Type-driven Incremental Semantic Parsing with Polymorphism

Kai Zhao
Liang Huang
Dept. EECS, Oregon State University

Oregon State UNIVERSITY

From Language to Meaning

more informative

Information Extraction

Extracts information about a set of pre-specified relations and entities

Example:
Relation Extraction

is_a(Barack_Obama, US_President)

(Artzi et al., 2013)

From Language to Meaning

more informative

Broad-Coverage Semantics

Focuses on specific phenomena (e.g., verb-argument matching)

Example:
Summarization

(Artzi et al., 2013)

From Language to Meaning

more informative

Semantic Parsing

Extracts complete meaning representation

Example:
Database Query from Natural Language

Which states border Texas?

semantic parsing

$\lambda x.state(x) \wedge borders(x,texas)$

execution over database

LA, AR, OK, NM

(Artzi et al., 2013)

Semantic Parsing

- **fully supervised**
  - analogous to MT

Input

`What states border Texas?`

Output

$\lambda x.state(x) \wedge borders(x,texas)$

- **weakly supervised**
  - aka. parsing from Q/A pairs

Input

`What states border Texas?`

Output

`{LA,AR,OK,NM}`

Challenges

✓ Unknown Derivation

i) which parsing tree leads to the correct MR?

ii) treated as Latent Variable

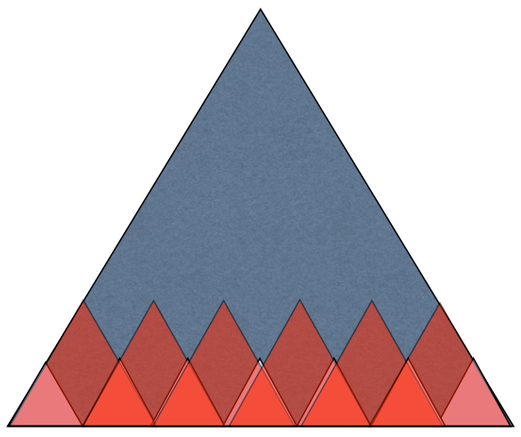
✓ Unknown Grammar

iii) i.e., the correspondences b/w English phrases & predicates

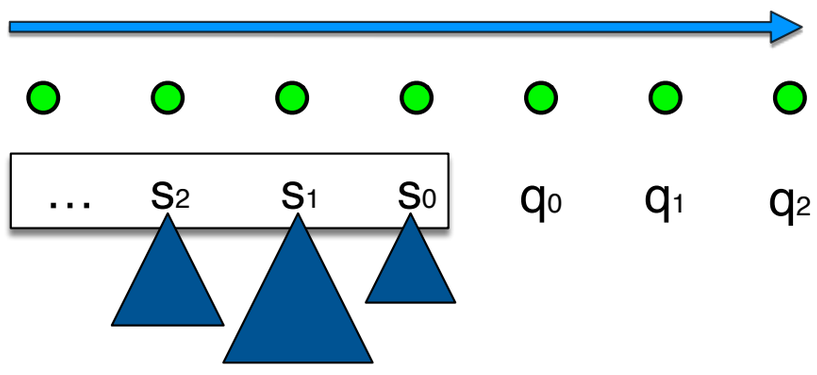
✓ Learn both derivation & grammar

From Bottom-Up to Incremental

- Conventional Parsing Algorithms:
  - CKY-based bottom-up parsing
  - cubic time



- Incremental Parsing
  - popular in constituent/ dependency parsing
  - linear time



Our Contributions

✓ **incremental** parsing (aka shift-reduce)

✓ abandon CCG, use **types** to guide parsing

- CCG: Combinatory Categorical Grammar
  A synchronous grammar b/w syntax & semantics
- type-driven: uses type checking to avoid unnecessary branching in searching

Incremental Parsing

✓ Type-driven Incremental Parsing

i) maintains **Stack** + **Queue**

ii) Actions:

