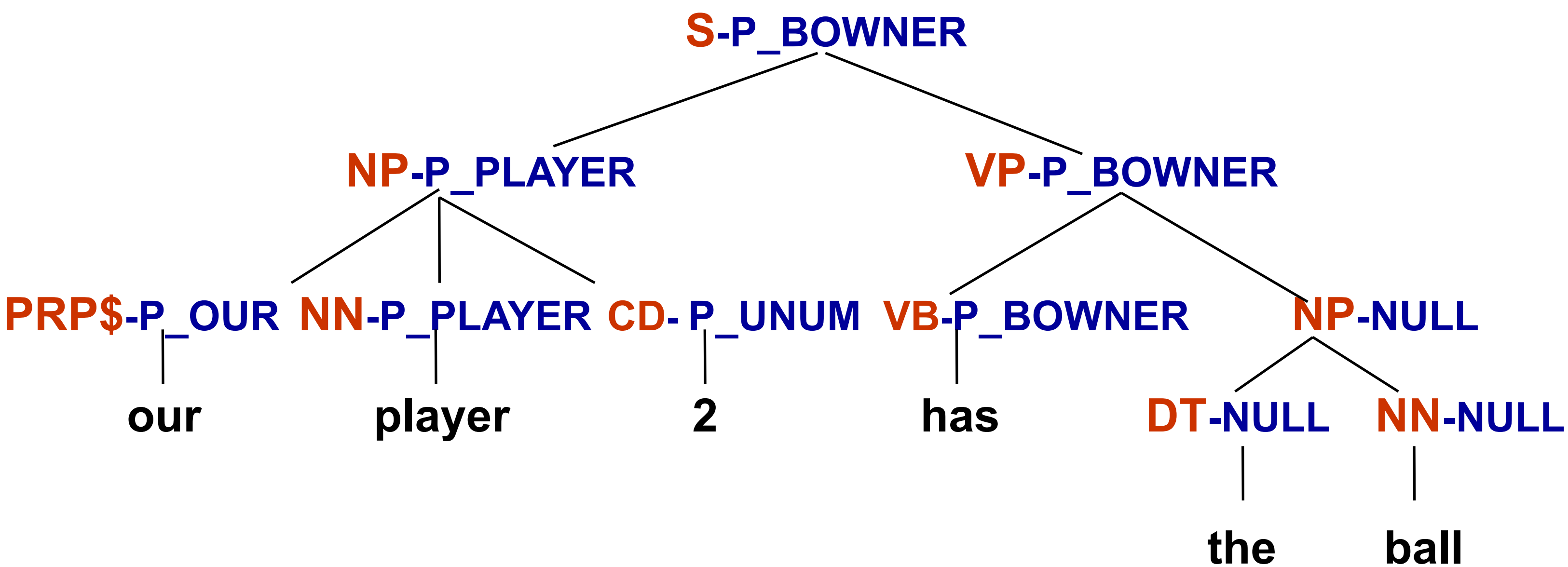
SCISSOR Ge & Mooney (2005)

- ***S***emantic ***C***omposition that ***I***ntegrates ***S***yntax and ***S***emantics to get ***O***ptimal ***R***epresentations
- **Integrated syntactic-semantic parsing**
  - Allows both **syntax and semantics** to be used **simultaneously** to obtain an accurate combined syntactic-semantic analysis
- A statistical parser is used to generate a **semantically augmented parse tree (SAPT)**

Syntactic Parse

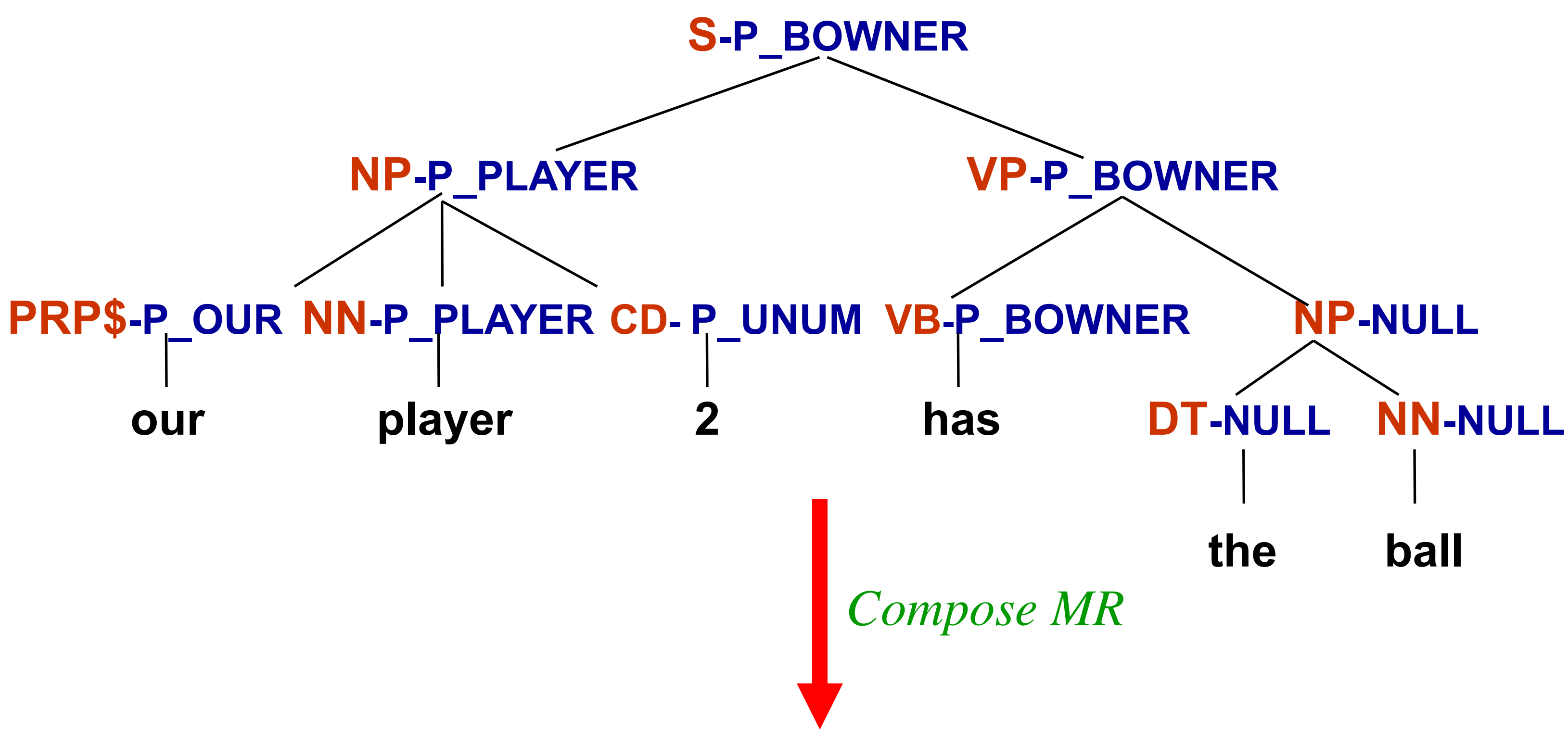


SAPT



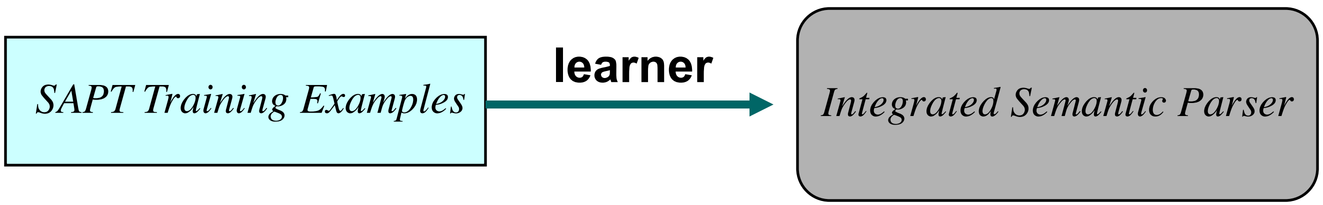
Non-terminals now have both syntactic and semantic labels

SAPT



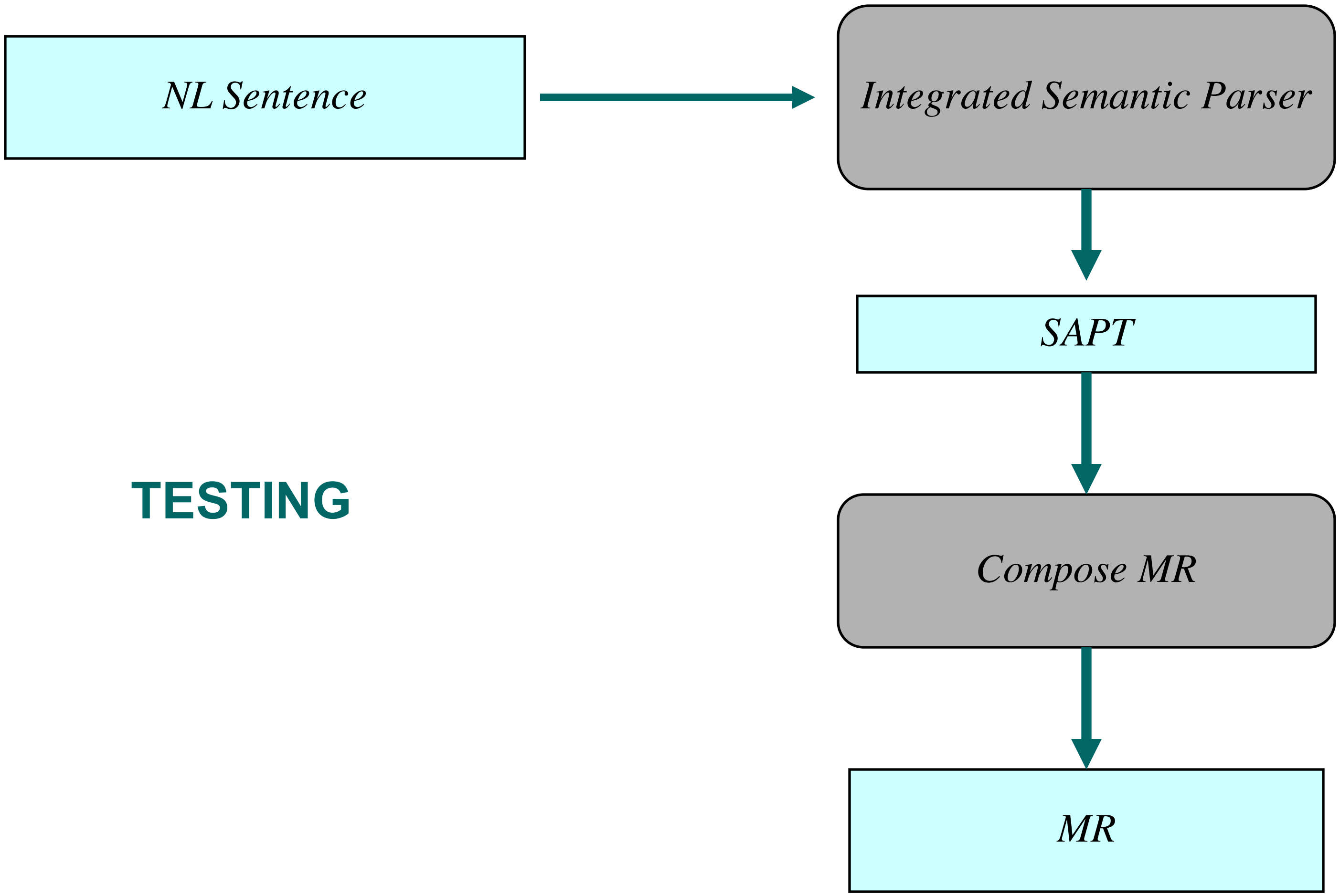
MR: (bowner (player our {2}))

SCISSOR Overview



TRAINING

SCISSOR Overview



TESTING

Integrated Syntactic-Semantic Parsing

- Find a SAPT with the maximum probability
- A *lexicalized head-driven* syntactic parsing model
- Extended Collins (1997) syntactic parsing model to *generate semantic labels simultaneously with syntactic labels*
- Smoothing
  - Each label in SAPT is the combination of a syntactic label and a semantic label which increases *data sparsity*
  - Break the parameters down

$$\begin{aligned} & P_h(\mathrm{H} \mid \mathrm{P}, \mathrm{w}) \\ &= P_h(\mathrm{H}_{syn}, \mathrm{H}_{sem} \mid \mathrm{P}, \mathrm{w}) \\ &= P_h(\mathrm{H}_{syn} \mid \mathrm{P}, \mathrm{w}) \times P_h(\mathrm{H}_{sem} \mid \mathrm{P}, \mathrm{w}, \mathrm{H}_{syn}) \end{aligned}$$

Experimental Corpora

- CLang (Kate, Wong & Mooney, 2005)
  - 300 pieces of coaching advice
  - 22.52 words per sentence
- Geoquery (Zelle & Mooney, 1996)
  - 880 queries on a geography database
  - 7.48 word per sentence
  - MRL: Prolog and FunQL

Prolog vs. FunQL (Wong & Mooney, 2007b)

What are the rivers in Texas?

Prolog:

$answer(x_1, (river(x_1), loc(x_1,x_2), equal(x_2,stateid(texas))))$

FunQL:

answer(river(loc_2(stateid(texas))))

Logical forms: widely used as MRLs in computational semantics, support reasoning

Prolog vs. FunQL (Wong & Mooney, 2007b)

What are the rivers in Texas?

Flexible order

Prolog:

answer(x_1, ($\boxed{\text{river}(x_1), \text{loc}(x_1,x_2), \text{equal}(x_2,\text{stateid}(\text{texas}))}$))

FunQL:

answer($\boxed{\text{river}(\text{loc\_2}(\text{stateid}(\text{texas})))}$)

Strict order

Better generalization on Prolog

Experimental Methodology

- Standard 10-fold cross validation
- Correctness
  - CLang: exactly matches the correct MR
  - Geoquery: retrieves the same answers as the correct MR
- Metrics
  - **Precision**: % of the returned MRs that are correct
  - **Recall**: % of NLs with their MRs correctly returned
  - **F-measure**: harmonic mean of precision and recall

Compared Systems

- **COCKTAIL** (Tang & Mooney, 2001)
  - **Deterministic**, inductive logic programming
- **WASP** (Wong & Mooney, 2006)
  - **Semantic grammar**, machine translation
- **KRISP** (Kate & Mooney, 2006)
  - **Semantic grammar**, string kernels
- **Z&C** (Zettleymoyer & Collins, 2007)
  - **Syntax-based**, combinatory categorial grammar (CCG)
- **LU** (Lu et al., 2008)
  - **Semantic grammar**, generative parsing model
- **SCISSOR** (Ge & Mooney 2005)
  - **Integrated syntactic-semantic parsing**

Compared Systems

- **COCKTAIL** (Tang & Mooney, 2001) → Hand-built lexicon for *Geoquery*
  - **Deterministic**, inductive logic programming
- **WASP** (Wong & Mooney, 2006)
  - **Semantic grammar**, machine translation
- **KRISP** (Kate & Mooney, 2006)
  - **Semantic grammar**, string kernels
- **Z&C** (Zettleymoyer & Collins, 2007) → Small part of the lexicon hand-built
  - **Syntax-based**, combinatory categorial grammar (CCG)
- **LU** (Lu et al., 2008)
  - **Semantic grammar**, generative parsing model
- **SCISSOR** (Ge & Mooney 2005)
  - **Integrated syntactic-semantic parsing**

Compared Systems

- **COCKTAIL** (Tang & Mooney, 2001)
  - **Deterministic**, inductive logic programming
- **WASP** (Wong & Mooney, 2006, 2007b) → **λ-WASP**, handling logical forms
  - **Semantic grammar**, machine translation
- **KRISP** (Kate & Mooney, 2006)
  - **Semantic grammar**, string kernels
- **Z&C** (Zettleymoyer & Collins, 2007)
  - **Syntax-based**, combinatory categorial grammar (CCG)
- **LU** (Lu et al., 2008)
  - **Semantic grammar**, generative parsing model
- **SCISSOR** (Ge & Mooney 2005)
  - **Integrated syntactic-semantic parsing**

Results on CLang

	Precision	Recall	F-measure
COCKTAIL	-	-	-
SCISSOR	89.5	73.7	80.8
WASP	88.9	61.9	73.0
KRISP	85.2	61.9	71.7
Z&C	-	-	-
LU	82.4	57.7	67.8

Memory overflow

Not reported

(LU: F-measure after reranking is 74.4%)

Results on CLang

	Precision	Recall	F-measure
SCISSOR	89.5	73.7	80.8
WASP	88.9	61.9	73.0
KRISP	85.2	61.9	71.7
LU	82.4	57.7	67.8

(LU: F-measure after reranking is 74.4%)

Results on Geoquery

	Precision	Recall	F-measure
SCISSOR	92.1	72.3	81.0
WASP	87.2	74.8	80.5
KRISP	93.3	71.7	81.1
LU	86.2	81.8	84.0
COCKTAIL	89.9	79.4	84.3
λ-WASP	92.0	86.6	89.2
Z&C	95.5	83.2	88.9

FunQL

Prolog

(LU: F-measure after reranking is 85.2%)

Results on Geoquery (FunQL)

	Precision	Recall	F-measure
SCISSOR	92.1	72.3	81.0
WASP	87.2	74.8	80.5
KRISP	93.3	71.7	81.1
LU	86.2	81.8	84.0

competitive

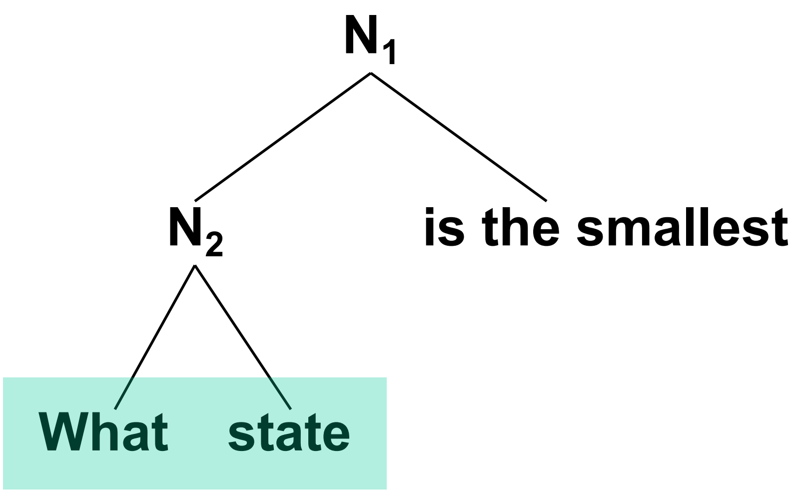
(LU: F-measure after reranking is 85.2%)

When the Prior Knowledge of Syntax Does Not Help

- Geoquery: 7.48 word per sentence
- Short sentence
  - Sentence structure can be feasibly learned from NLs paired with MRs
- **Gain from knowledge of syntax** vs. **flexibility loss**

Limitation of Using Prior Knowledge of Syntax

Traditional syntactic analysis

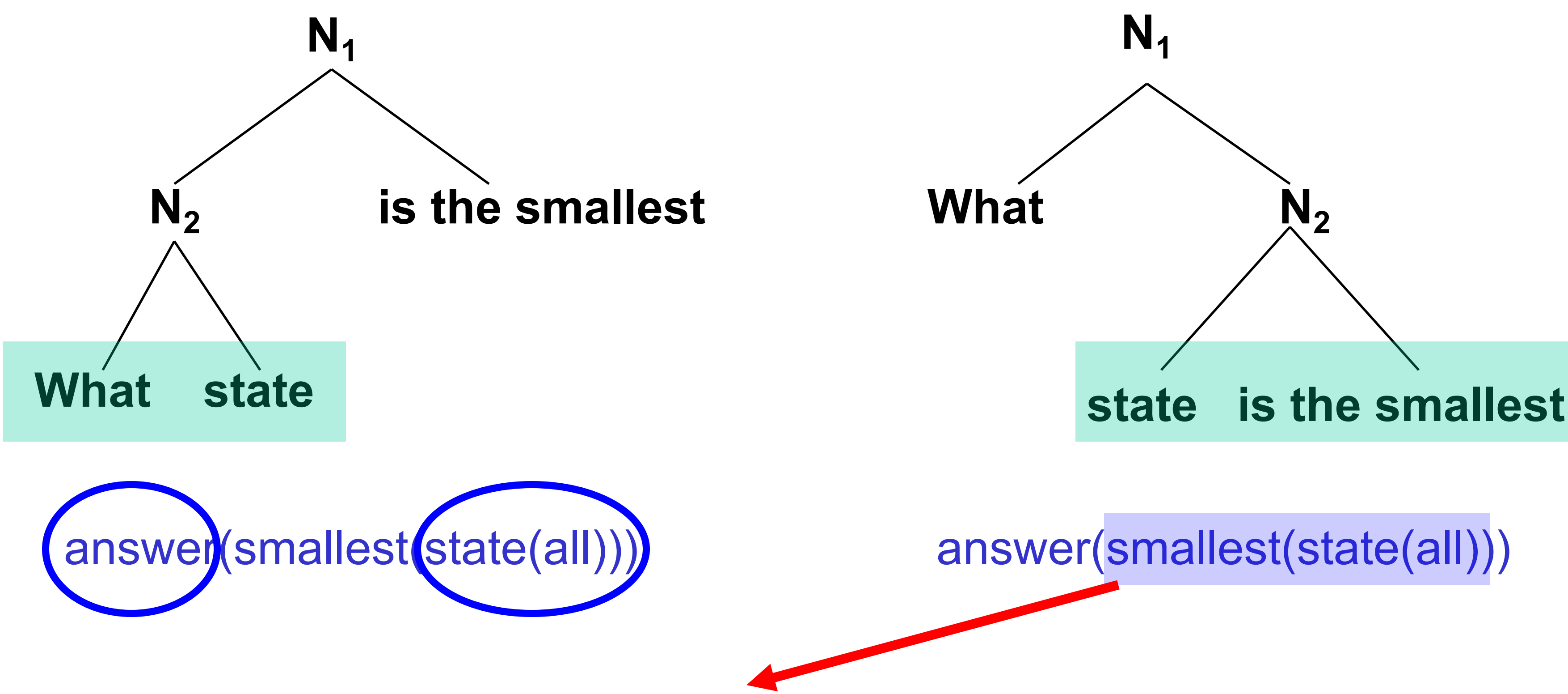


answer(smallest(state(all)))

