Broad-coverage CCG Semantic Parsing with AMR

Yoav Artzi
Cornell University

Kenton Lee
University of Washington

Luke Zettlemoyer

Combinatory Categorial Grammar

Abstract Meaning Representation
Semantic Parsing

Show me all papers about semantic parsing

$$\lambda x.\text{paper}(x) \land \text{topic}(x, \text{SEMANTIC\_PARSING})$$

- **Less Supervision**
  - Answers
  - Demonstrations
  - Conversations

- **More Domains**
  - Databases
  - Large Knowledge-bases
  - Instructions
  - Web Tables
  - Time

- **Situated Parsing**
  - Spatial Observations
  - Linguistic Context
  - Database Content
Semantic Parsing

Show me all papers about semantic parsing

$$\lambda x. \text{paper}(x) \land \text{topic}(x, \text{SEMANTIC_PARSING})$$

- Less Supervision
- More Domains
- Situated Parsing
- Non-compositional Semantics
- Broad-coverage Grammar Induction
Pyongyang officials denied their involvement
Pyongyang officials denied their involvement
Pyongyang officials denied their involvement.
Pyongyang officials denied their involvement.
Pyongyang officials denied their involvement.
Pyongyang officials denied their involvement.
AMR and Combinatory Categorial Grammar

Great opportunity study CCG semantic parsing at scale

Challenges:

• Distant non-compositional dependencies
• Longer sentences
• Higher syntactic variability
Parsing Approach

• Use CCG to recover compositional parse structure

• Second stage to resolve non-compositional phenomena, such as co-reference resolution
Combinatory Categorial Grammar

**Category**

\[ S \backslash NP/NP : \lambda x.\lambda y.\lambda d.\text{deny-01}(d) \land \text{ARG0}(d, y) \land \text{ARG1}(d, x) \]

**Lexicon**

Assign category to words

**Combinators**

Unary and binary operators to combine categories

**Syntax**

**Semantics**
CCG

Entries from Lexicon

\[
\begin{align*}
\text{NP} & \quad \text{is} & \quad \text{fun} \\
\text{CCG} & \quad S \setminus NP/ADJ & \quad ADJ \\
\lambda f. \lambda x. f(x) & \quad \lambda x. \text{fun}(x) \\
\end{align*}
\]

Logical Form

Parse Steps

Combinators

Lexicon

Learned

CCG is fun

\[
\begin{align*}
S \setminus NP & \quad \lambda x. \text{fun}(x) \\
\text{fun}(\text{CCG}) & < \ \\
\end{align*}
\]
Pyongyang officials denied their involvement
AMR to Lambda Calculus

Pyongyang officials denied their involvement

\[ A_1(\lambda d.\text{deny-01}(d)) \land \\
\text{ARG0}(d, A_2(\lambda p.\text{person}(p)) \land \\
\text{ARG0-of}(p, A_3(\lambda h.\text{have-org-role-01}(h)) \land \\
\text{ARG1}(h, A_4(\lambda c.\text{city}(c)) \land \\
\text{name}(c, A_5(\lambda n.\text{name}(n) \land \text{op1}(n, \text{PYONGYANG})))))) \land \\
\text{ARG2}(h, A_6(\lambda o.\text{official}(o)))))) \land \\
\text{ARG1}(d, A_7(\lambda i.\text{involve-01}(i) \land \text{ARG1}(i, R(2))))))) \]
Pyongyang officials denied their involvement

\[ \mathcal{A}_1(\lambda d.\text{deny-01}(d)) \wedge \\
\text{ARG0}(d, \mathcal{A}_2(\lambda p.\text{person}(p)) \wedge \\
\text{ARG0-of}(p, \mathcal{A}_3(\lambda h.\text{have-org-role-91}(h)) \wedge \\
\text{ARG1}(h, \mathcal{A}_4(\lambda c.\text{city}(c)) \wedge \\
\text{name}(c, \mathcal{A}_5(\lambda n.\text{name}(n) \wedge \text{op1}(n, \text{PYONGYANG})))))) \wedge \\
\text{ARG2}(h, \mathcal{A}_6(\lambda o.\text{official}(o))))))) \wedge \\
\text{ARG1}(d, \mathcal{A}_7(\lambda i.\text{involve-01}(i) \wedge \text{ARG1}(i, \mathcal{R}(2))))))) \]
AMR Lambda Calculus: Relations