- **SHIFT**: pops a word from queue, pushes its grounded semantic expr. onto stack
  use templates triggered by POS tags/patterns
- **REDUCE**: function application (type-driven)
- **SKIP**

A Running Example

INIT

Stack

ϕ

Queue

What	is	the	capital	of	the	largest	state	by	area
WP	VBZ	DT	NN	IN	DT	JJS	NN	IN	NN

A Running Example

SKIP: What

Stack

What

WP

ϕ

Queue

is	the	capital	of	the	largest	state	by	area
VBZ	DT	NN	IN	DT	JJS	NN	IN	NN

A Running Example

SKIP: is

Stack

What is

WP VBZ

ϕ

Queue

the	capital	of	the	largest	state	by	area
DT	NN	IN	DT	JJS	NN	IN	NN

A Running Example

SKIP: the

Stack

What is the

WP VBZ DT

ϕ

Queue

capital	of	the	largest	state	by	area
NN	IN	DT	JJS	NN	IN	NN

A Running Example

SHIFT: capital (NN)

Stack

What	is	the	capital
WP	VBZ	DT	NN

$\mathbf{capital}$
$\mathtt{e} \rightarrow \mathtt{e}$

Queue

of	the	largest	state	by	area
IN	DT	JJS	NN	IN	NN

POS Tag NN triggers templates:

$$\lambda P : \mathtt{e} \rightarrow \mathtt{e} \; . \; P$$

$$\lambda P : \mathtt{e} \rightarrow \mathtt{t} \; . \; P$$

$$\lambda P : \mathtt{e} \rightarrow \mathtt{i} \; . \; P$$

predicate $\mathbf{capital}$ $\mathtt{e} \rightarrow \mathtt{e}$ in database

New York	Albany
New Jersey	Trenton
Pennsylvania	Harrisburg
…	…

A Running Example

SKIP: of

Stack

What	is	the	capital	of
WP	VBZ	DT	NN	IN

capital

$e \rightarrow e$

Queue

the	largest	state	by	area
DT	JJS	NN	IN	NN

A Running Example

SKIP: the

Stack

What	is	the	capital	of	the
WP	VBZ	DT	NN	IN	DT

capital

e → e

Queue

largest	state	by	area
JJS	NN	IN	NN

A Running Example

SHIFT: largest (JJS)

Stack

What	is	the	capital	of	the	largest
WP	VBZ	DT	NN	IN	DT	JJS

capital $\mathtt{e} \to \mathtt{e}$

argmax $(\mathtt{e} \to \mathtt{t}) \to (\mathtt{e} \to \mathtt{i}) \to \mathtt{e}$

Queue

state	by	area
NN	IN	NN

POS Tag JJS triggers template:

$$\lambda P : (\mathtt{e} \to \mathtt{t}) \to (\mathtt{e} \to \mathtt{i}) \to \mathtt{e} \;.\; P$$

$$\underset{\mathtt{e}\to\mathtt{t}\;\mathtt{e}\to\mathtt{i}}{\mathbf{argmax}}\;\; f \;\; g \triangleq \underset{x:f(x)}{\arg\max}\;\; g(x)$$

A Running Example

TRY REDUCE?

Stack

What	is	the	capital	of	the	largest
WP	VBZ	DT	NN	IN	DT	JJS

capital $\mathtt{e} \to \mathtt{e}$

argmax $(\mathtt{e} \to \mathtt{t}) \to (\mathtt{e} \to \mathtt{i}) \to \mathtt{e}$

Queue

state	by	area
NN	IN	NN

type checking:

- left-reduce?
  - $\mathtt{e} \to \mathtt{e}$ does not match $\mathtt{e} \to \mathtt{t}$
- right reduce?
  - $(\mathtt{e} \to \mathtt{t}) \to (\mathtt{e} \to \mathtt{i}) \to \mathtt{e}$ does not match $\mathtt{e}$

A Running Example

SHIFT: state (NN)