Limitation of Using Prior Knowledge of Syntax

Traditional syntactic analysis

Semantic grammar



Isomorphic syntactic structure with MR

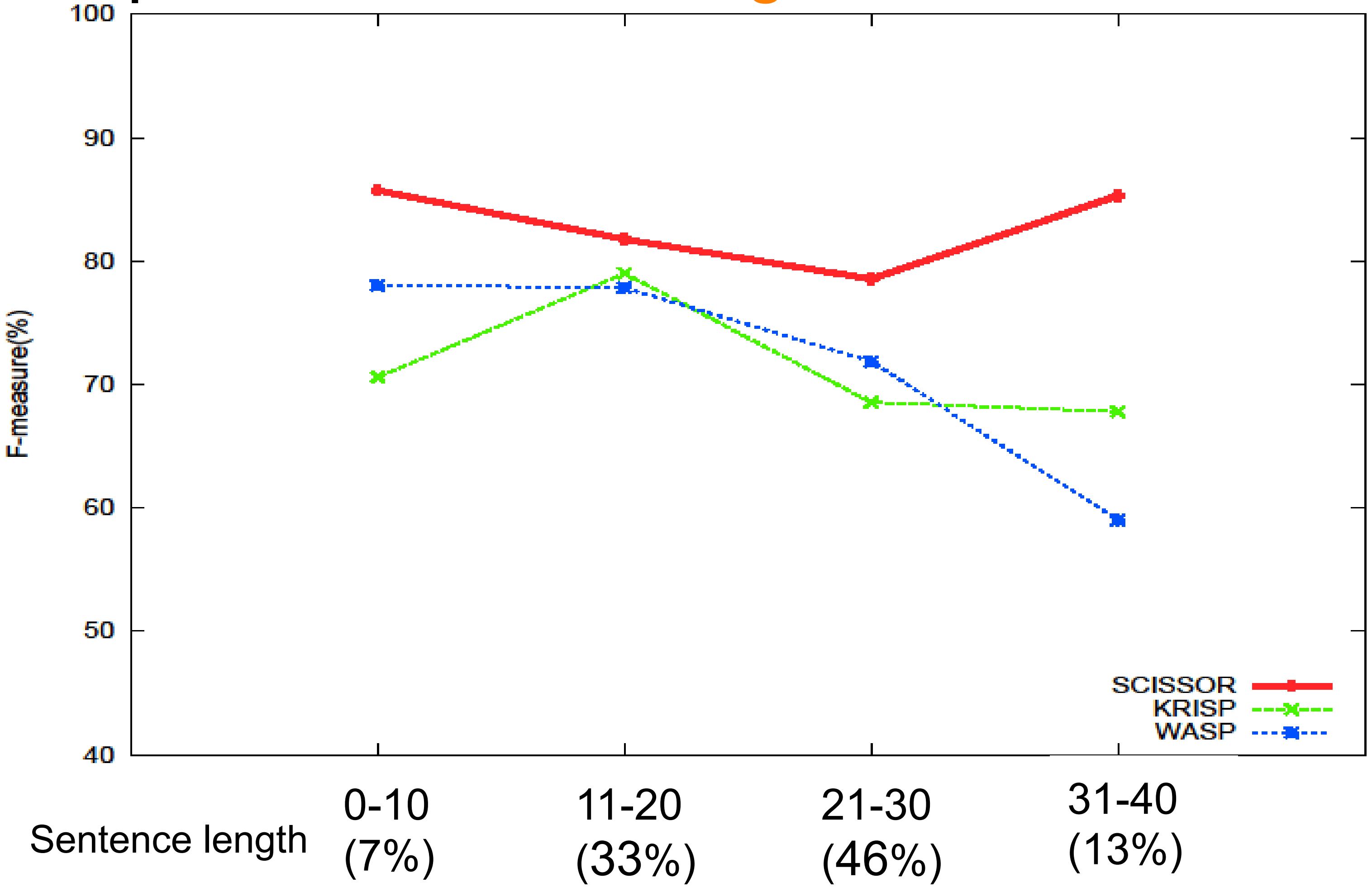
Better generalization

When the Prior Knowledge of Syntax Does Not Help

- Geoquery: 7.48 word per sentence
- Short sentence
  - Sentence structure can be feasibly learned from NLs paired with MRs
- **Gain from knowledge of syntax** vs. **flexibility loss**

Clang Results with Sentence Length

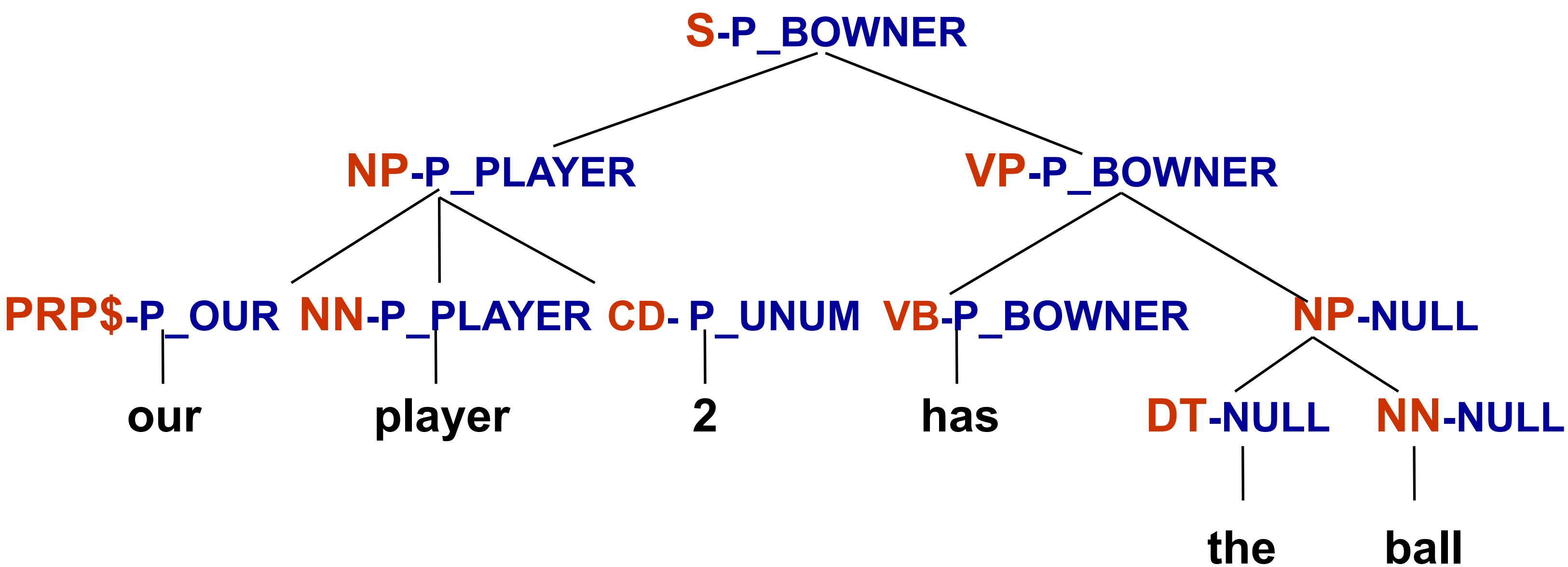
Knowledge of syntax improves performance on *long sentences*



SYNSEM Ge & Mooney (2009)

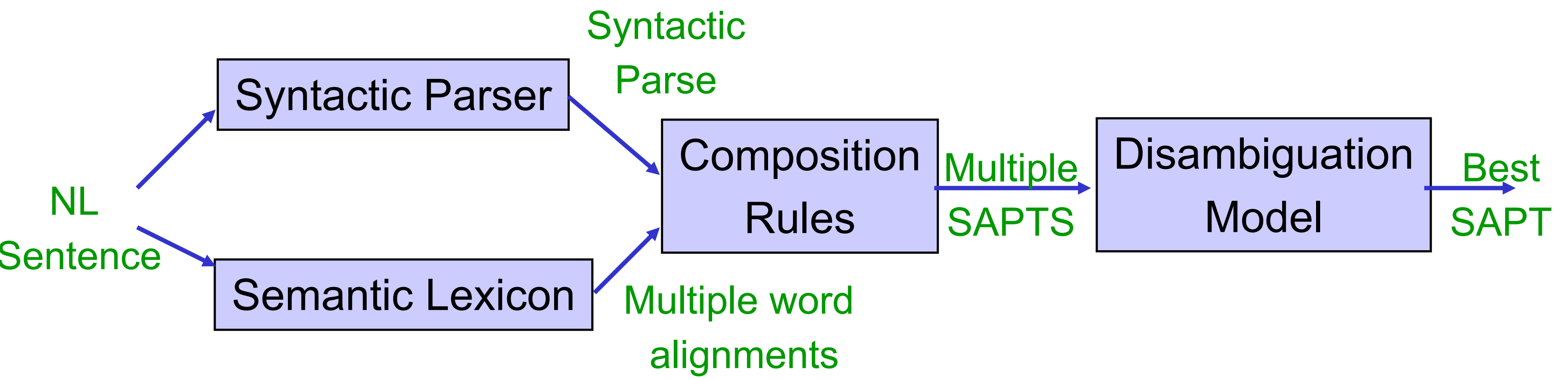
- **SCISSOR** requires extra SAPT annotation for training
- Must learn both syntax and semantics from same limited training corpus
- High performance **syntactic parsers** are available that are trained on existing large corpora (Collins, 1997; Charniak & Johnson, 2005)

SCISSOR Requires SAPT Annotation

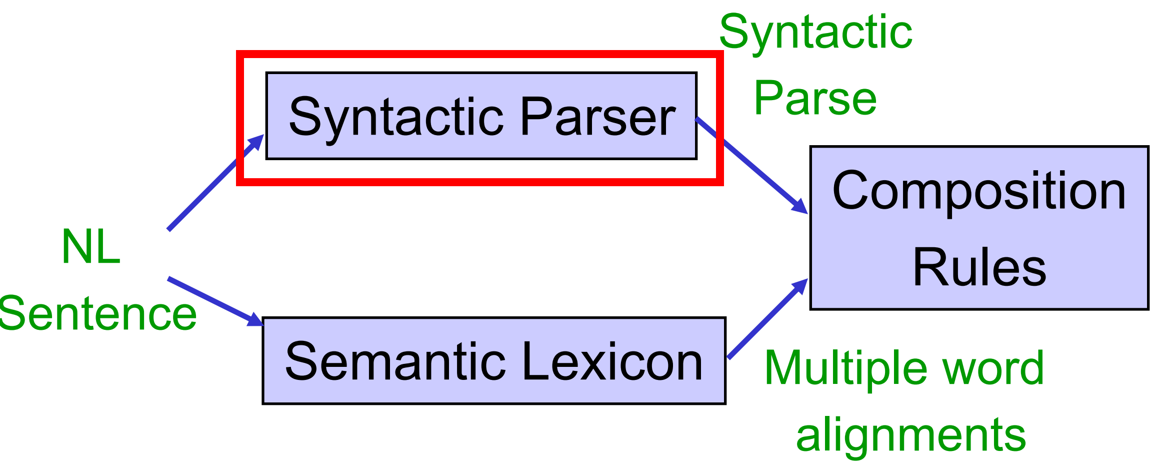


Time consuming.
Automate it!

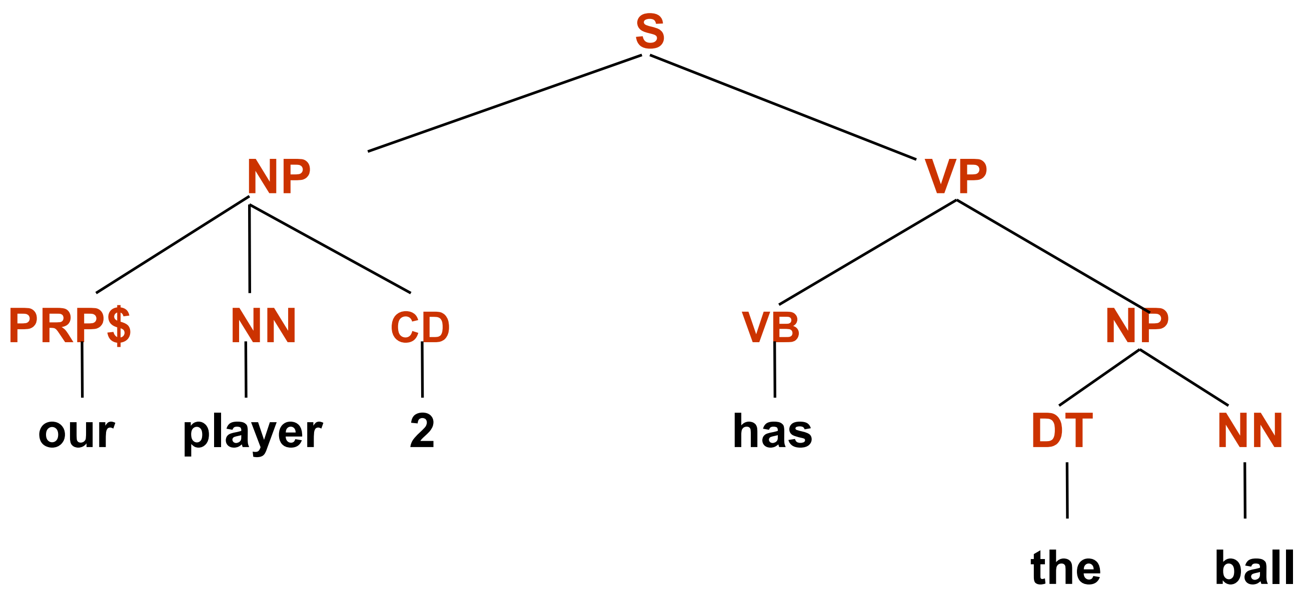
SYNSEM Overview Ge & Mooney (2009)



SYNSEM Training : Learn Semantic Knowledge

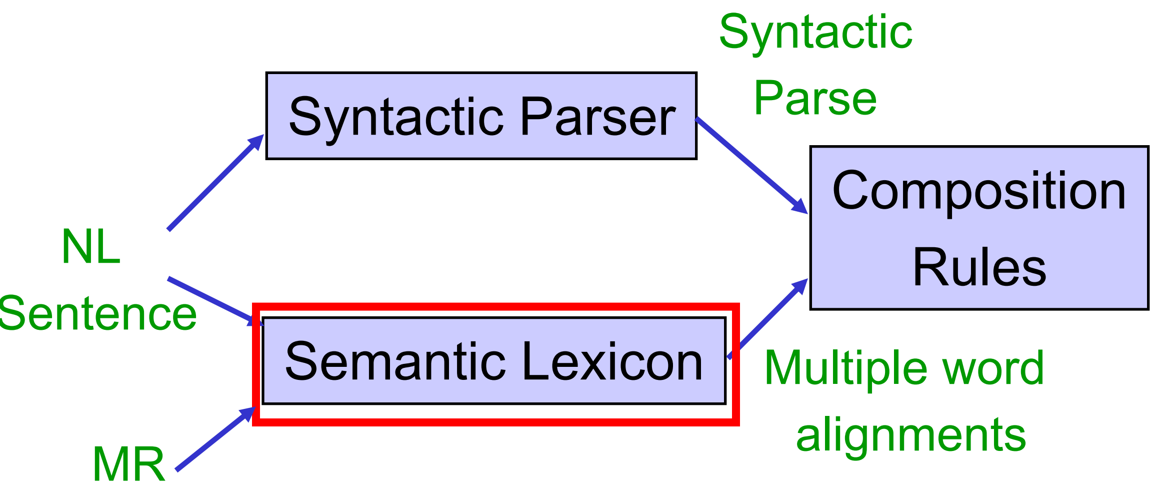


Syntactic Parser

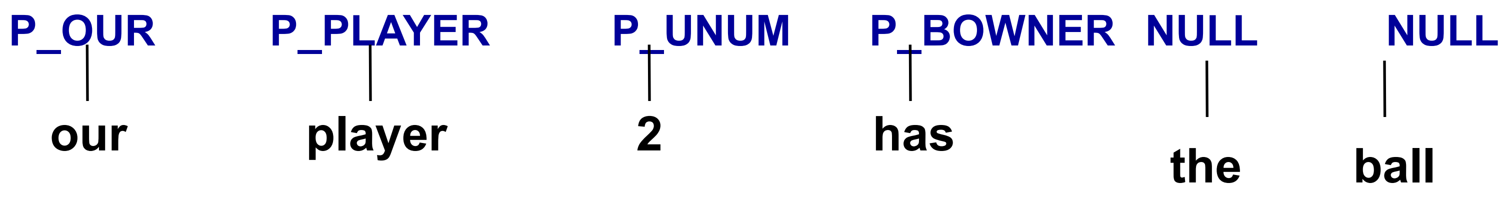


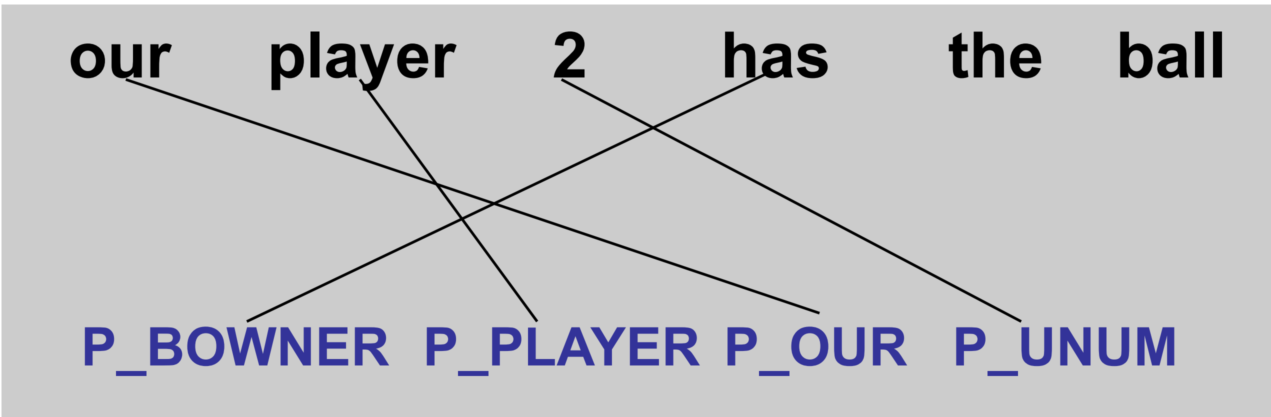
Use a statistical syntactic parser

SYNSEM Training: Learn Semantic Knowledge



Semantic Lexicon



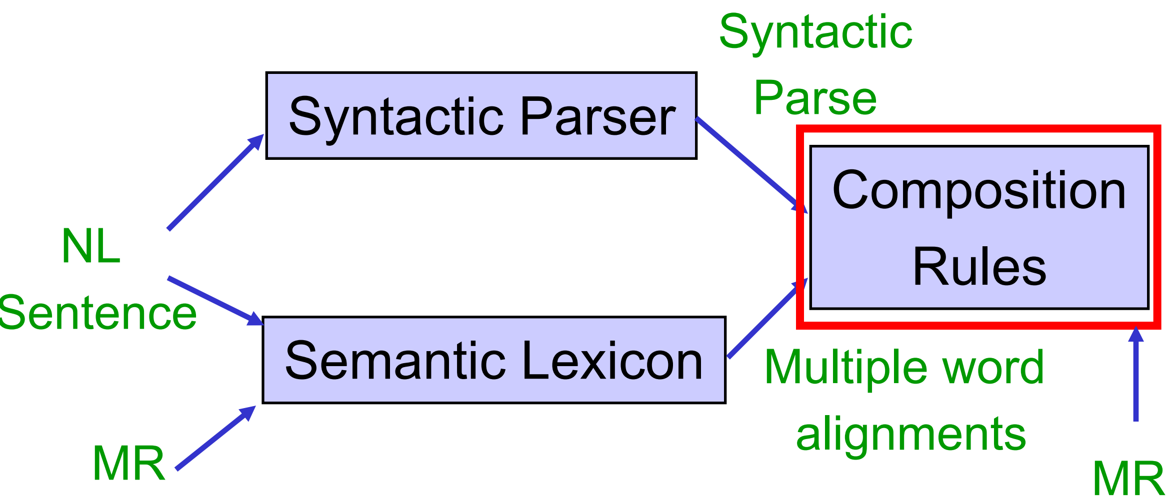


Use a word alignment model (Wong and Mooney (2006))