Stack

What	is	the	capital	of	the	largest	state
WP	VBZ	DT	NN	IN	DT	JJS	NN

capital	**argmax**	**state**
$\mathtt{e} \to \mathtt{e}$	$(\mathtt{e} \to \mathtt{t}) \to (\mathtt{e} \to \mathtt{i}) \to \mathtt{e}$	$\mathtt{e} \to \mathtt{t}$

Queue

by	area
IN	NN

POS Tag NN triggers templates:

$$\lambda P : \mathtt{e} \to \mathtt{e} \;.\; P$$

$$\lambda P : \mathtt{e} \to \mathtt{t} \;.\; P$$

$$\lambda P : \mathtt{e} \to \mathtt{i} \;.\; P$$

predicate	**state** $\mathtt{e} \to \mathtt{t}$	in database
New York State		TRUE
New York City		FALSE
Pennsylvania		TRUE
…		…

A Running Example

TRY REDUCE?

Stack

What	is	the	capital	of	the	largest	state
WP	VBZ	DT	NN	IN	DT	JJS	NN

capital	**argmax**	**state**
$\mathtt{e} \to \mathtt{e}$	$(\mathtt{e} \to \mathtt{t}) \to (\mathtt{e} \to \mathtt{i}) \to \mathtt{e}$	$\mathtt{e} \to \mathtt{t}$

Queue

by	area
IN	NN

type checking:

- left-reduce?
  - $(\mathtt{e} \to \mathtt{t}) \to (\mathtt{e} \to \mathtt{i}) \to \mathtt{e}$ does not match $\mathtt{e}$
- right reduce?
  - $\mathtt{e} \to \mathtt{t}$ does match $\mathtt{e} \to \mathtt{t}$

A Running Example

REDUCE

Stack

What	is	the	capital	of	the	largest	state
WP	VBZ	DT	NN	IN	DT	JJS	NN

capital
$\mathtt{e} \rightarrow \mathtt{e}$

(argmax state)
$(\mathtt{e} \rightarrow \mathtt{i}) \rightarrow \mathtt{e}$

Queue

by	area
IN	NN

A Running Example

SKIP: by

Stack

What	is	the	capital	of	the	largest	state	by
WP	VBZ	DT	NN	IN	DT	JJS	NN	IN

Queue

area
NN

capital
$\mathtt{e} \rightarrow \mathtt{e}$

(argmax state)
$(\mathtt{e} \rightarrow \mathtt{i}) \rightarrow \mathtt{e}$

A Running Example

SHIFT: area (NN)

Stack

What	is	the	capital	of	the	largest	state	by	area
WP	VBZ	DT	NN	IN	DT	JJS	NN	IN	NN

$\mathbf{capital}$ $\mathtt{e} \rightarrow \mathtt{e}$

$\mathbf{(argmax\ \ state)}$ $(\mathtt{e} \rightarrow \mathtt{i}) \rightarrow \mathtt{e}$

$\mathbf{size}$ $\mathtt{e} \rightarrow \mathtt{i}$

Queue

ϕ

POS Tag NN triggers templates:

$$\lambda P : \mathtt{e} \rightarrow \mathtt{e} \ .\ P$$

$$\lambda P : \mathtt{e} \rightarrow \mathtt{t} \ .\ P$$

$$\lambda P : \mathtt{e} \rightarrow \mathtt{i} \ .\ P$$

predicate	$\mathbf{size}$ $\mathtt{e} \rightarrow \mathtt{i}$	in database
New York State		54,556
New York City		304.6
Pennsylvania		46,055
…		…

A Running Example

REDUCE

Stack

What	is	the	capital	of	the	largest	state	by	area
WP	VBZ	DT	NN	IN	DT	JJS	NN	IN	NN

capital $\quad$ **(argmax** $\quad$ **state** $\quad$ **size)**

e $\rightarrow$ e $\quad$ e

Queue

ϕ

A Running Example

REDUCE

Stack

What	is	the	capital	of	the	largest	state	by	area
WP	VBZ	DT	NN	IN	DT	JJS	NN	IN	NN

$$\textbf{(capital} \quad \textbf{(argmax} \quad \textbf{state} \quad \textbf{size))}$$

e

Queue

ϕ

Grounding Ambiguity

Who is the mayor of New York?

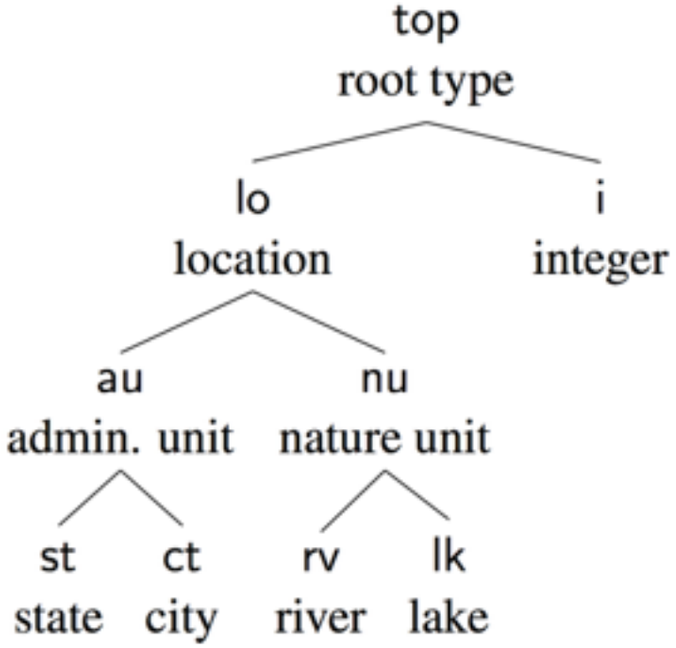
$$\begin{array}{ccc} & \textbf{new}\ \textbf{york}\ \textbf{city} & \\ & \mathsf{e} & \\ \textbf{mayor} & & \\ \mathsf{e} \rightarrow \mathsf{e} & \textbf{new}\ \textbf{york}\ \textbf{state} & \\ & \mathsf{e} & \end{array}$$

Who is the governor of New York?

$$\begin{array}{ccc} & \textbf{new}\ \textbf{york}\ \textbf{city} & \\ & \mathsf{e} & \\ \textbf{governor} & & \\ \mathsf{e} \rightarrow \mathsf{e} & \textbf{new}\ \textbf{york}\ \textbf{state} & \\ & \mathsf{e} & \end{array}$$

Subtyping

Type Hierarchy



Typed Function Application

$f : \mathtt{t}_1 \to \mathtt{t}_2$ takes argument $x : \mathtt{t}_3$

iff.

$\mathtt{t}_3$ is a subtype of $\mathtt{t}_1$

$\mathtt{t}_3 <: \mathtt{t}_1$

e.g.

population	**new york city**	✓
$\mathtt{au} \to \mathtt{i}$	$\mathtt{ct}$	
population	**new york state**	✓
$\mathtt{au} \to \mathtt{i}$	$\mathtt{st}$	
population	**hudson river**	✗
$\mathtt{au} \to \mathtt{i}$	$\mathtt{rv}$	

Subtyping

Who is the mayor of New York?

mayor
ct → pr

new york city
ct

~~**new york state**~~
st

Who is the governor of New York?

governor
st → pr

~~**new york city**~~
ct

new york state
st

Think again about argmax

$$\underset{\mathtt{e}\to\mathtt{t}\;\mathtt{e}\to\mathtt{i}}{\mathbf{argmax}}\;\; f\;\; g \stackrel{\Delta}{=} \underset{x:f(x)}{\arg\max}\;\; g(x)$$

$$\mathbf{argmax}$$
$$(\mathtt{e}\to\mathtt{t})\to(\mathtt{e}\to\mathtt{i})\to\mathtt{e}$$

✓ argmax is defined to accommodate the context

i) returns `ct` in largest city; `rv` in largest river

ii) can be defined as polymorphic type

$$\mathbf{argmax}$$
$$('\mathtt{a}\to\mathtt{t})\to('\mathtt{a}\to\mathtt{i})\to'\mathtt{a}$$

iv) type is bound at the parsing time on-the-fly

Our Contributions

- ✓ **incremental** parsing (aka shift-reduce)
- ✓ abandon CCG, use **type** to guide parsing
- ✓ subtyping **hierarchy**
- ✓ **polymorphic** functions

A Running Example

TRY REDUCE?