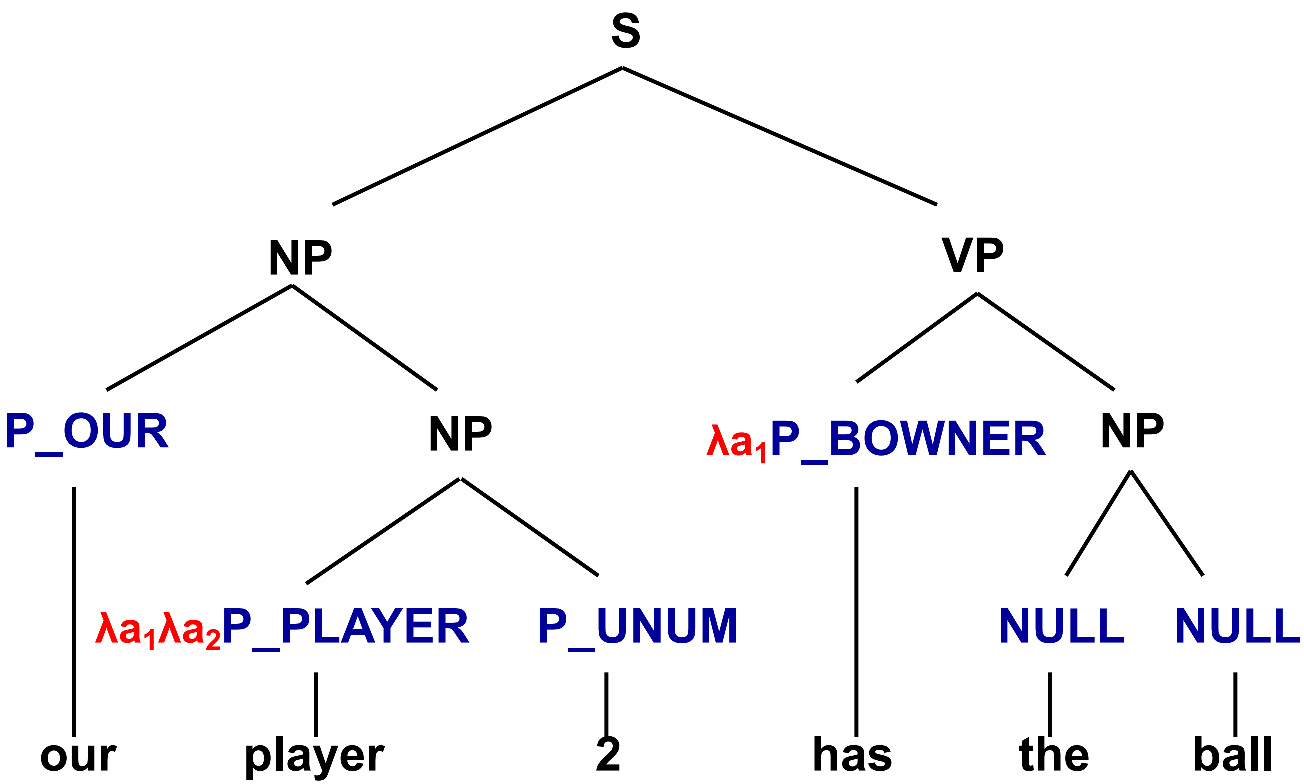
Learning a Semantic Lexicon

- IBM Model 5 word alignment (GIZA++)
- Top 5 word/predicate alignments for each training example
- Assume each word alignment and syntactic parse defines a possible SAPT for composing the correct MR

SYNSEM Training : Learn Semantic Knowledge

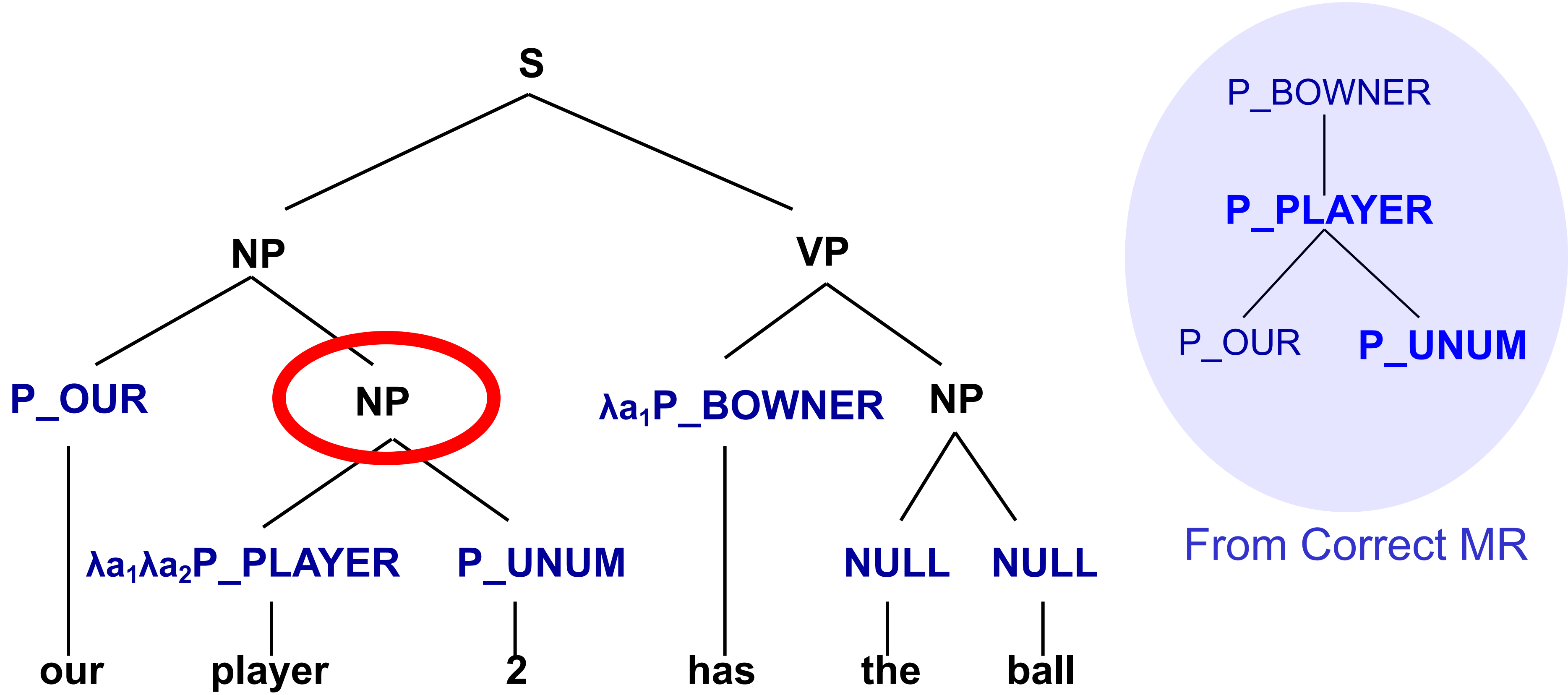


Introduce λ Variables



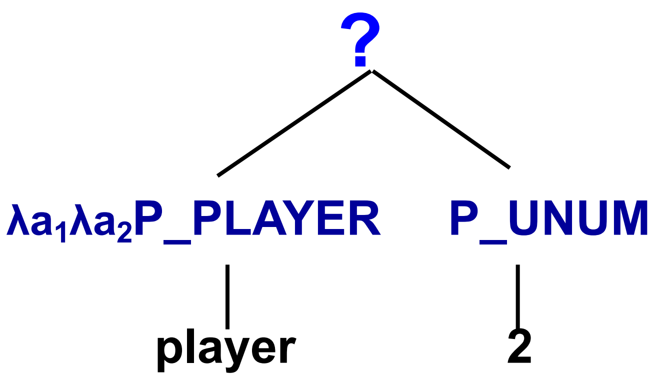
Introducing λvariables in semantic labels for missing arguments
(a_1: the first argument)

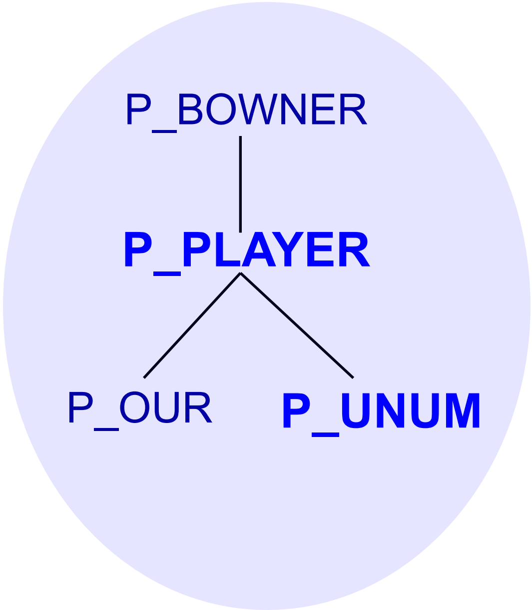
Internal Semantic Labels



How to choose the dominant predicates?

Collect Semantic Composition Rules

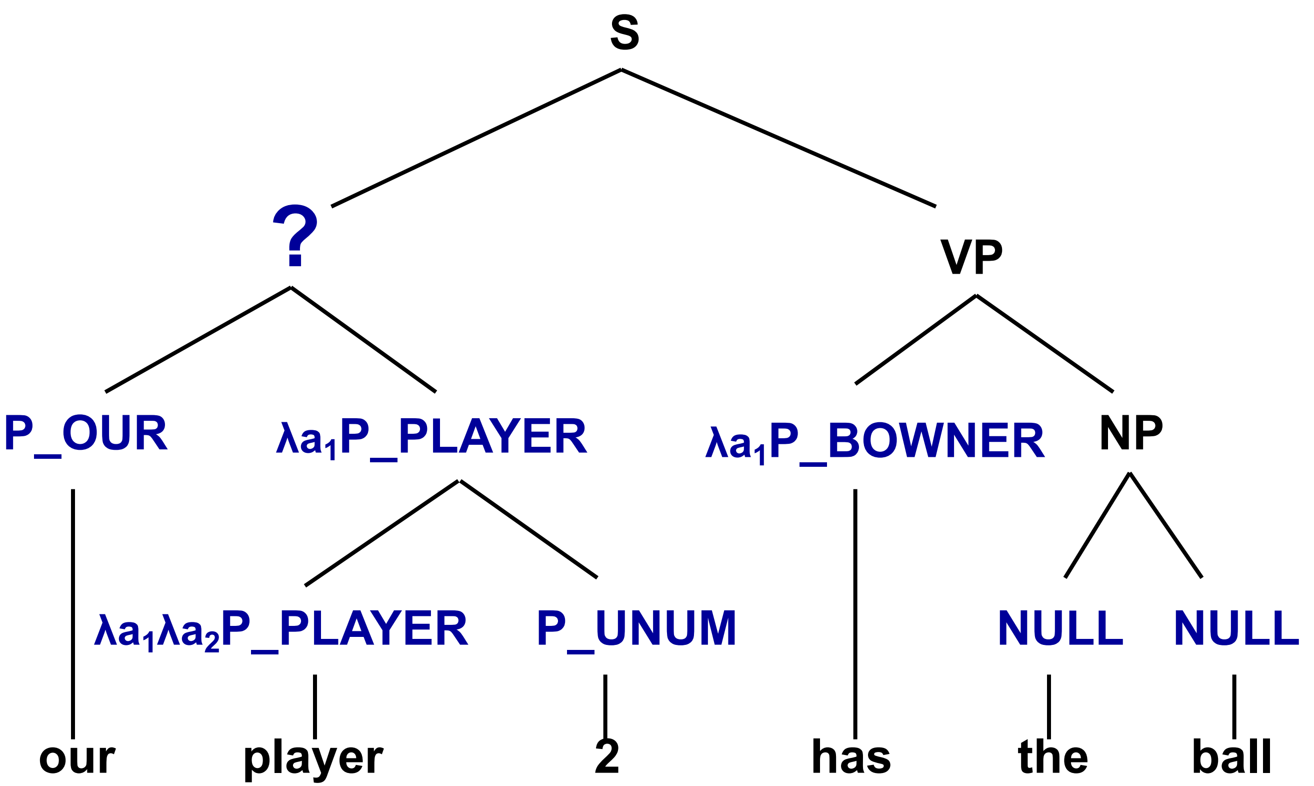


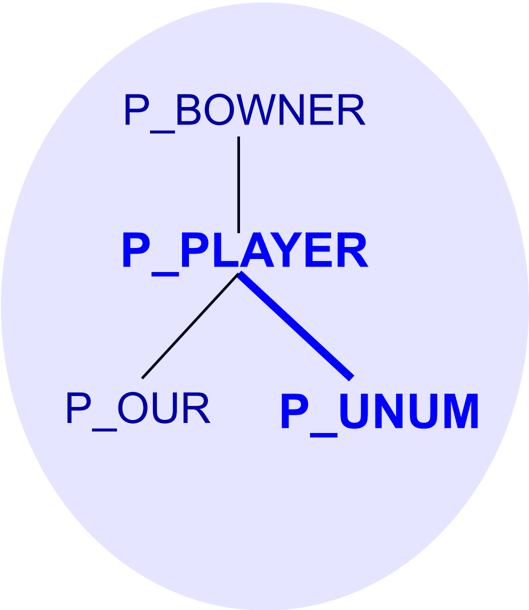


$\lambda a_1 \lambda a_2$**PLAYER** **+ P_UNUM** $\rightarrow$ λa_1 **P_PLAYER ,** $a_2 = c_2$

(c_2: child 2)

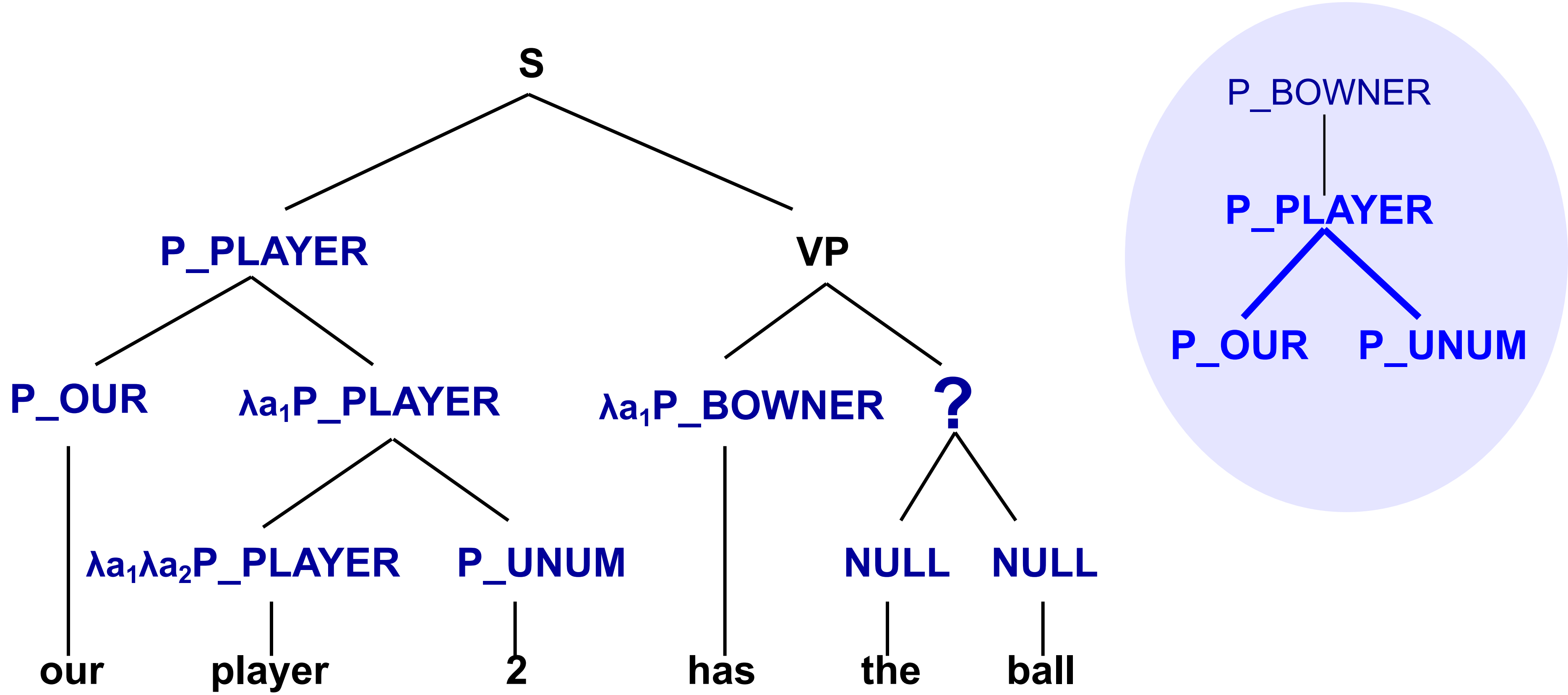
Collect Semantic Composition Rules





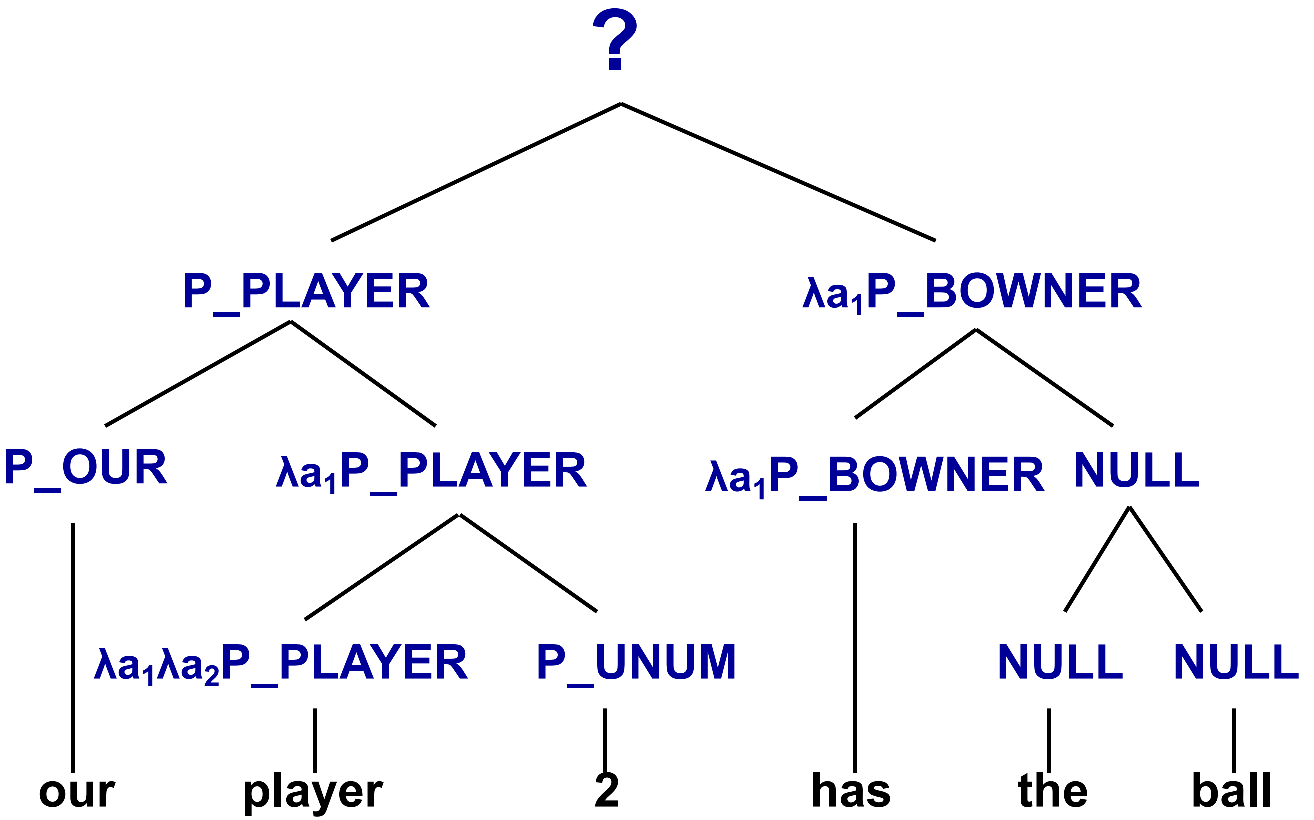
$\lambda a_1 \lambda a_2$PLAYER + P_UNUM → {λa_1P_PLAYER, $a_2 = c_2$}