Stack

What	is	the	capital	of	the	largest
WP	VBZ	DT	NN	IN	DT	JJS

Queue

state	by	area
NN	IN	NN

capital $\mathtt{st} \to \mathtt{ct}$

argmax $(\mathtt{'a} \to \mathtt{t}) \to (\mathtt{'a} \to \mathtt{i}) \to \mathtt{'a}$

type checking:

- left-reduce?
  - $\mathtt{st} \to \mathtt{ct}$ does not match $\mathtt{'a} \to \mathtt{t}$
    - although $\mathtt{'a}$ can be bound to $\mathtt{st}$, $\mathtt{ct}$ does not match $\mathtt{t}$
- right reduce?
  - $(\mathtt{'a} \to \mathtt{t}) \to (\mathtt{'a} \to \mathtt{i}) \to \mathtt{'a}$ does not match $\mathtt{st}$

Running Example Revisited

TRY REDUCE?

Stack

What	is	the	capital	of	the	largest	state
WP	VBZ	DT	NN	IN	DT	JJS	NN

Queue

by	area
IN	NN

capital	**argmax**	**state**
$\mathtt{st} \to \mathtt{ct}$	$(\mathtt{'a} \to \mathtt{t}) \to (\mathtt{'a} \to \mathtt{i}) \to \mathtt{'a}$	$\mathtt{st} \to \mathtt{t}$

type checking:

- left-reduce?
  - $(\mathtt{'a} \to \mathtt{t}) \to (\mathtt{'a} \to \mathtt{i}) \to \mathtt{'a}$ does not match $\mathtt{st}$
- right reduce?
  - $\mathtt{st} \to \mathtt{t}$ can match $\mathtt{'a} \to \mathtt{t}$ as long as $\mathtt{'a}$ is bound to $\mathtt{st}$

Running Example Revisited

TRY REDUCE?

Stack

What	is	the	capital	of	the	largest	state
WP	VBZ	DT	NN	IN	DT	JJS	NN

Queue

by	area
IN	NN

capital	**argmax**	**state**
$\mathtt{st} \to \mathtt{ct}$	$(\mathtt{st} \to \mathtt{t}) \to (\mathtt{st} \to \mathtt{i}) \to \mathtt{st}$	$\mathtt{st} \to \mathtt{t}$

type checking:

- left-reduce?
  - $('\mathtt{a} \to \mathtt{t}) \to ('\mathtt{a} \to \mathtt{i}) \to '\mathtt{a}$ does not match $\mathtt{st}$
- right reduce?
  - $\mathtt{st} \to \mathtt{t}$ can match $'\mathtt{a} \to \mathtt{t}$ as long as $'\mathtt{a}$ is bound to $\mathtt{st}$

Running Example Revisited

REDUCE

Stack

What	is	the	capital	of	the	largest	state
WP	VBZ	DT	NN	IN	DT	JJS	NN

Queue

by	area
IN	NN

capital

$\texttt{st} \rightarrow \texttt{ct}$

(argmax state)

$(\texttt{st} \rightarrow \texttt{i}) \rightarrow \texttt{st}$

A Running Example

TRY REDUCE?

Stack

What	is	the	capital	of	the	largest	state	by	area
WP	VBZ	DT	NN	IN	DT	JJS	NN	IN	NN

Queue: ϕ

capital $\texttt{st} \rightarrow \texttt{ct}$

(argmax state) $(\texttt{st} \rightarrow \texttt{i}) \rightarrow \texttt{st}$

size $\texttt{lo} \rightarrow \texttt{i}$

type checking:

- right-reduce?
  - does $\texttt{lo} \rightarrow \texttt{i}$ match $\texttt{st} \rightarrow \texttt{i}$

YES, due to the contravariant rule in type theory

$$\frac{A <: B}{B \rightarrow C <: A \rightarrow C}$$

A Running Example

REDUCE

Stack

What	is	the	capital	of	the	largest	state	by	area
WP	VBZ	DT	NN	IN	DT	JJS	NN	IN	NN

capital **(argmax state size)**

st → ct

st

Queue

ϕ

A Running Example

REDUCE

Stack

What	is	the	capital	of	the	largest	state	by	area
WP	VBZ	DT	NN	IN	DT	JJS	NN	IN	NN

$$\textbf{(capital} \quad \textbf{(argmax} \quad \textbf{state} \quad \textbf{size))}$$

ct

Queue

ϕ

Learning

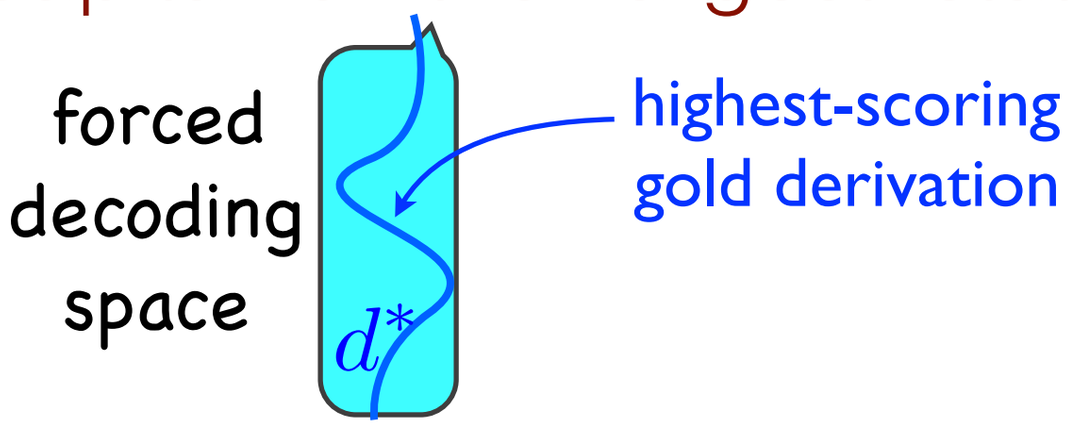
- ✓ Both derivation/Grammar are Unknown
- ✓ Spurious Ambiguity
  - i) Various derivations/groundings lead to the same logical form
- ✓ Latent Variable
  - ii) Structured Perceptron => Latent Variable Structured Perceptron

Learning

(size (argmin city population))

PENALIZE

What is the capital of the largest state by area ?



REWARD

(capital (argmax state size))

Learning

(size (argmin city population))

full search

$\hat{d}$

highest-scoring derivation

PENALIZE

beam

$$\mathbf{w} \leftarrow \mathbf{w} + \Phi(x, d^*) - \Phi(x, \hat{d})$$

reward correct

penalize wrong

Wh ?

space

d^*

REWARD

(capital (argmax state size))

Experiments

✓ Datasets

i) GeoQuery

- which state is dallas in?
- what are the populations of the states through which the mississippi run?
- what states border states that border states that border states that border texas?