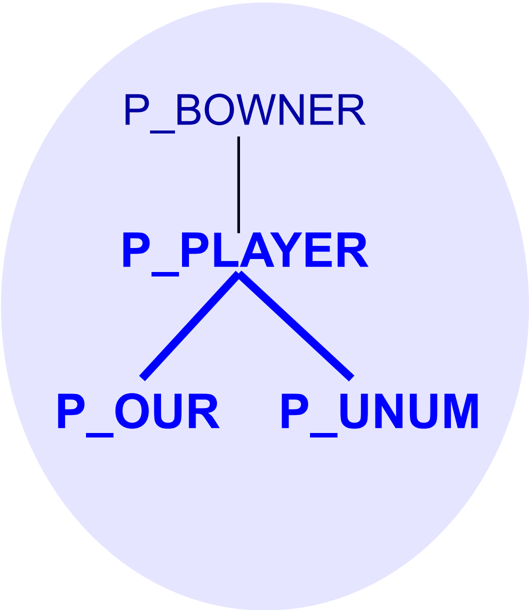
Collect Semantic Composition Rules



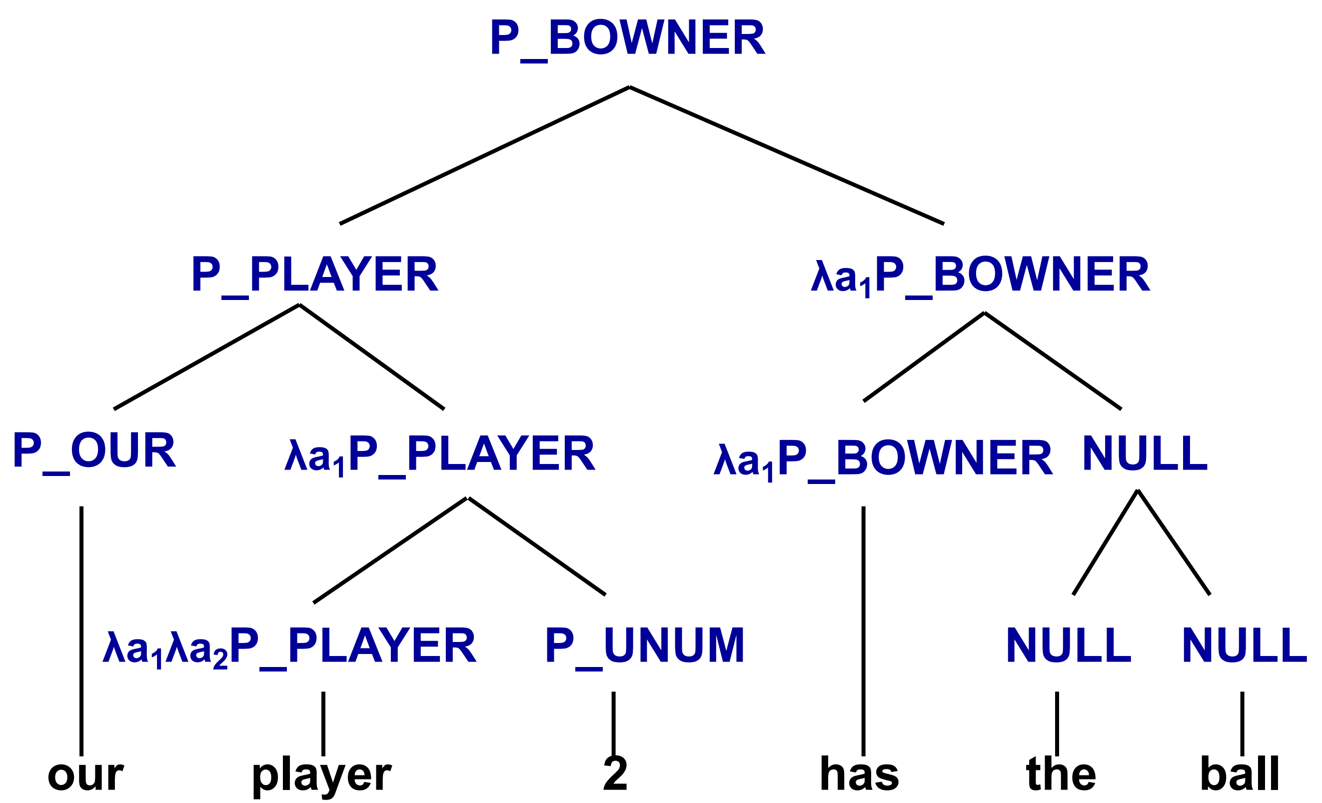
P_OUR +λa_1P_PLAYER → {P_PLAYER, $a_1=c_1$}

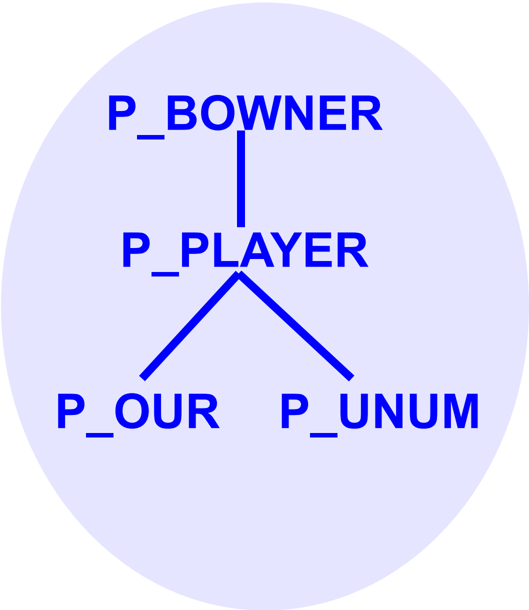
Collect Semantic Composition Rules





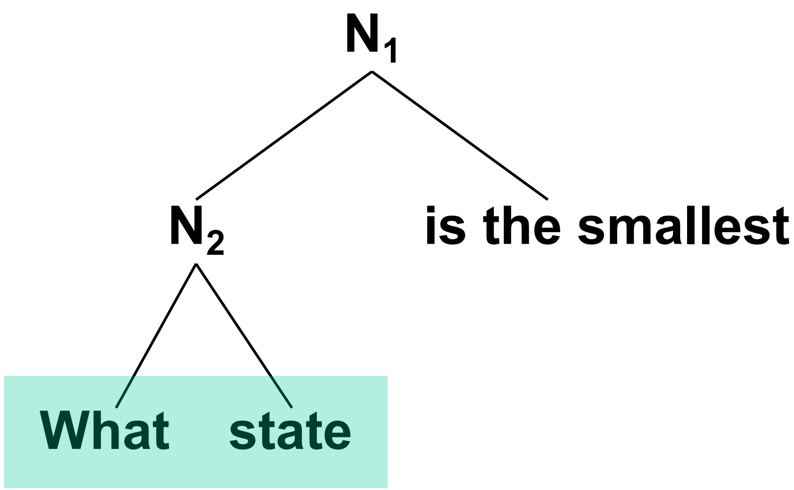
Collect Semantic Composition Rules





P_PLAYER + λa_1**P_BOWNER → {P_BOWNER,** $a_1 = c_1$**}**

Ensuring Meaning Composition



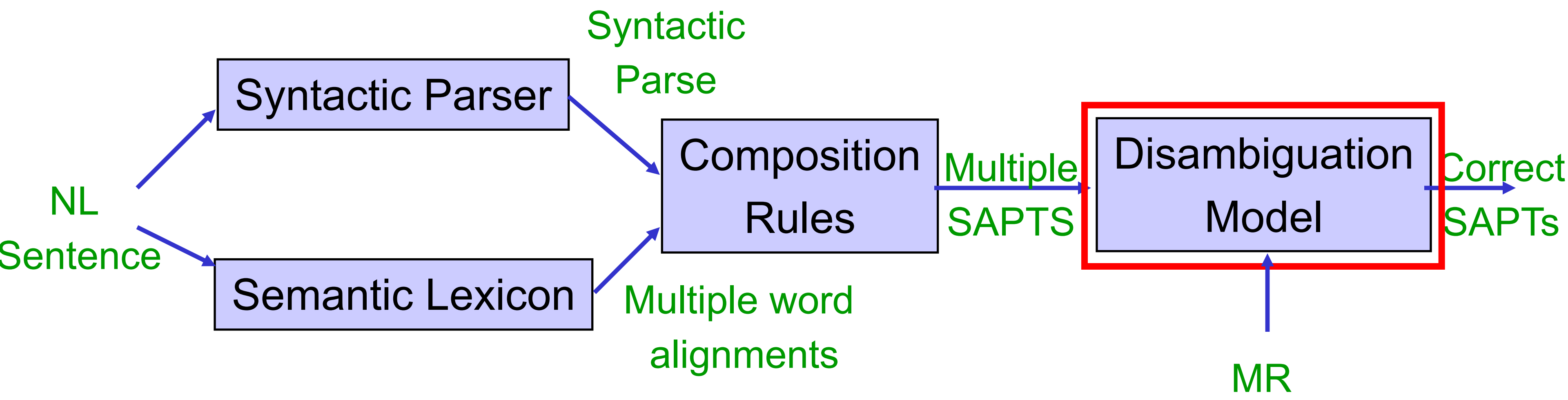
answer(smallest(state(all)))

Non-isomorphism

Ensuring Meaning Composition

- Non-isomorphism between NL parse and MR parse
  - Various linguistic phenomena
  - Word alignment between NL and MRL
  - Use automated syntactic parses
- Introduce ***macro-predicates*** that combine multiple predicates
- Ensure that MR can be composed using a syntactic parse and word alignment

SynSem Training: Learn Disambiguation Model



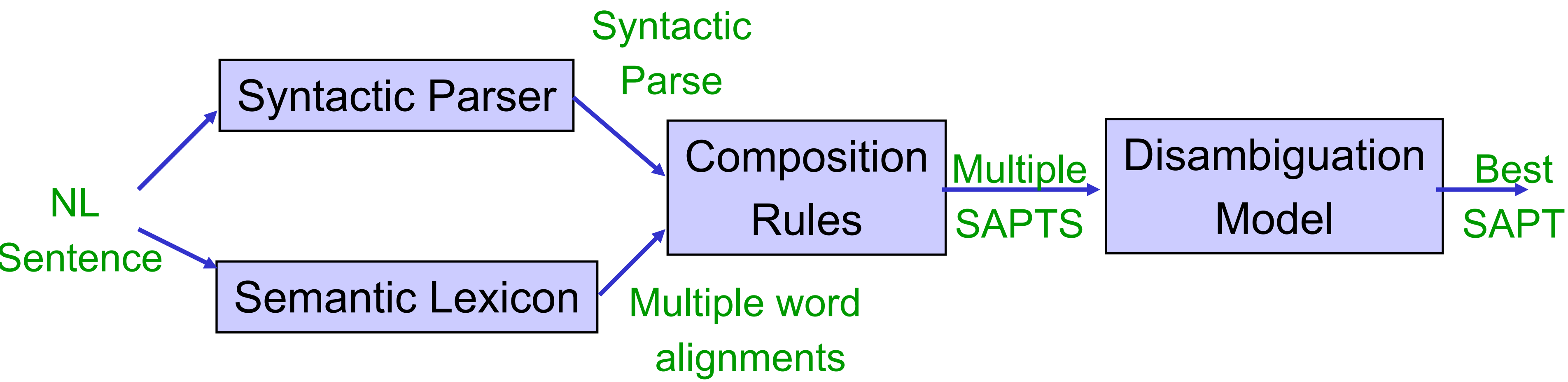
Parameter Estimation

- Apply the learned semantic knowledge to all training examples to generate possible SAPTs
- Use a standard maximum-entropy model similar to that of Zettlemoyer & Collins (2005), and Wong & Mooney (2006)
- Training finds a parameter that (approximately) maximizes the sum of the conditional log-likelihood of the training set including syntactic parses
- Incomplete data since SAPTs are hidden variables

Features

- Lexical features:
  - Unigram features: # that a word is assigned a predicate
  - Bigram features: # that a word is assigned a predicate given its previous/subsequent word.
- Rule features: # a composition rule applied in a derivation

SYNSEM Testing



Syntactic Parsers (Bikel,2004)

- WSJ only
  - CLang(**SYN0**): F-measure=82.15%
  - Geoquery(**SYN0**) : F-measure=76.44%
- WSJ + in-domain sentences
  - CLang(**SYN20**): 20 sentences, F-measure=88.21%
  - Geoquery(**SYN40**): 40 sentences, F-measure=91.46%
- Gold-standard syntactic parses (**GOLDSYN**)

Questions

- **Q1**. Can SYNSEM produce accurate semantic interpretations?
- **Q2**. Can more accurate Treebank syntactic parsers produce more accurate semantic parsers?

Results on CLang

	Precision	Recall	F-measure
GOLDSYN	**84.7**	**74.0**	**79.0**
SYN20	**85.4**	**70.0**	**76.9**
SYN0	**87.0**	**67.0**	**75.7**
SCISSOR	**89.5**	**73.7**	**80.8**
WASP	88.9	61.9	73.0
KRISP	85.2	61.9	71.7
LU	82.4	57.7	67.8

SYNSEM

SAPTs

(LU: F-measure after reranking is 74.4%)

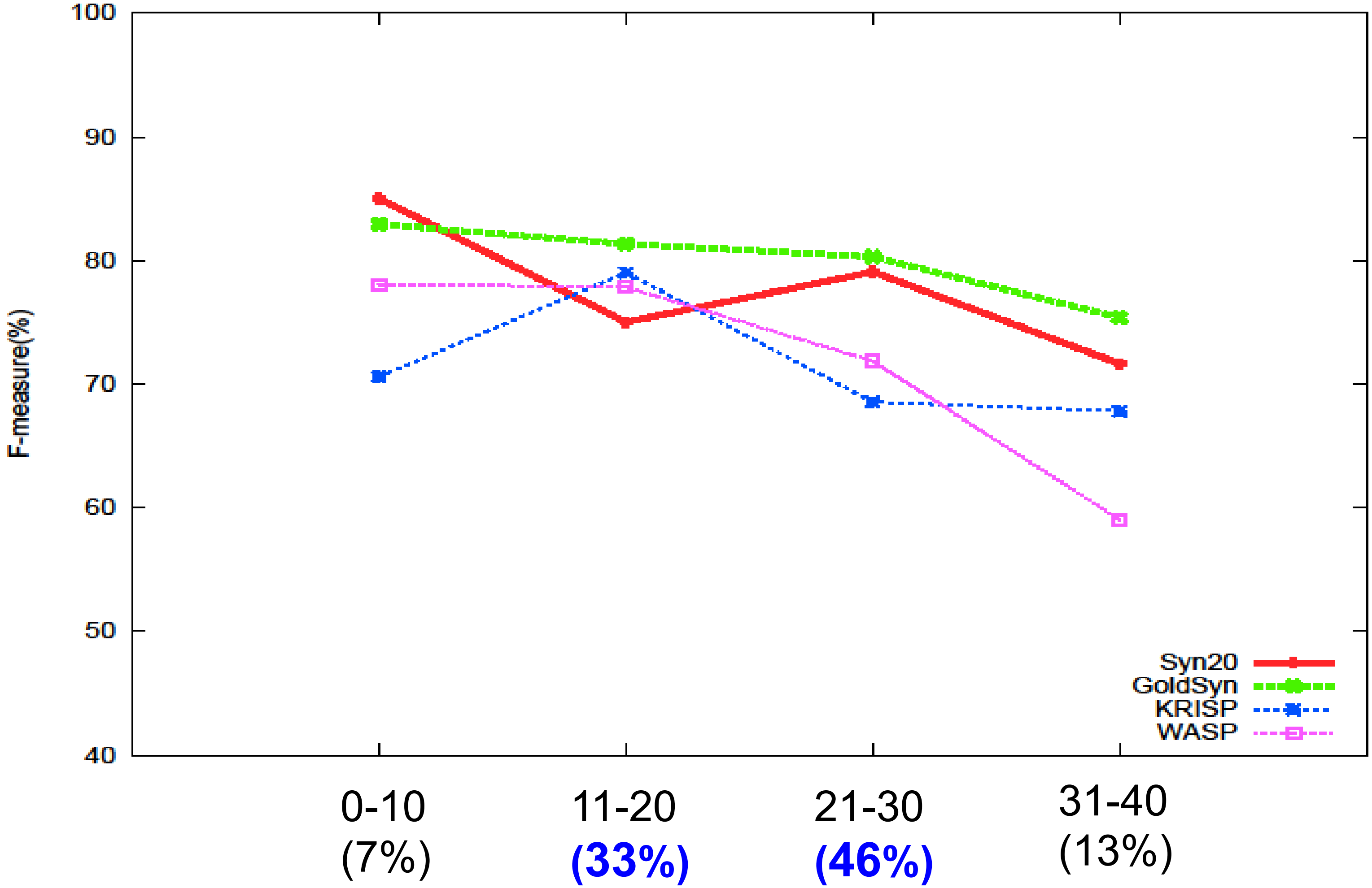
GOLDSYN > SYN20 > SYN0

Questions

- **Q1**. Can SynSem produce accurate semantic interpretations? **[yes]**
- **Q2**. Can more accurate Treebank syntactic parsers produce more accurate semantic parsers? **[yes]**
- **Q3**. Does it also improve on long sentences?

Detailed Clang Results on Sentence Length

↑ Prior Knowledge + ↓ Flexibility + ↓ Syntactic error = ?



Questions

- **Q1**. Can SynSem produce accurate semantic interpretations? **[yes]**
- **Q2**. Can more accurate Treebank syntactic parsers produce more accurate semantic parsers? **[yes]**
- **Q3**. Does it also improve on long sentences? **[yes]**
- **Q4**. Does it improve on limited training data due to the prior knowledge from large treebanks?

Results on Clang (training size = 40)

	Precision	Recall	F-measure
GOLDSYN	**61.1**	**35.7**	**45.1**
SYN20	**57.8**	**31.0**	**40.4**
SYN0	**53.5**	**22.7**	**31.9**
SCISSOR	**85.0**	**23.0**	**36.2**
WASP	88.0	14.4	24.7
KRISP	68.35	20.0	31.0

SYNSEM

SAPTs

The quality of syntactic parser is critically important!

Questions

- **Q1**. Can SynSem produce accurate semantic interpretations? **[yes]**
- **Q2**. Can more accurate Treebank syntactic parsers produce more accurate semantic parsers? **[yes]**
- **Q3**. Does it also improve on long sentences? **[yes]**
- **Q4**. Does it improve on limited training data due to the prior knowledge from large treebanks? **[yes]**

References

Bikel (2004). Intricacies of Collins' Parsing Model. *Computational Linguistics*, 30(4):479-511.

Eugene Charniak and Mark Johnson (2005). Coarse-to-fine n-best parsing and MaxEnt discriminative reranking. In *Proc. of ACL-2005*, pp. 173-180, Ann Arbor, MI, June 2005.

Michal J. Collins (1997). Three generative, lexicalized models for syntactic parsing. In *Proc. of ACL-97*, pages 16-23, Madrid, Spain.

Ruifang Ge and Raymond J. Mooney (2005). A statistical semantic parser that integrates syntax and semantics. In *Proc. of CoNLL-2005*, pp. 9-16, Ann Arbor, MI, June 2005.

Ruifang Ge and Raymond J. Mooney (2006). Discriminative reranking for semantic parsing. In *Proc. of COLING/ACL-2006*, pp. 263-270, Syndney, Australia, July 2006.

Ruifang Ge and Raymond J. Mooney (2009). Learning a compositional semantic parser using an existing syntactic parser. In *Proc. of ACL-2009*, pp. 611-619, Suntec, Singapore, August 2009.