ii) Jobs

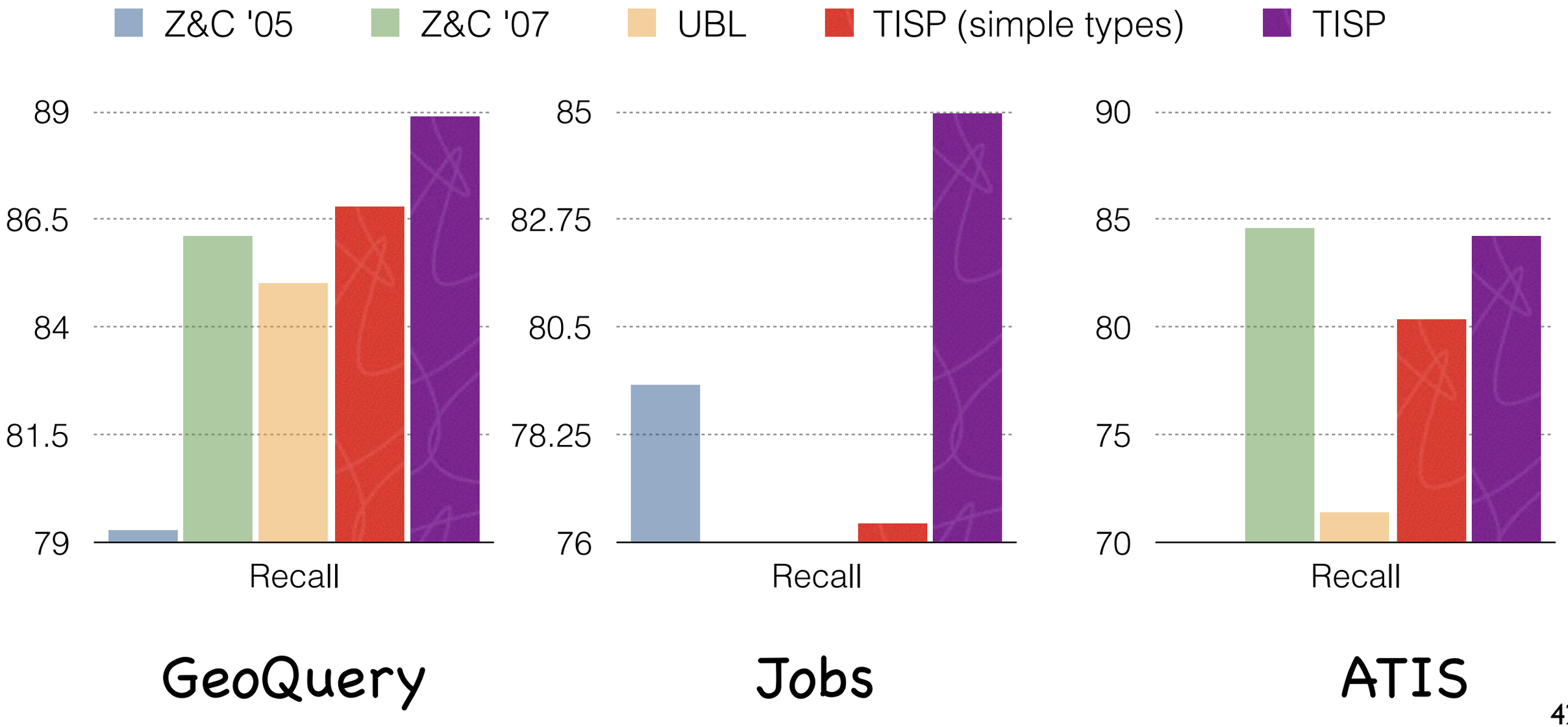
- are there any jobs using cpp with dell?
- are there any jobs in the us with the title verification engineer?

iii) ATIS

- show me the united flights from denver to baltimore
- what flights do you have in the morning of september twentieth on united airlines from pittsburgh to san francisco and a stopover in denver

Experiments

- ✓ High decoding speed; Linear in theory & practice
  - i) 0.5 sec/sentence
- ✓ recall (# correct parses / # sents)



GeoQuery Jobs ATIS

Conclusion

- ✓ Polymorphic typing guides the parsing
- ✓ Linear time incremental parsing
- ✓ Learning w/ Latent Variable Structured Perceptron
- ✓ Future Work:
  - Open Domain (Freebase)
  - Learning from Q/A pairs