Outline

1. Introduction to the task of semantic parsing
   a) Definition of the task
   b) Examples of application domains and meaning representation languages
   c) Distinctions from and relations to other NLP tasks
2. Semantic parsers
   a) Earlier hand-built systems
   b) Learning for semantic parsing
      I. Semantic parsing learning task
      II. Early semantic parser learners
      III. Recent semantic parser learners
      IV. Exploiting syntax for semantic parsing
      V. Underlying commonalities and differences between semantic parsers
   c) Various forms of supervision
3. Semantic parsing beyond a sentence
   a) Learning language from perceptual contexts
   b) Using discourse contexts for semantic parsing
4. Research challenges and future directions
   a) Machine reading of documents: Connecting with knowledge representation
   b) Applying semantic parsing techniques to the Semantic Web
   c) Future research directions
5. Conclusions

Underlying Commonalities and Differences between Semantic Parsers

Underlying Commonalities between Semantic Parsers

- A model to connect language and meaning

A Model to Connect Language and Meaning

- Zettlemoyer & Collins: CCG grammar with semantic types

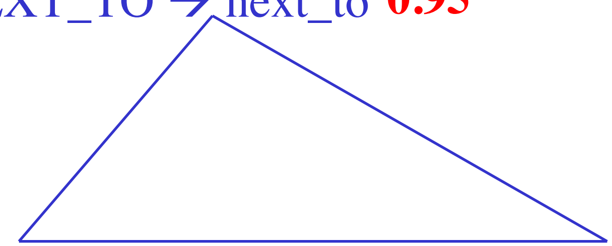
  borders := (S \ NP) / NP : λx.λy.borders(y, x)

- WASP: Synchronous CFG

  QUERY → **What is** CITY / answer(CITY)

- KRISP: Probabilistic string classifiers

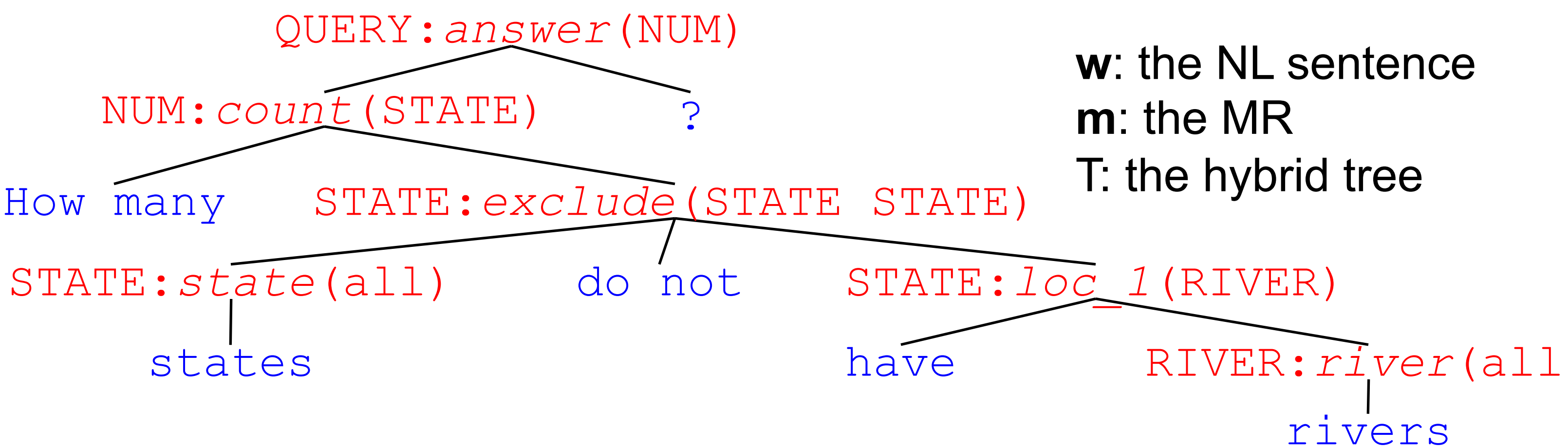
  NEXT_TO → next_to **0.95**



Which rivers run through the states bordering Texas?

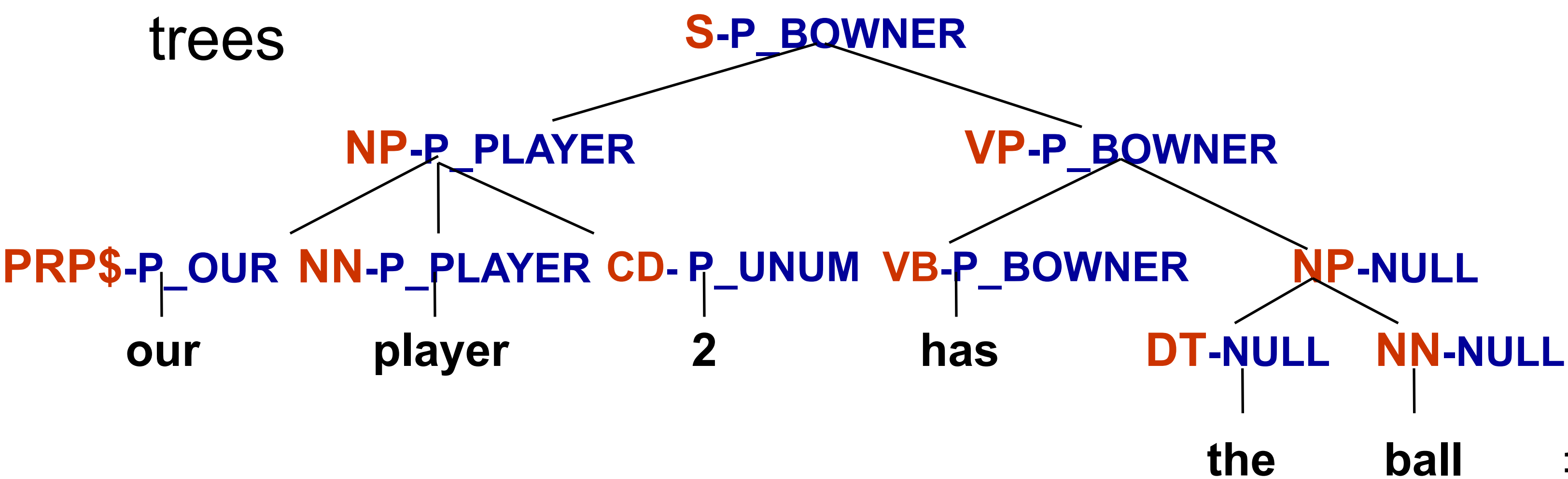
A Model to Connect Language and Meaning

- Lu et al.: Hybrid tree, hybrid patterns



w: the NL sentence
m: the MR
T: the hybrid tree

- SCISSOR/SynSem: Semantically annotated parse trees



Underlying Commonalities between Semantic Parsers

- A model to connect language and meaning
- A mechanism for meaning composition

A Mechanism for Meaning Composition

- Zettlemoyer & Collins: CCG parsing rules
- WASP: Meaning representation grammar
- KRISP: Meaning representation grammar
- Lu et al.: Meaning representation grammar
- SCISSOR: Semantically annotated parse trees
- SynSem: Syntactic parses

Underlying Commonalities between Semantic Parsers

- A model to connect language and meaning
- A mechanism for meaning composition
- Parameters for selecting a meaning representation out of many

Parameters to Select a Meaning Representation

- Zettlemoyer & Collins: Weights for lexical items and CCG parsing rules
- WASP: Weights for grammar productions
- KRISP: SVM weights
- Lu et al.: Generative model parameters
- SCISSOR: Parsing model weights
- SynSem: Parsing model weights

Underlying Commonalities between Semantic Parsers

- A model to connect language and meaning
- A mechanism for meaning composition
- Parameters for selecting a meaning representation out of many
- An iterative EM-like method for training to find the right associations between NL and MR components

An Iterative EM-like Method for Training

- Zettlemoyer & Collins: Stochastic gradient ascent, structured perceptron
- WASP: Quasi-Newton method (L-BFGS)
- KRISP: Re-parse the training sentences to find more refined positive and negative examples
- Lu et al.: Inside-outside algorithm with EM
- SCISSOR: None
- SynSem: Quasi-Newton method (L-BFGS)

Underlying Commonalities between Semantic Parsers

- A model to connect language and meaning
- A mechanism for meaning composition
- Parameters for selecting a meaning representation out of many
- An iterative EM-like loop in training to find the right associations between NL and MR components
- A generalization mechanism

A Generalization Mechanism

- Zettlemoyer & Collins: Different combinations of lexicon items and parsing rules
- WASP: Different combinations of productions
- KRISP: Different combinations of meaning productions, string similarity
- Lu et al.: Different combinations of hybrid patterns
- SCISSOR: Different combinations of parsing productions
- SynSem: Different combinations of parsing productions

Underlying Commonalities between Semantic Parsers

- A model to connect language and meaning
- A mechanism for meaning composition
- Parameters for selecting a meaning representation out of many
- An iterative EM-like loop in training to find the right associations between NL and MR components
- A generalization mechanism

Differences between Semantic Parsers

- Learn lexicon or not

Learn Lexicon or Not

- Learn it as a first step:
  - Zettlemoyer & Collins
  - WASP
  - SynSem
  - COCKTAIL
- Do not learn it:
  - KRISP
  - Lu at el.
  - SCISSOR

Differences between Semantic Parsers

- Learn lexicon or not
- Utilize knowledge of natural language or not

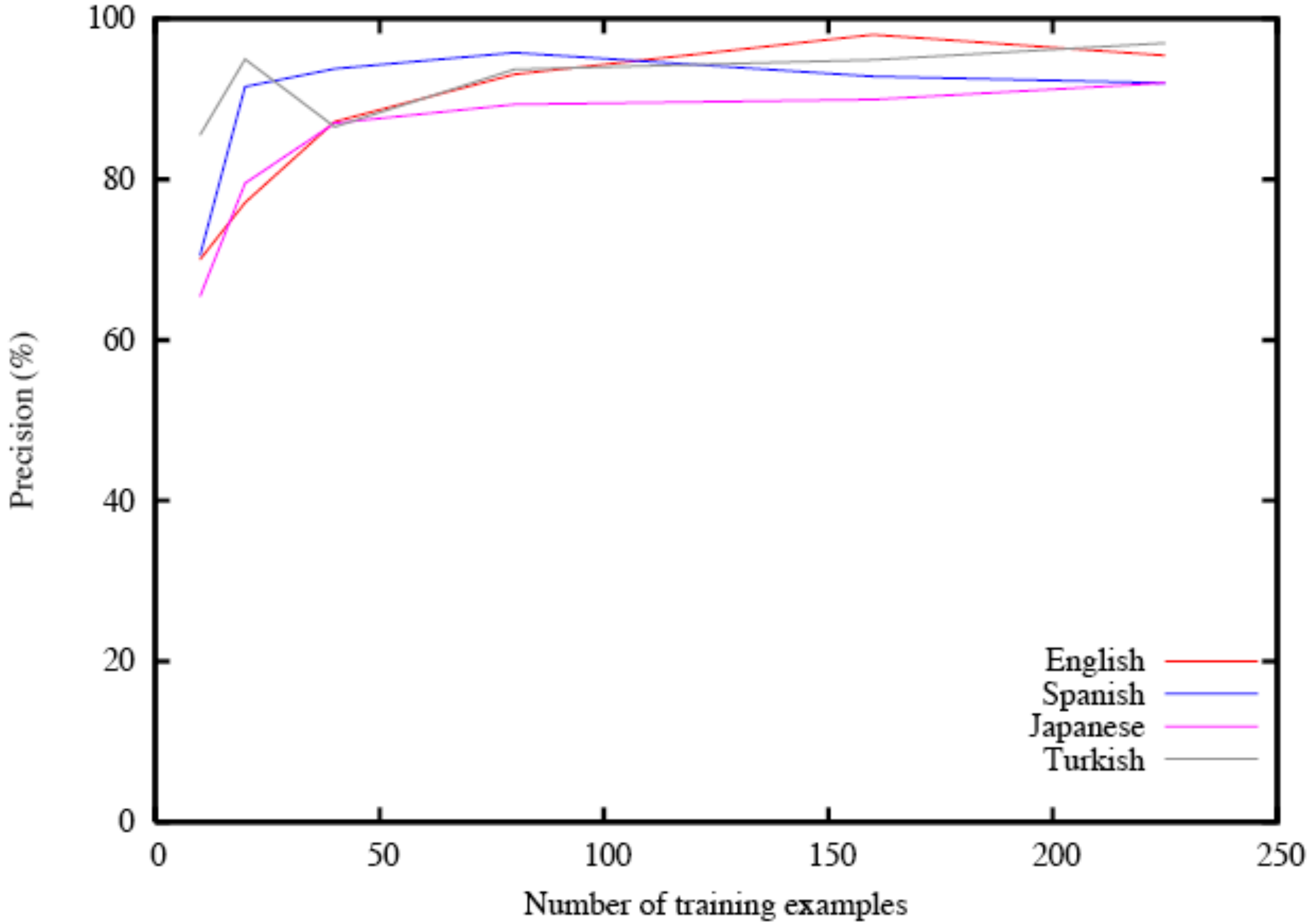
Utilize Knowledge of Natural Language or Not

- Utilize knowledge of English syntax:
  - Zettlemoyer & Collins: CCG
  - SCISSOR: Phrase Structure Grammar
  - SynSem: Phrase Structure Grammar

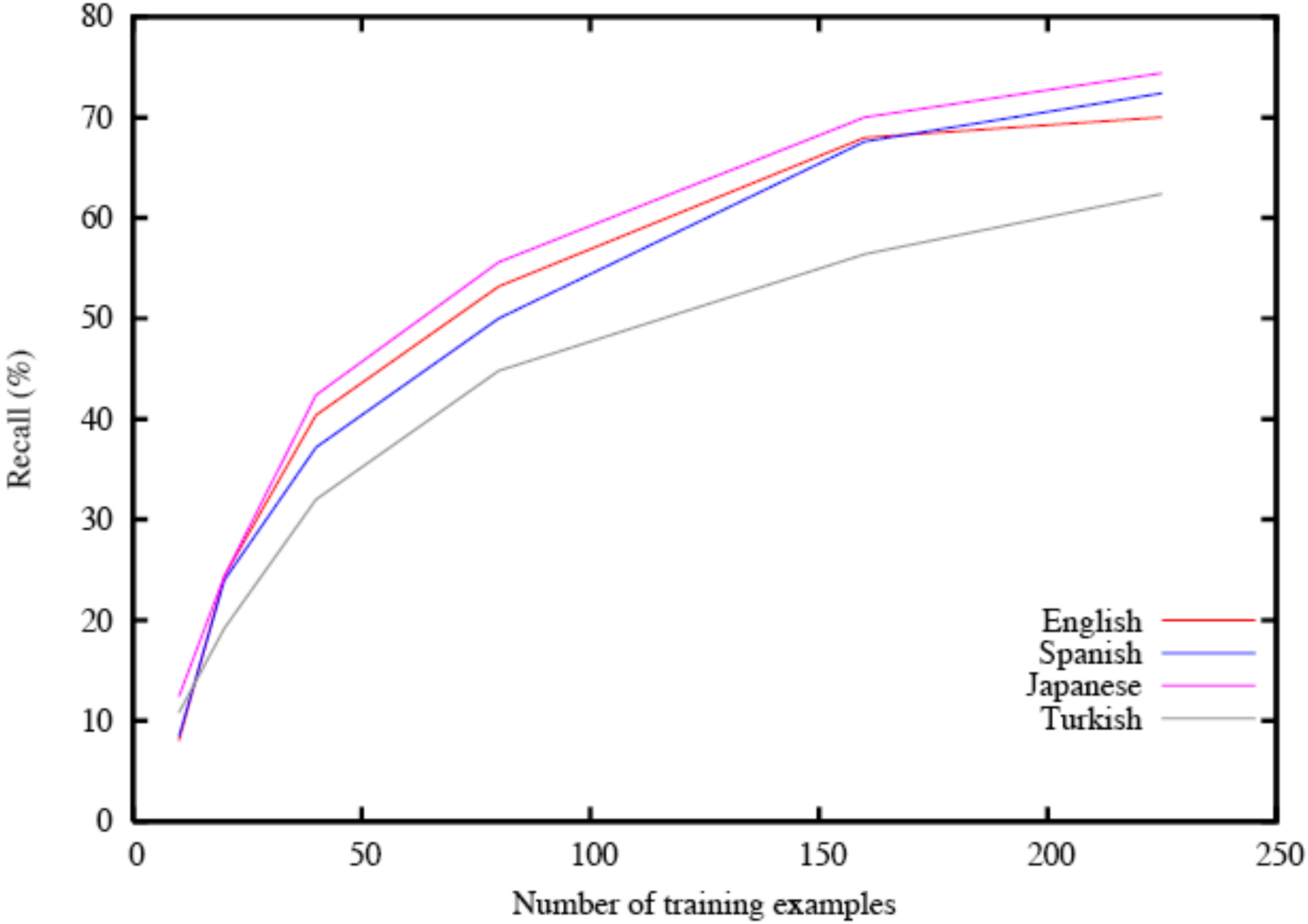
  Leverage from natural language knowledge
- Do not utilize:
  - WASP
  - KRISP
  - Lu et al.

  Portable to other natural languages

Precision Learning Curve for GeoQuery (WASP)



Recall Learning Curve for GeoQuery (WASP)



Differences between Semantic Parsers

- Learn lexicon or not
- Utilize general syntactic parsing grammars or not
- Use matching patterns or not

Use Matching Patterns or Not

- Use matching patterns:
  - Zettlemoter & Collins
  - WASP
  - Lu et al.
  - SCISSOR/SynSem

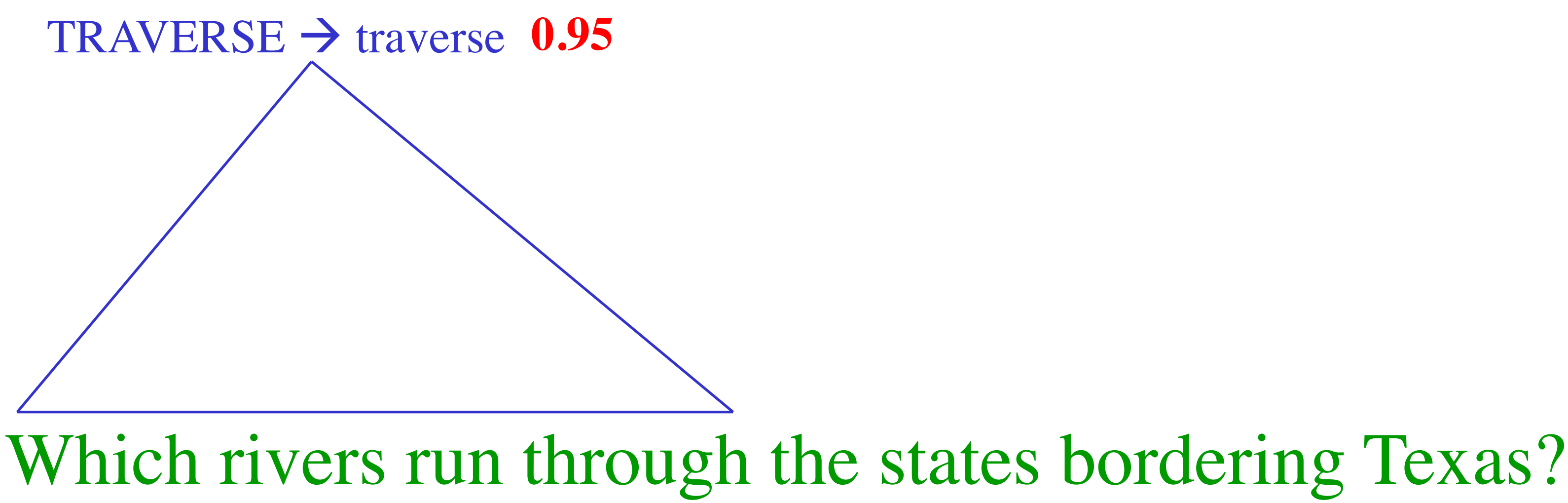
  The systems can be inverted to form a generation system, for e.g. WASP^{-1}
- Do not use matching patterns:
  - KRISP

  Makes it robust to noise

Robustness of KRISP

- KRISP does not use matching patterns
- String-kernel-based classification softly captures wide range of natural language expressions

$\Rightarrow$ Robust to rephrasing and noise

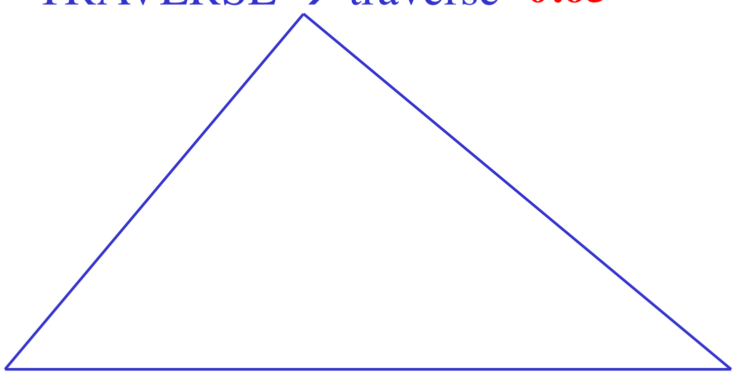


Robustness of KRISP

- KRISP does not use matching patterns
- String-kernel-based classification softly captures wide range of natural language expressions

$\Rightarrow$ Robust to rephrasing and noise

TRAVERSE → traverse **0.65**

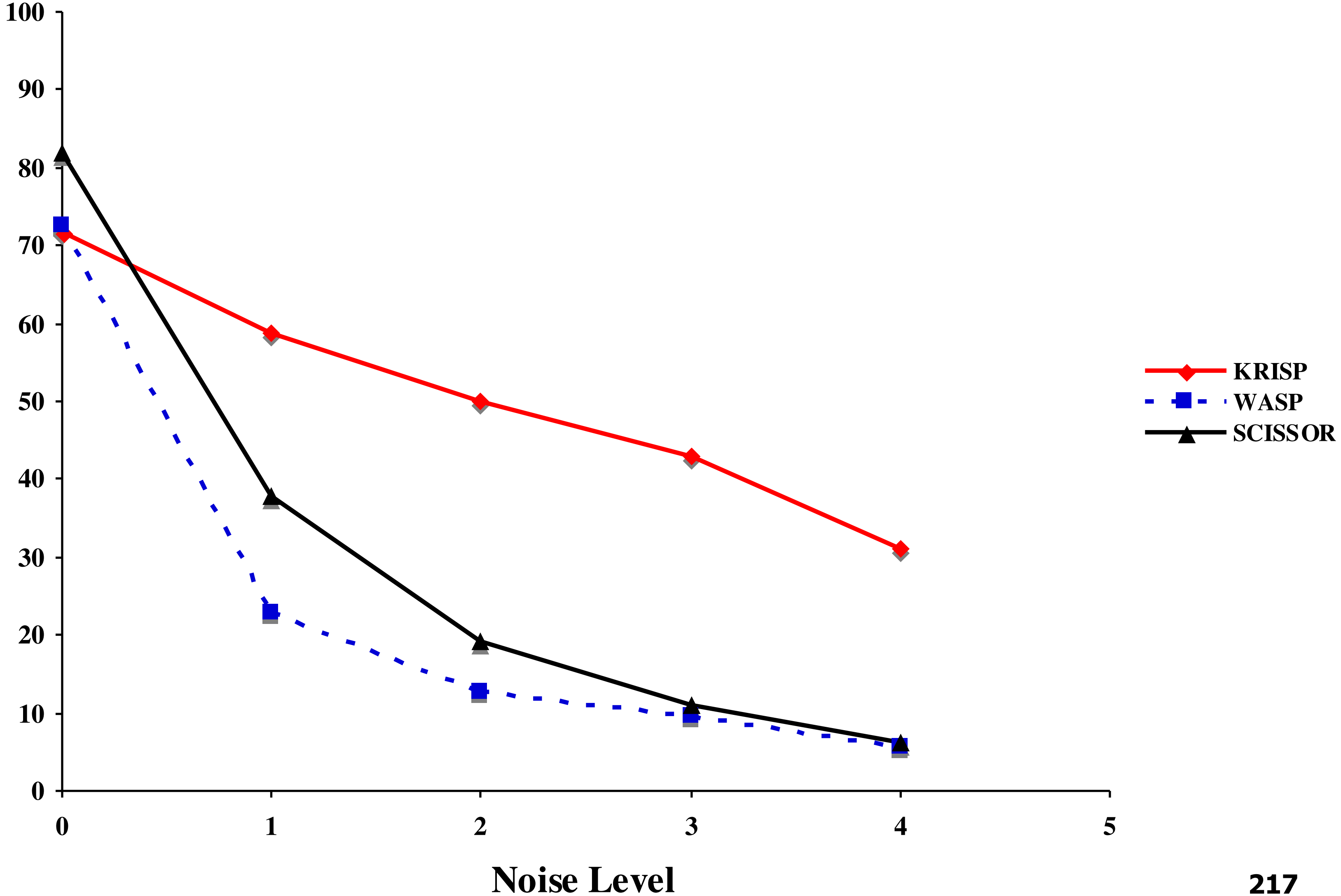


Which rivers through the states bordering Texas?

Experiments with Noisy NL Sentences contd.

- Noise was introduced in the NL sentences by:
  - Adding extra words chosen according to their frequencies in the BNC
  - Dropping words randomly
  - Substituting words with phonetically close high frequency words
- Four levels of noise was created by increasing the probabilities of the above
- We show best F-measures (harmonic mean of precision and recall)

Results on Noisy CLang Corpus



Differences between Semantic Parsers

- Learn lexicon or not
- Utilize general syntactic parsing grammars or not
- Use matching patterns or not
- Different amounts of supervision

Different Amounts of Supervision

- Zettlemoyer & Collins
  - NL-MR pairs, CCG category rules, a small number of predefined lexical items
- WASP, KRISP, Lu et al.
  - NL-MR pairs, MR grammar
- SCISSOR
  - Semantically annotated parse trees
- SynSem
  - NL-MR pairs, syntactic parser

Differences between Semantic Parsers

- Learn lexicon or not
- Utilize general syntactic parsing grammars or not
- Use matching patterns or not
- Different forms of supervision

Outline

1. Introduction to the task of semantic parsing
   a) Definition of the task
   b) Examples of application domains and meaning representation languages
   c) Distinctions from and relations to other NLP tasks
2. Semantic parsers
   a) Earlier hand-built systems
   b) Learning for semantic parsing
   c) Various forms of supervision
3. Semantic parsing beyond a sentence
   a) Learning language from perceptual contexts
   b) Using discourse contexts for semantic parsing
4. Research challenges and future directions
   a) Machine reading of documents: Connecting with knowledge representation
   b) Applying semantic parsing techniques to the Semantic Web
   c) Future research directions
5. Conclusions

Various Forms of Supervision

Semi-Supervised Learning for Semantic Parsing (Kate & Mooney, 2007a)

Supervised Corpus

Which rivers run through the states bordering Texas?

answer(traverse(next_to(stateid(‘texas’))))

What is the lowest point of the state with the largest area?

answer(lowest(place(loc(largest_one(area(state(all)))))))

What is the largest city in states that border California?

answer(largest(city(loc(next_to(stateid('california'))))))

……

Unsupervised Corpus

Which states have a city named Springfield?

What is the capital of the most populous state?

How many rivers flow through Mississippi?

How many states does the Mississippi run through?

How high is the highest point in the smallest state?

Which rivers flow through the states that border California?

…….

- Building supervised training data is expensive
- Extended KRISP to utilize NL sentences not annotated with their MRs, usually cheaply available

SEMISUP-KRISP: Semi-Supervised Semantic Parser Learner (Kate & Mooney, 2007a)

Semantic Parsing

Supervised Corpus

Which rivers run through the states bordering Texas?

answer(traverse(next_to(stateid('texas'))))

What is the lowest point of the state with the largest area?

answer(lowest(place(loc(largest_one(area(state(all)))))))

What is the largest city in states that border California?

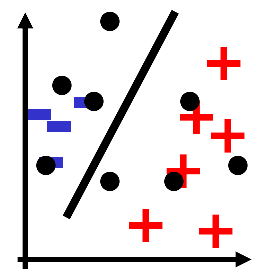
answer(largest(city(loc(next_to(stateid('california'))))))

……

Collect labeled examples

Transductive

SVM classifiers



Learned Semantic parser

Unsupervised Corpus

Which states have a city named Springfield?

What is the capital of the most populous state?

How many rivers flow through Mississippi?

How many states does the Mississippi run through?

How high is the highest point in the smallest state?

Which rivers flow through the states that border California?

…….

Collect *unlabeled* examples

Achieves the same performance with 25% less supervised training data

Semantic Parsing

Response-Driven Learning for Semantic Parsing (Clarke et al. 2010)

- Learning semantic parsers without any annotated meaning representations
- Supervision in the form of binary feedback
  - Does the predicted meaning representation when executed produces the desired response for a given input sentence?
- Uses SVM with squared-hinge loss as base classifier
- Geoquery: 80% accuracy
  - Competitive compared to full supervision (86%)

Unambiguous Supervision for Learning Semantic Parsers

- The training data for semantic parsing consists of hundreds of natural language sentences *unambiguously* paired with their meaning representations

Unambiguous Supervision for Learning Semantic Parsers

- The training data for semantic parsing consists of hundreds of natural language sentences *unambiguously* paired with their meaning representations

Which rivers run through the states bordering Texas?

answer(traverse(next_to(stateid(‘texas’))))

What is the lowest point of the state with the largest area?

answer(lowest(place(loc(largest_one(area(state(all)))))))

What is the largest city in states that border California?

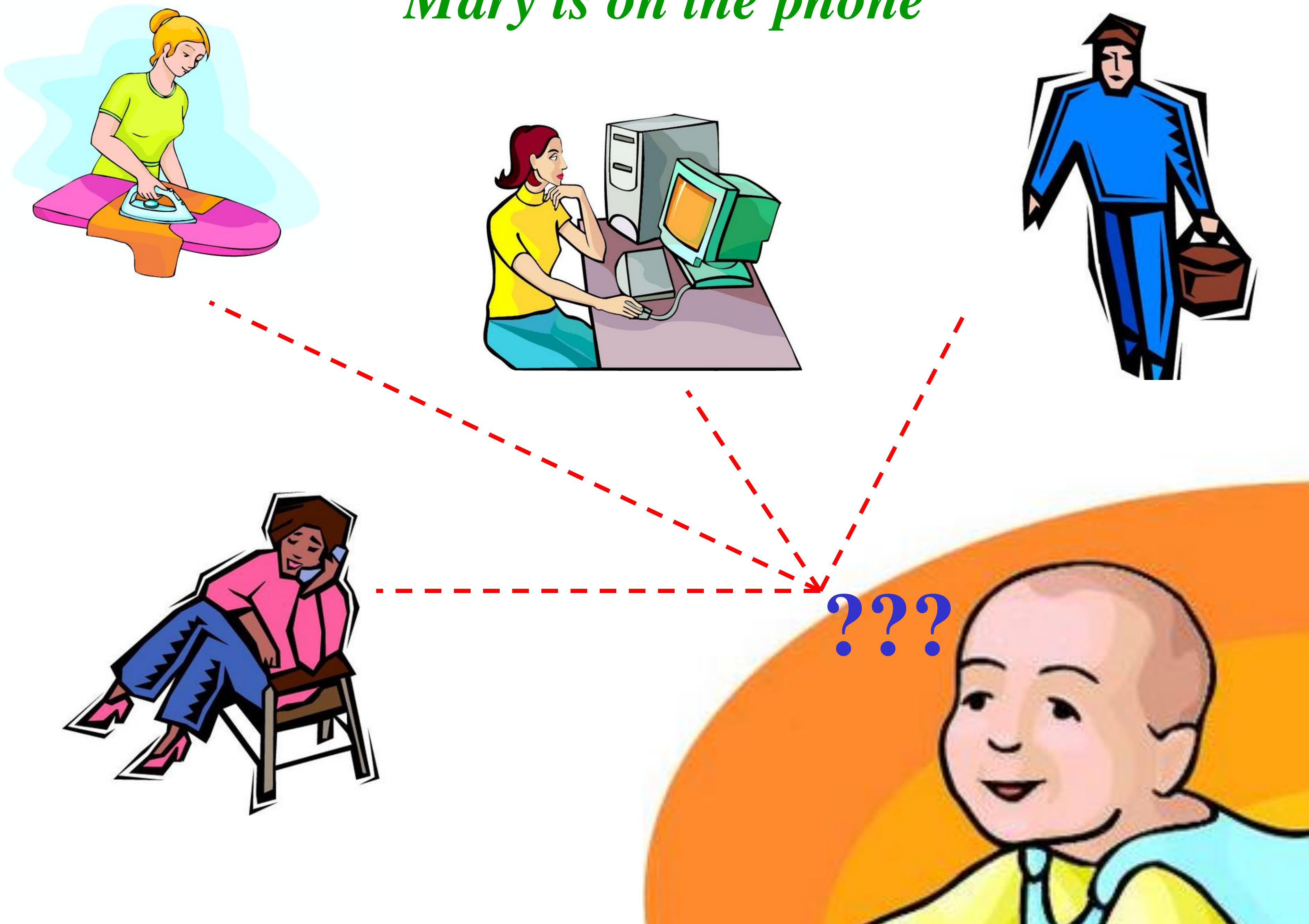
answer(largest(city(loc(next_to(stateid('california'))))))

……

Shortcomings of Unambiguous Supervision

- It requires considerable human effort to annotate each sentence with its correct meaning representation
- Does not model the type of supervision children receive when they are learning a language
  - Children are not taught meanings of individual sentences
  - They learn to identify the correct meaning of a sentence from several meanings possible in their perceptual context

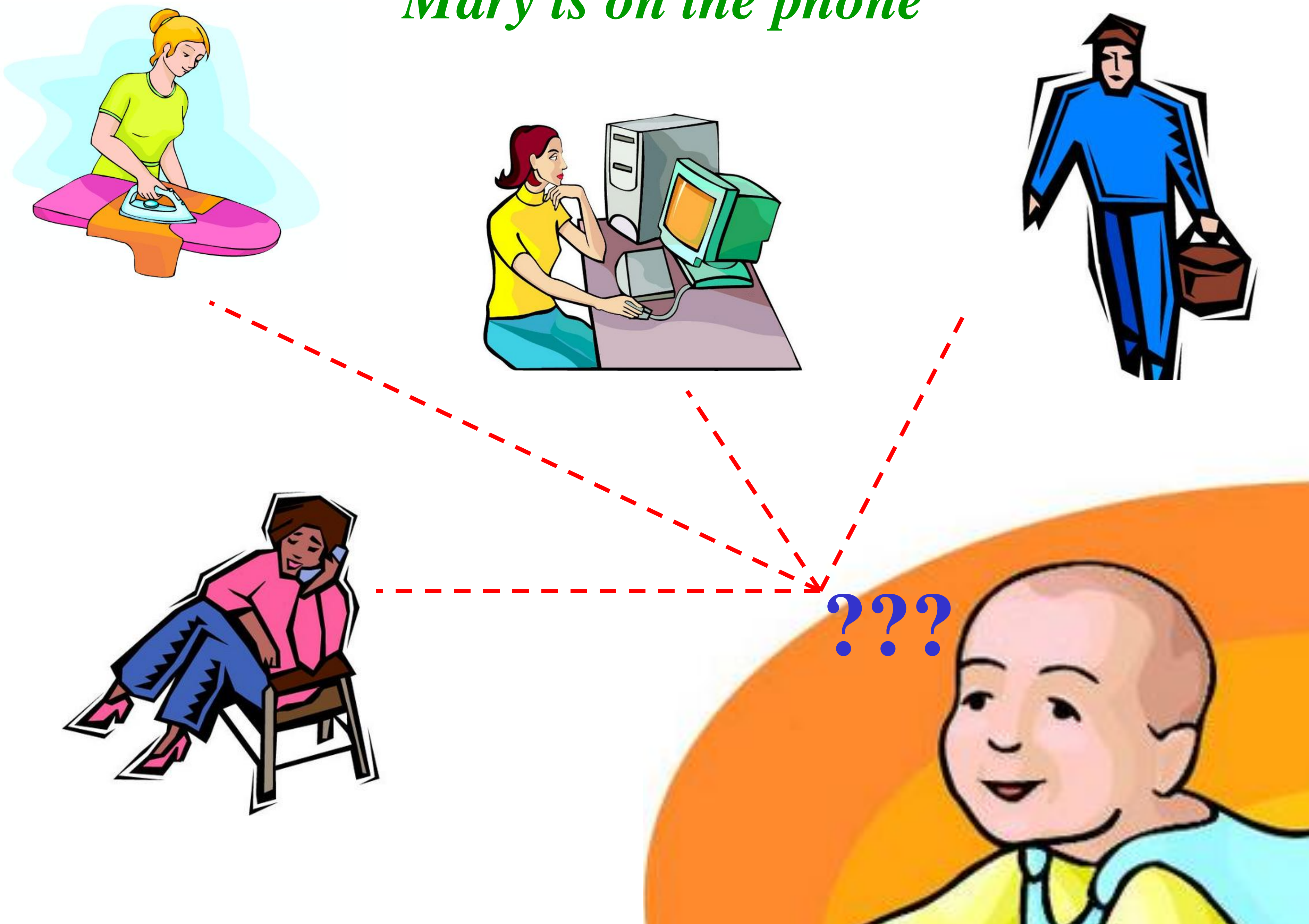
“Mary is on the phone”

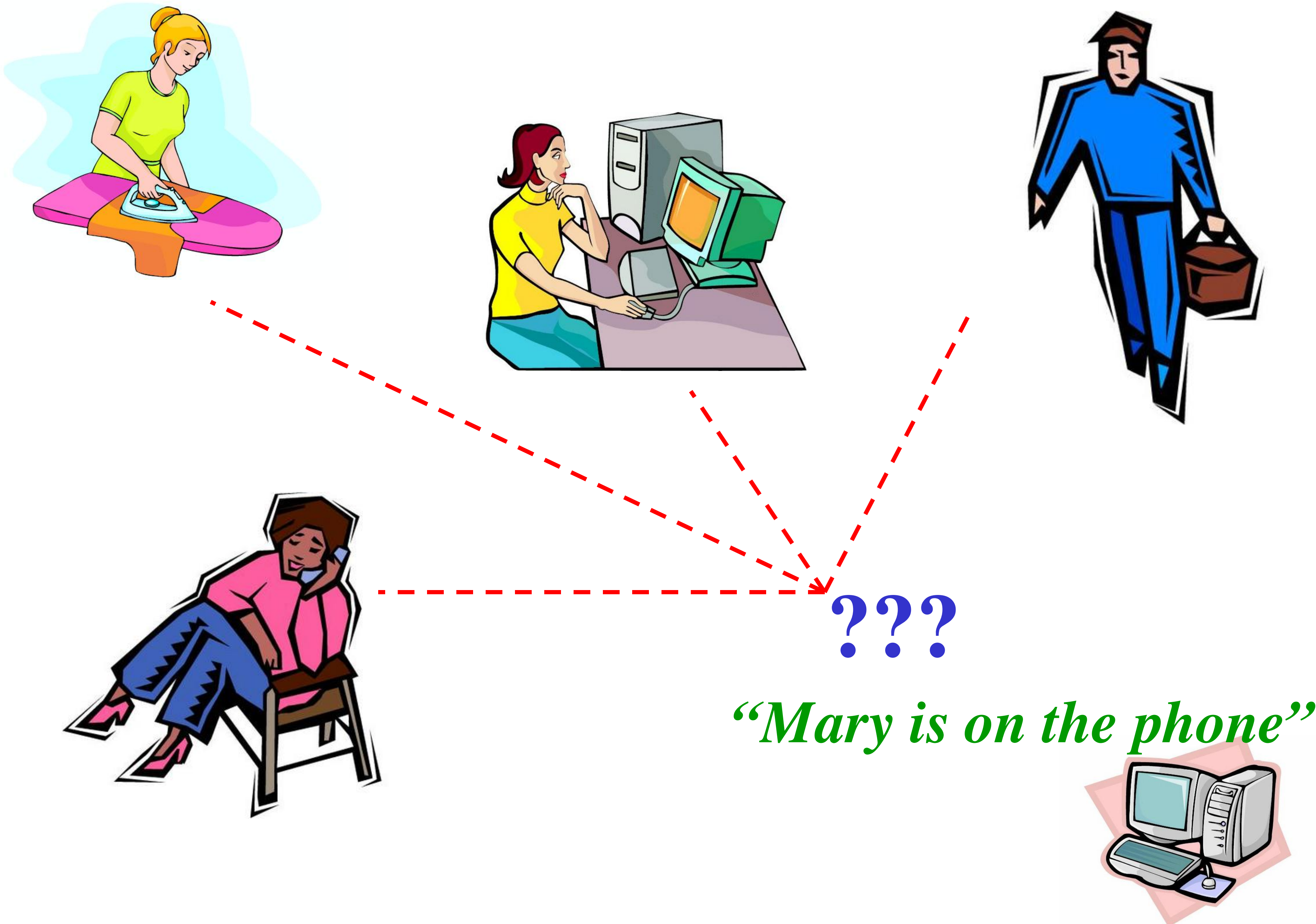


Ambiguous Supervision for Learning Semantic Parsers

- A computer system simultaneously exposed to perceptual contexts and natural language utterances should be able to learn the underlying language semantics
- We consider *ambiguous* training data of sentences associated with multiple potential meaning representations
  - Siskind (1996) uses this type “referentially uncertain” training data to learn meanings of words
  - We use ambiguous training data to learn meanings of sentences
- Capturing meaning representations from perceptual contexts is a difficult unsolved problem
  - Our system directly works with symbolic meaning representations

"Mary is on the phone"

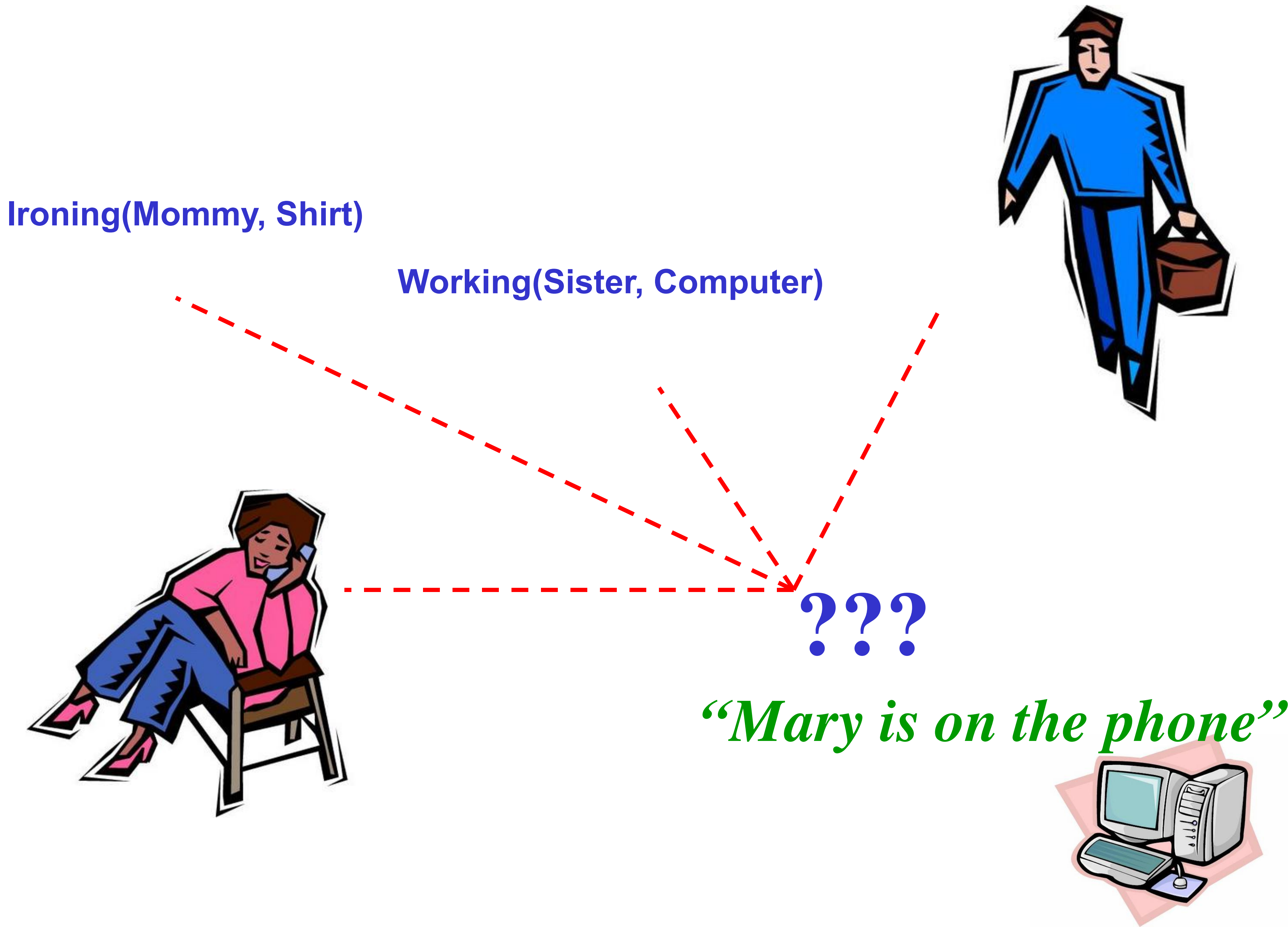



???
"Mary is on the phone"

Ironing(Mommy, Shirt)

???

"Mary is on the phone"


Ironing(Mommy, Shirt)
Working(Sister, Computer)
???
"Mary is on the phone"

Ironing(Mommy, Shirt)

Carrying(Daddy, Bag)

Working(Sister, Computer)

???

"Mary is on the phone"

Ambiguous Training Example

Ironing(Mommy, Shirt)

Carrying(Daddy, Bag)

Working(Sister, Computer)

Talking(Mary, Phone)

Sitting(Mary, Chair)

???

"Mary is on the phone"

Next Ambiguous Training Example

Ironing(Mommy, Shirt)

Working(Sister, Computer)

Talking(Mary, Phone)

Sitting(Mary, Chair)

???

"Mommy is ironing shirt"

Sample Ambiguous Corpus

Daisy gave the clock to the mouse.

Mommy saw that Mary gave the hammer to the dog.

The dog broke the box.

John gave the bag to the mouse.

The dog threw the ball.

gave(daisy, clock, mouse)

ate(mouse, orange)

ate(dog, apple)

saw(mother, gave(mary, dog, hammer))

broke(dog, box)

gave(woman, toy, mouse)

gave(john, bag, mouse)

threw(dog, ball)

runs(dog)

saw(john, walks(man, dog))

KRISPER: KRISP with **E**M-like **R**etraining
(Kate & Mooney, 2007b)

- Extension of KRISP that learns from *ambiguous* supervision
- Uses an iterative Expectation-Maximization-like method [Dempster et al. 1977] to gradually converge on a correct meaning for each sentence
- Successfully learns semantic parser with ambiguous supervision

References

J. Clarke, D. Goldwasser, M. Chang, D. Roth (2010). Driving Semantic Parsing from the World's Response. To appear in *Proc. of CoNLL-2010*, Uppsala, Sweden.

Rohit J. Kate and Raymond J. Mooney (2007a). Semi-supervised learning for semantic parsing using support vector machines. In *Proc. of NAACL-HLT 2007*, pp. 81-84, Rochester, NY, April 2007.

Rohit J. Kate and Raymond J. Mooney (2007b). Learning language semantics from ambiguous supervision. In *Proc. of AAAI-2007*, pp. 895-900, Vancouver, BC, July 2007.

Jeffrey M. Siskind (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition* 61(1):29-91.

Outline

1. Introduction to the task of semantic parsing
   a) Definition of the task
   b) Examples of application domains and meaning representation languages
   c) Distinctions from and relations to other NLP tasks
2. Semantic parsers
   a) Earlier hand-built systems
   b) Learning for semantic parsing
   c) Various forms of supervision

---

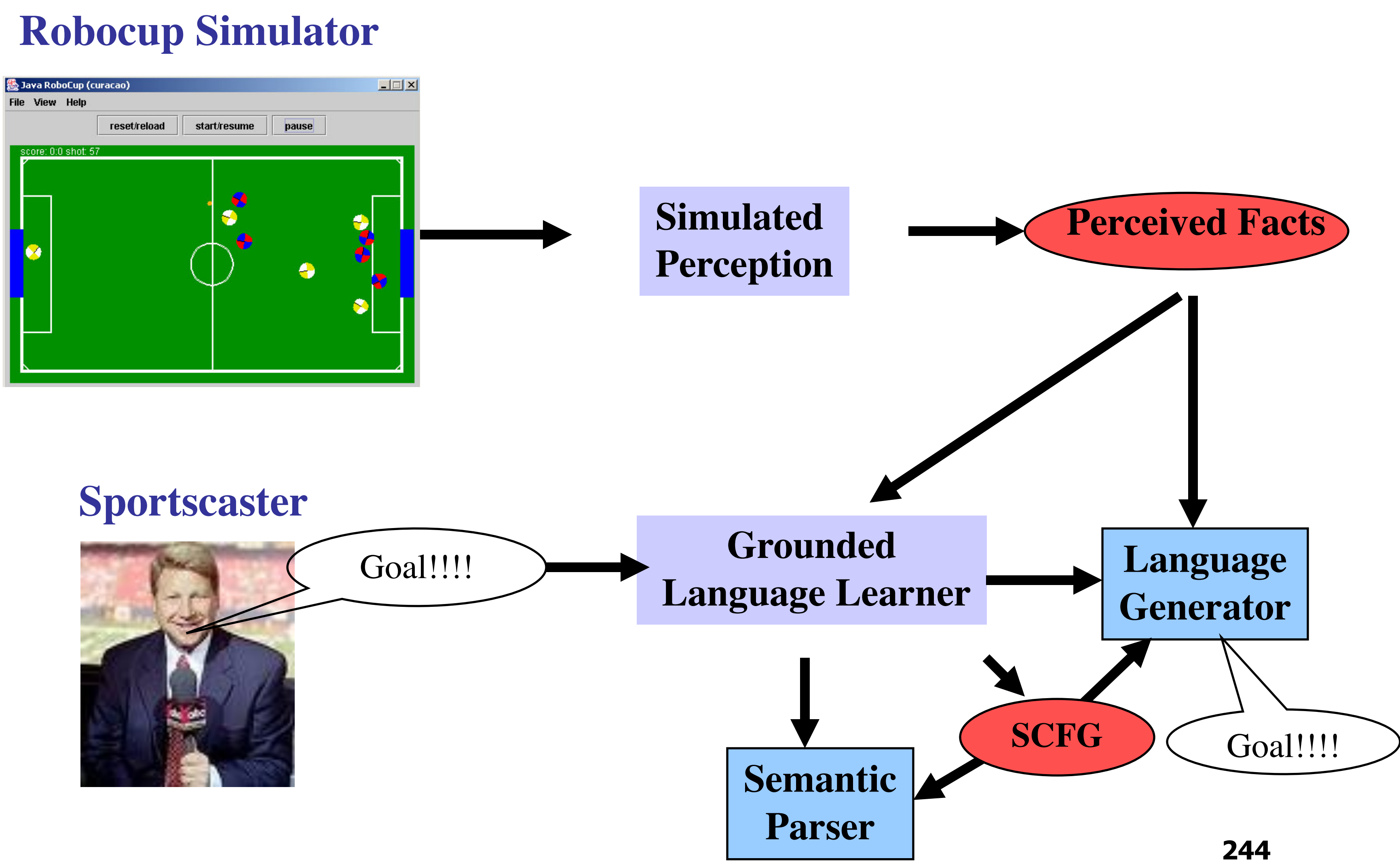
3. Semantic parsing beyond a sentence
   a) Learning language from perceptual contexts
   b) Using discourse contexts for semantic parsing
4. Research challenges and future directions
   a) Machine reading of documents: Connecting with knowledge representation
   b) Applying semantic parsing techniques to the Semantic Web
   c) Future research directions
5. Conclusions

Learning from Perceptual Context

Tractable Challenge Problem: Learning to Be a Sportscaster

- **Goal**: Learn from realistic data of natural language used in a representative context while avoiding difficult issues in computer perception (i.e. speech and vision).
- **Solution**: Learn from textually annotated traces of activity in a simulated environment.
- **Example**: Traces of games in the Robocup simulator paired with textual sportscaster commentary.

Grounded Language Learning in Robocup (Chen & Mooney, 2008)



Robocup Sportscaster Trace

Time

Natural Language Commentary

Purple goalie turns the ball over to Pink8

Purple team is very sloppy today

Pink8 passes the ball to Pink11

Pink11 looks around for a teammate

Pink11 makes a long pass to Pink8

Pink8 passes back to Pink11

Meaning Representation

badPass (Purple1, Pink8)

turnover (Purple1, Pink8)

kick (Pink8)

pass (Pink8, Pink11)

kick (Pink11)

kick (Pink11)

ballstopped

kick (Pink11)

pass (Pink11, Pink8)

kick (Pink8)

pass (Pink8, Pink11)

Robocup Sportscaster Trace

Natural Language Commentary

Meaning Representation

Time

badPass (Purple1, Pink8)

Purple goalie turns the ball over to Pink8

turnover (Purple1, Pink8)

kick (Pink8)

Purple team is very sloppy today

pass (Pink8, Pink11)

Pink8 passes the ball to Pink11

kick (Pink11)

Pink11 looks around for a teammate

kick (Pink11)

ballstopped

Pink11 makes a long pass to Pink8

kick (Pink11)

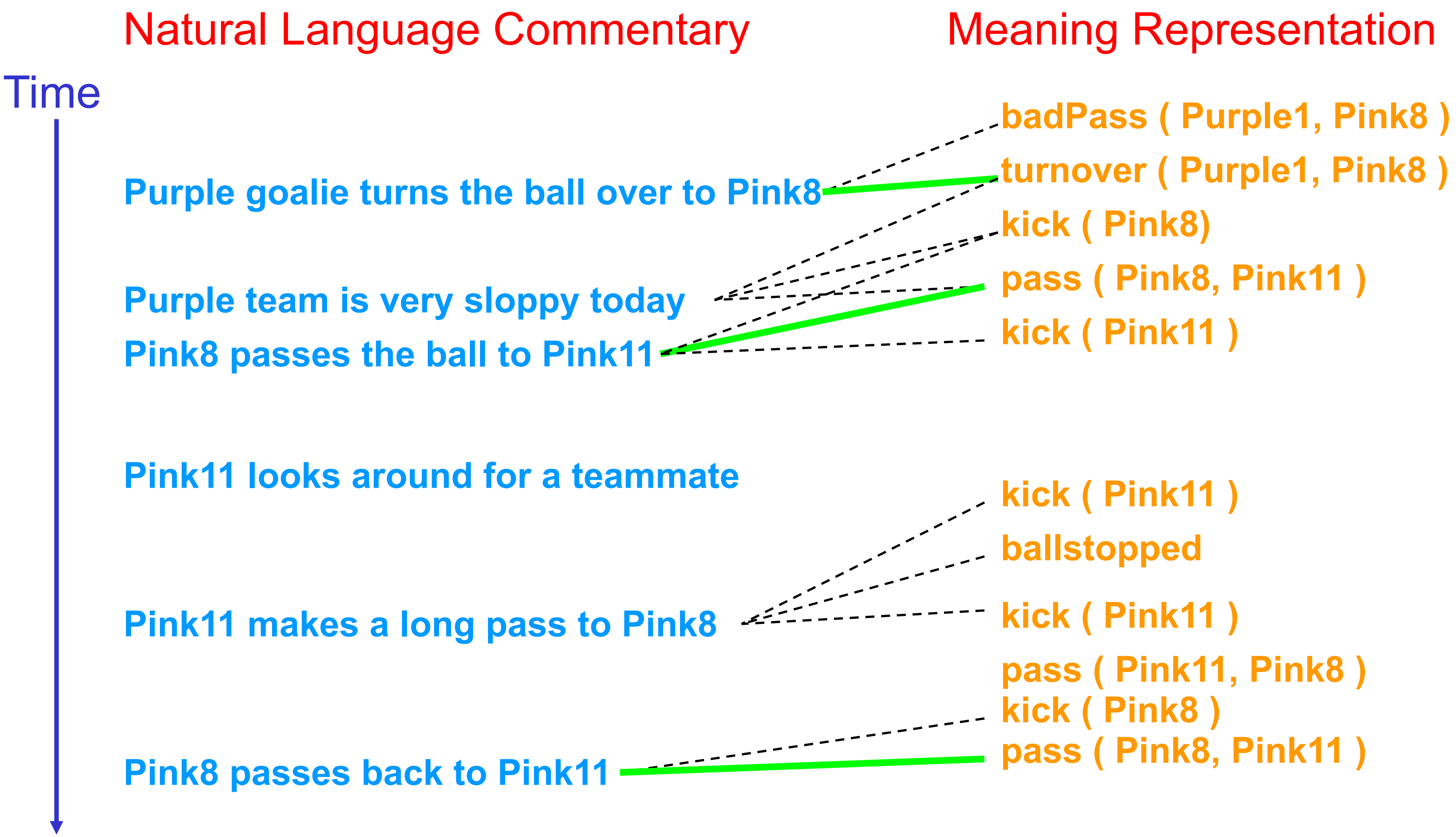
pass (Pink11, Pink8)

kick (Pink8)

Pink8 passes back to Pink11

pass (Pink8, Pink11)

Robocup Sportscaster Trace



Robocup Sportscaster Trace

Natural Language Commentary	Meaning Representation
	P6 (C1, C19)
Purple goalie turns the ball over to Pink8	P5 (C1, C19)
	P1(C19)
Purple team is very sloppy today	P2 (C19, C22)
Pink8 passes the ball to Pink11	P1 (C22)
Pink11 looks around for a teammate	
	P1 (C22)
	P0
Pink11 makes a long pass to Pink8	P1 (C22)
	P2 (C22, C19)
	P1 (C19)
Pink8 passes back to Pink11	P2 (C19, C22)

Time

Sportscasting Data

- Collected human textual commentary for the 4 Robocup championship games from 2001-2004.
  - Avg # events/game = 2,613
  - Avg # sentences/game = 509
- Each sentence matched to all events within previous 5 seconds.
  - Avg # MRs/sentence = 2.5 (min 1, max 12)

KRISPER-WASP

- First iteration of EM-like training produces very noisy training data (> 50% errors).
- KRISP is better than WASP at handling noisy training data.
  - SVM prevents overfitting.
  - String kernel allows partial matching.
- But KRISP does not support language generation.
- First train KRISPER just to determine the best NL→MR matchings.
- Then train WASP on the resulting unambiguously supervised data.

WASPER-GEN

- In KRISPER, the correct MR for each sentence is chosen based on maximizing the confidence of semantic parsing (NL→MR).
- Instead, WASPER-GEN determines the best matching based on **generation** (MR→NL).
- Score each potential NL/MR pair by using the currently trained WASP^{-1} generator.

Strategic Generation

- Generation requires not only knowing ***how*** to say something (***tactical generation***) but also ***what*** to say (***strategic generation***).
- For automated sportscasting, one must be able to effectively choose which events to describe.

Example of Strategic Generation

pass (purple7 , purple6)

ballstopped

kick (purple6)

pass (purple6 , purple2)

ballstopped

kick (purple2)

pass (purple2 , purple3)

kick (purple3)

badPass (purple3 , pink9)

turnover (purple3 , pink9)

Example of Strategic Generation

pass (purple7 , purple6)

ballstopped

kick (purple6)

pass (purple6 , purple2)

ballstopped

kick (purple2)

pass (purple2 , purple3)

kick (purple3)

badPass (purple3 , pink9)

turnover (purple3 , pink9)

Learning for Strategic Generation

- For each event type (e.g. pass, kick) estimate the probability that it is described by the sportscaster.
- Use Iterative Generation Strategy Learning (IGSL), a variant of EM, to estimate the probabilities from the ambiguous training data.

Human Evaluation (Quasi Turing Test)

- Asked 4 fluent English speakers to evaluate overall quality of sportscasts.
- Randomly picked a 2 minute segment from each of the 4 games.
- Each human judge evaluated 8 commented game clips, each of the 4 segments shown twice: commented once by a human and once by the machine when tested on that game (and trained on the 3 other games).
- The 8 clips presented to each judge were shown in random counter-balanced order.
- Judges were not told which ones were human or machine generated.

Human Evaluation Metrics

Score	English Fluency	Semantic Correctness	Sportscasting Ability
5	Flawless	Always	Excellent
4	Good	Usually	Good
3	Non-native	Sometimes	Average
2	Disfluent	Rarely	Bad
1	Gibberish	Never	Terrible

Results on Human Evaluation

Commentator	English Fluency	Semantic Correctness	Sportscasting Ability
Human	3.94	4.25	3.63
Machine (WASPER-GEN)	3.44	3.56	2.94
Difference	–0.5	–0.69	–0.69

References

David L. Chen and Raymond J. Mooney (2008). Learning to sportscast: A test of grounded language acquisition. In *Proc. of ICML-2008*, Helsinki, Finland, July 2008.

David L. Chen, Joohyun Kim, Raymond J. Mooney (2010). Training a multilingual sportscaster: Using perceptual context to learn language. *Journal of Artificial Intelligence Research* 37 (2010): 397-435.

Fleischman, M., & Roy, D. (2007). Situated models of meaning for sports video retrieval. In *Proc. of NAACL-HLT-2007,* Rochester, NY, April 2007.

Yu, C., & Ballard, D. H. (2004). On the integration of grounding language and learning objects. In *Proc. of AAAI-2004*, pp. 488-493.

Outline

1. Introduction to the task of semantic parsing
   a) Definition of the task
   b) Examples of application domains and meaning representation languages
   c) Distinctions from and relations to other NLP tasks
2. Semantic parsers
   a) Earlier hand-built systems
   b) Learning for semantic parsing
   c) Various forms of supervision
3. Semantic parsing beyond a sentence
   a) Learning language from perceptual contexts
   b) Using discourse contexts for semantic parsing
4. Research challenges and future directions
   a) Machine reading of documents: Connecting with knowledge representation
   b) Applying semantic parsing techniques to the Semantic Web
   c) Future research directions
5. Conclusions

Using Discourse Contexts for Semantic Parsing

Discourse Contexts

- May I see all the flights from Cleveland to Dallas?
- Which of those leave before noon?
- Show me the cheapest.
- What about afternoon?

Discourse Contexts

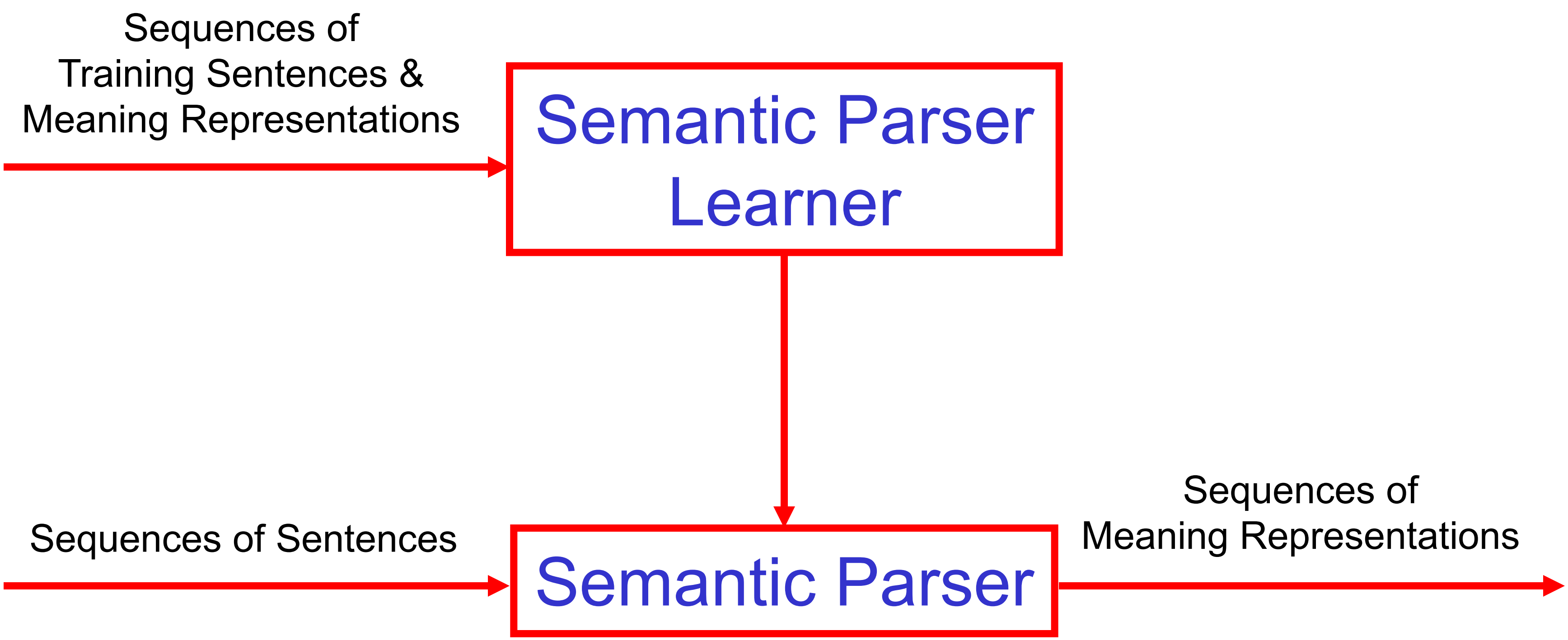
- May I see all the flights from Cleveland to Dallas?
  $\lambda x.flight(x) \wedge from(x, CLE) \wedge to(x, DFW)$
- Which of those leave before noon?
  $\lambda x.flight(x) \wedge from(x, CLE) \wedge to(x, DFW) \wedge morning(x)$
- Show me the cheapest.
  $argmin(\lambda x.flight(x) \wedge from(x, CLE) \wedge to(x, DFW) \wedge morning(x), \lambda y.cost(y))$
- What about afternoon?
  $argmin(\lambda x.flight(x) \wedge from(x, CLE) \wedge to(x, DFW) \wedge afternoon(x), \lambda y.cost(y))$

A Supervised Learning Problem

- Training examples: sequences of sentences and meaning representations

- May I see all the flights from Cleveland to Dallas?
  $\lambda x.flight(x) \wedge from(x, CLE) \wedge to(x, DFW)$
- Which of those leave before noon?
  $\lambda x.flight(x) \wedge from(x, CLE) \wedge to(x, DFW) \wedge morning(x)$
- Show me the cheapest.
  $argmin(\lambda x.flight(x) \wedge from(x, CLE) \wedge to(x, DFW) \wedge morning(x), \lambda y.cost(y))$
- What about afternoon?
  $argmin(\lambda x.flight(x) \wedge from(x, CLE) \wedge to(x, DFW) \wedge afternoon(x), \lambda y.cost(y))$

A Supervised Learning Problem



Example Analysis

- May I see all the flights from Cleveland to Dallas?
  $\boxed{\lambda x.flight(x) \wedge from(x, CLE) \wedge to(x, DFW)}$

- Which of those leave before noon?
  $\lambda x.\boxed{flight(x) \wedge from(x, CLE) \wedge to(x, DFW)} \wedge morning(x)$

Example Analysis

- May I see all the flights from Cleveland to Dallas?
  $\lambda x.flight(x) \wedge from(x, CLE) \wedge to(x, DFW)$
- Which of those leave before noon?
  $\boxed{\lambda x.flight(x) \wedge from(x, CLE) \wedge to(x, DFW) \wedge morning(x)}$
- Show me the cheapest.
  $argmin(\boxed{\lambda x.flight(x) \wedge from(x, CLE) \wedge to(x, DFW) \wedge morning(x)}, \lambda y.cost(y))$

Example Analysis

- May I see all the flights from Cleveland to Dallas?
  $\lambda x.\mathrm{flight}(x) \wedge \mathrm{from}(x, \mathrm{CLE}) \wedge \mathrm{to}(x, \mathrm{DFW})$
- Which of those leave before noon?
  $\lambda x.\mathrm{flight}(x) \wedge \mathrm{from}(x, \mathrm{CLE}) \wedge \mathrm{to}(x, \mathrm{DFW}) \wedge \mathrm{morning}(x)$
- Show me the cheapest.
  $\boxed{\mathrm{argmin}(\lambda x.\mathrm{flight}(x) \wedge \mathrm{from}(x, \mathrm{CLE}) \wedge \mathrm{to}(x, \mathrm{DFW}) \wedge \mathrm{morning}(x), \lambda y.\mathrm{cost}(y))}$
- What about afternoon?
  $\mathrm{argmin}(\lambda x.\mathrm{flight}(x) \wedge \mathrm{from}(x, \mathrm{CLE}) \wedge \mathrm{to}(x, \mathrm{DFW}) \wedge \boxed{\mathrm{afternoon}(x),} \lambda y.\mathrm{cost}(y))$

Context-Dependent Parsing using CCG

Zettlemoyer & Collins (2009)

- Given a sequence of sentences $x_1, \ldots, x_n$:
  - Each sentence x_i is parsed using CCG to form an intermediate, context-independent logical form y_i
  - Resolve all references in y_i
  - Optionally, perform an elaboration
  - Final logical form is z_i
  - Last two steps depend on the context $z_1, \ldots, z_{i-1}$
- A linear model for scoring derivations
- Treat $y_1, \ldots, y_n$ as hidden variables

Context-Independent Parsing

- Add referential lexical items

ones := N : $\lambda x.!f(x)$
it := NP : $!e$

Which of those leave before noon?
$\lambda x.!f(x) \wedge morning(x)$

Elliptical Expressions

- Add type-shifting rules:

$$A / B : g \quad \Rightarrow \quad A : g(\lambda x.!f(x))$$

$$A \backslash B : g \quad \Rightarrow \quad A : g(\lambda x.!f(x))$$

$$\frac{\dfrac{\text{the cheapest}}{NP / N \quad \lambda g.\mathrm{argmin}(g, \lambda y.\mathrm{cost}(y))}}{NP \quad \mathrm{argmin}(\lambda x.!f(x), \lambda y.\mathrm{cost}(y))}$$

Resolving References

- For each reference:
  - Select an expression from the context
  - Substitute into current logical form
- For each logical form in context, enumerate `e` and `<e,t>` type subexpressions

$\lambda x.flight(x) \wedge from(x, CLE) \wedge to(x, DFW)$

$\lambda x.flight(x) \wedge from(x, CLE) \wedge to(x, DFW) \wedge morning(x)$

CLE

DFW

$\lambda x.flight(x) \wedge from(x, CLE) \wedge to(x, DFW) \wedge morning(x)$

$\lambda x.flight(x) \wedge from(x, CLE) \wedge to(x, DFW)$

$\lambda x.flight(x) \wedge from(x, CLE) \wedge morning(x)$

$\lambda x.flight(x) \wedge to(x, DFW) \wedge morning(x)$

$\lambda x.from(x, CLE) \wedge to(x, DFW) \wedge morning(x)$

...

Elaboration

- Use current logical form to expand an embedded variable
- Can do deletions before elaboration

$\lambda x.flight(x) \wedge from(x, CLE) \wedge to(x, DFW)$

$\lambda x.flight(x) \wedge from(x, CLE) \wedge to(x, DFW) \wedge morning(x)$

$argmin(\lambda x.flight(x) \wedge from(x, CLE) \wedge to(x, DFW) \wedge morning(x), \lambda y.cost(y))$

what about afternoon

$\boxed{\lambda g}.argmin(\lambda x.flight(x) \wedge from(x, CLE) \wedge to(x, DFW) \wedge morning(x) \boxed{\wedge\, g(x),} \lambda y.cost(y))$

$\lambda x.afternoon(x)$

$\lambda g.argmin(\lambda x.flight(x) \wedge from(x, CLE) \wedge to(x, DFW) \wedge \cancel{morning(x) \wedge} g(x), \lambda y.cost(y))$

$argmin(\lambda x.flight(x) \wedge from(x, CLE) \wedge to(x, DFW) \wedge \boxed{afternoon(x),} \lambda y.cost(y))$

Scoring Derivations

- Linear model with the following features:
  - Parsing features: Zettlemoyer & Collins (2007)
  - Context features:
    - Distance indicators: Resolve references with expressions from recent sentences?
    - Copy indicators: Resolve references with expressions that contain a predicate p?
    - Deletion indicators: Override a predicate p_1 from the context expression with a predicate p_2 in the current expression?

Inference and Learning

- Use beam search to compute:
  - Best derivation
  - Best derivation with final logical form z
- Error-driven, perceptron-style parameter update

Results

- Test for correct logical forms
- For ATIS, 83.7% accuracy
  - Significantly higher than previous state-of-the-art (Miller et al., 1996)

References

S. Miller, D. Stallard, R. Bobrow, R. Schwartz (1996). A fully statistical approach to natural language interfaces. In *Proc. of ACL*, pp. 55-61. Santa Cruz, CA.

L. Zettlemoyer, M. Collins (2009). Learning context-dependent mappings from sentences to logical form. In *Proc. of ACL-IJCNLP*, pp. 976-984. Singapore.

Outline

1. Introduction to the task of semantic parsing
    a) Definition of the task
    b) Examples of application domains and meaning representation languages
    c) Distinctions from and relations to other NLP tasks
2. Semantic parsers
    a) Earlier hand-built systems
    b) Learning for semantic parsing
    c) Various forms of supervision
3. Semantic parsing beyond a sentence
    a) Learning language from perceptual contexts
    b) Using discourse contexts for semantic parsing
4. Research challenges and future directions
    a) Machine reading of documents: Connecting with knowledge representation
    b) Applying semantic parsing techniques to the Semantic Web
    c) Future research directions
5. Conclusions

Machine Reading of Documents: Connecting with Knowledge Representation

- A wealth of knowledge is available in natural language text
- Current question-answering systems can only detect answers from the text
- Knowledge representation and reasoning (KR&R) systems can answer complex questions that require inference and also provide explanations (Barker et al. 2004)
- Extract formal representation from text to augment KR&R systems

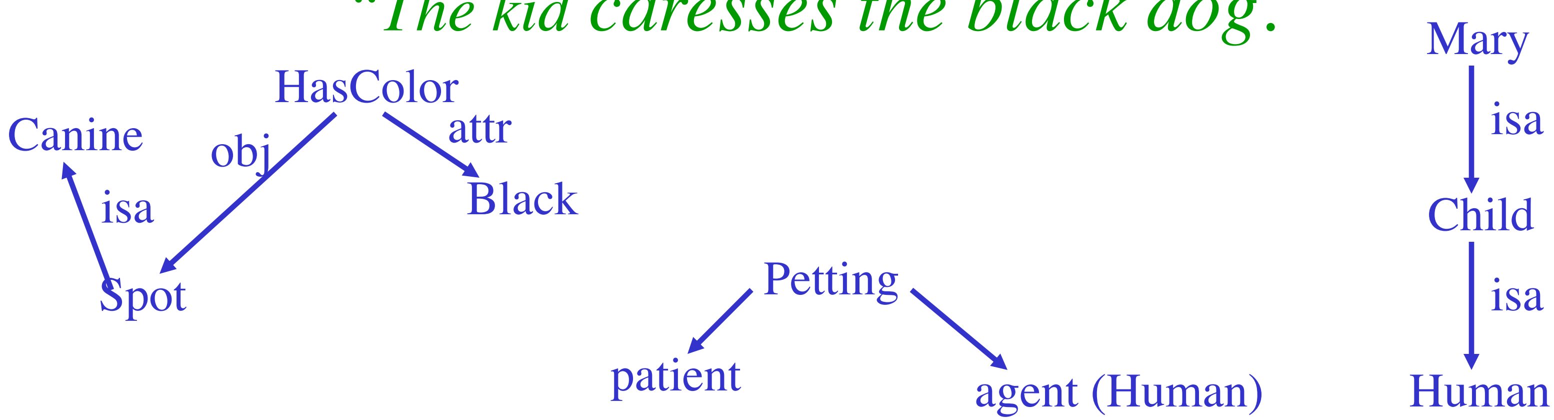
Machine Reading of Documents: Connecting with Knowledge Representation

- Semantic parsing gives the capability to interpret a sentence for a domain
- Extend it to interpret natural language passages for specific domains at a deeper semantic level than that of information extraction
- Use it for *machine reading* text passages to build knowledge bases (Barker et al. 2007)
  - Better semantic parsers can capture more knowledge from text

Machine Reading of Documents: Connecting with Knowledge Representation

- What will be the meaning representation language?
- The concepts and relations of an existing ontology for a domain could be treated as the formal meaning representation language
- Use semantic parsing and discourse context to build meaning representations
- Knowledge representation can in turn aid semantic parsing

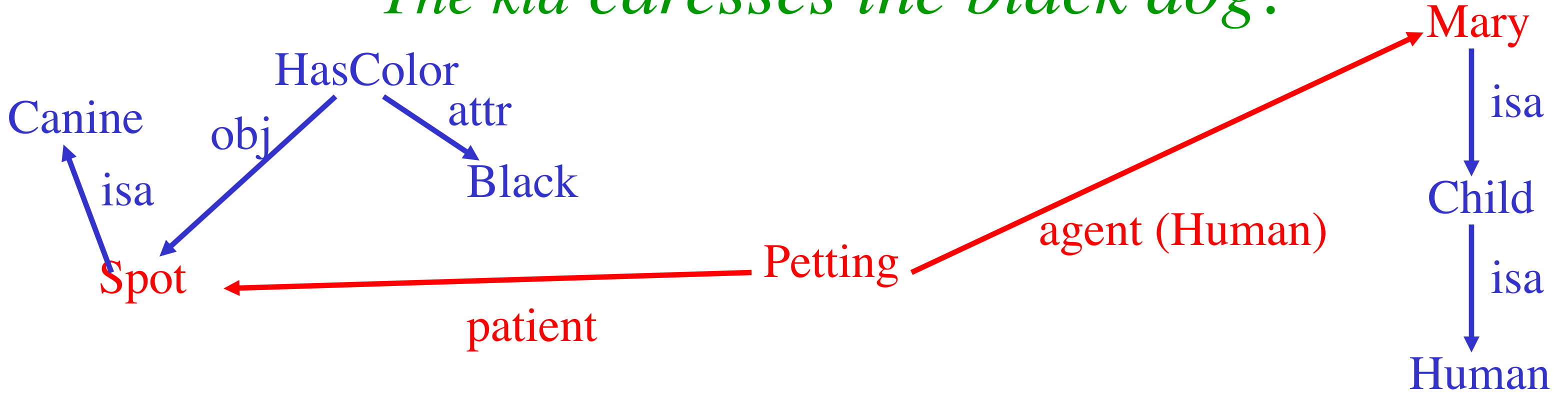
"The kid caresses the black dog."



Machine Reading of Documents: Connecting with Knowledge Representation

- What will be the meaning representation language?
- The concepts and relations of an existing ontology for a domain could be treated as the formal meaning representation language
- Use semantic parsing and discourse context to build meaning representations
- Knowledge representation can in turn aid semantic parsing

"The kid caresses the black dog."



Machine Reading of Documents: Connecting with Knowledge Representation

- How to get the training data?
- Use semi-supervised or unsupervised techniques
- Use domain adaptation techniques
- Gather “noisy” training examples by correlating text and existing knowledge bases or databases

The Semantic Web

- Extending the Web to allow the automation of information processing
- Presupposes large amounts of formally structured relational data, and the ability to make inferences over this data
- The fundamental decentralized information on the Web is text
- This ever growing body of text must form the basis of the Semantic Web if it is to succeed (Wilks & Brewster, 2009)

Annotating the Semantic Web

- Extract information from semi-structured text (Lerman et al., 2004)
- Extract relational tuples from free text (Lin & Pantel, 2001; Banko et al., 2007)

Natural Language Interface to Semantic Web

- Semantic web uses representations that require formal languages like RDF, OWL, SPARQL for querying
- Natural language interfaces offer familiar and convenient way for querying semantic web (Schwitter & Tilbrook, 2004; Kaufmann & Bernstein, 2007)
- Semantic parsing techniques could be used to query the semantic web in natural language

Unsupervised Semantic Parsing

Poon & Domingos (2009)

- Complete formal meaning of sentences
- Target predicates and constants are viewed as clusters of variations of the same meaning
  - borders, is next to, share the border with
  - Utah, the beehive state
- Clustering over forms that are built recursively
  - Forms used in composition with the same forms are encouraged to cluster together
- Uses Markov logic
  - A probabilistic extension of first-order logic

More Future Research Directions

- Interactive communication with computers in natural language
  - Semantic parsing is one-way
  - Combine with generation and discourse processing
- Reduce the requirement of training data
  - Utilize word similarity, paraphrases
  - Exploit active learning, domain adaptation etc.
- Develop corpora for new domains

References

Barker et al. (2004). A question-answering system for AP chemistry. In *Principles of Knowledge Representation and Reasoning 2004*.

Barker et al. (2007). Learning by reading: A prototype system, performance baseline and leassons learned. In *Proc. Of AAAI-2007*, Vancouver, BC, July 2007.

M. Banko, M. Cafarella, S. Soderland, M. Broadhead, O. Etzioni (2007). Open information extraction from the Web. In *Proc. of IJCAI*, pp. 2670-2676. Hyderabad, India.

K. Lerman, C. Gazen, S. Minton, C. Knoblock (2004). Populating the semantic web. In *Proc. of AAAI Workshop on Advances in Text Extraction and Mining*. San Jose, CA.

Esther Kaufmann and Abraham Bernstein (2007). How useful are natural language interfaces to the semantic web for casual end-users? In *Proc. of ISWC-2007*, pp. 281-294.

D. Lin, P. Pantel (2001). DIRT - Discovery of inference rules from text. In *Proc. of KDD*, pp. 323-328. San Francisco, CA.

References

H. Poon, P. Domingos (2009). Unsupervised semantic parsing. In *Proc. of EMNLP*, pp. 1-10. Singapore.

R. Schwitter, M. Tilbrook (2004). Controlled Natural Language meets the Semantic Web. In: A. Asudeh, C. Paris, S. Wan (eds.), *Proc. of the Australasian Language Technology Workshop*, pp. 55-62. Sydney, Australia.

Y. Wilks, C. Brewster (2009). *Natural language processing as a foundation of the semantic web*. Now Publishers Inc.

Conclusions

- Semantic parsing is an interesting and challenging task and is critical for developing computing systems that can understand and process natural language input
- The state-of-the-art in semantic parsing has been significantly advanced in recent years using a variety of statistical machine learning techniques and grammar formalisms
- Semantic parsing has also been extended to work beyond a sentence
- There are several promising future research directions which can have significant impact
  - Machine reading of documents
  - Applying to the Semantic Web
  - Interactive communication with computers in natural language

Thank You!

Questions??

Resources:

Geoquery and CLang data:

http://www.cs.utexas.edu/users/ml/nldata/geoquery.html

http://www.cs.utexas.edu/users/ml/nldata/clang.html

WASP and KRISP semantic parsers:

http://www.cs.utexas.edu/~ml/wasp/

http://www.cs.utexas.edu/~ml/krisp/

Geoquery Demo:

http://www.cs.utexas.edu/users/ml/geo.html