Pyongyang officials denied their involvement

\[ A_1(\lambda d.\text{deny-01}(d) \land \\
\text{ARG0}(d, A_2(\lambda p.\text{person}(p) \land \\
\text{ARG0-of}(p, A_3(\lambda h.\text{have-org-role-91}(h) \land \\
\text{ARG1}(h, A_4(\lambda c.\text{city}(c) \land \\
\text{name}(c, A_5(\lambda n.\text{name}(n) \land \text{op1}(n, \text{PYONGYANG})))))) \land \\
\text{ARG2}(h, A_6(\lambda o.\text{official}(o)))))) \land \\
\text{ARG1}(d, A_7(\lambda i.\text{involve-01}(i) \land \text{ARG1}(i, \mathcal{R}(2)))))) \]
Pyongyang officials denied their involvement

\[ A_1(\lambda d.\text{deny-01}(d)) \land \]
\[ \text{ARG0}(d, A_2(\lambda p.\text{person}(p)) \land \]
\[ \text{ARG0-of}(p, A_3(\lambda h.\text{have-org-role-91}(h)) \land \]
\[ \text{ARG1}(h, A_4(\lambda c.\text{city}(c)) \land \]
\[ \text{name}(c, A_5(\lambda n.\text{name}(n) \land \text{op1}(n, \text{PYONGYANG})))))) \land \]
\[ \text{ARG2}(h, A_6(\lambda o.\text{official}(o)))))) \land \]
\[ \text{ARG1}(d, A_7(\lambda i.\text{involve-01}(i) \land \text{ARG1}(i, R(2)))))) \]
AMR Lambda Calculus: References

Pyongyang officials denied their involvement

\( \mathcal{A}_1(\lambda d.\text{deny-01}(d)) \land 
\text{ARG0}(d, \mathcal{A}_2(\lambda p.\text{person}(p)) \land 
\text{ARG0-of}(p, \mathcal{A}_3(\lambda h.\text{have-org-role-91}(h)) \land 
\text{ARG1}(h, \mathcal{A}_4(\lambda c.\text{city}(c)) \land 
\text{name}(c, \mathcal{A}_5(\lambda n.\text{name}(n) \land \text{op1}(n, \text{PYONGYANG}))))))) \land 
\text{ARG2}(h, \mathcal{A}_6(\lambda o.\text{official}(o))))))) \land 
\text{ARG1}(d, \mathcal{A}_7(\lambda i.\text{involve-01}(i) \land \text{ARG1}(i, \mathcal{R}(2))))))) \)
Pyongyang officials denied their involvement
Pyongyang officials denied their involvement

Underspecified Logical Form

Model

CCG Parse

Constant Mapping
Pyongyang officials denied their involvement

\[ \mathcal{A}_1(\lambda d. \text{deny-01}(d) \land \text{ARG0}(d, \mathcal{A}_2(\lambda p. \text{person}(p)) \land \text{REL-of}(p, \mathcal{A}_3(\lambda h. \text{have-org-role-91}(h)) \land \text{ARG1}(h, \mathcal{A}_4(\lambda c. \text{city}(c)) \land \text{name}(c, \mathcal{A}_5(\lambda n. \text{name}(n) \land \text{op1}(n, \text{PYONGYANG})))))) \land \text{REL}(h, \mathcal{A}_6(\lambda o. \text{official}(o)))))) \land \text{ARG1}(d, \mathcal{A}_7(\lambda i. \text{involve-01}(i) \land \text{ARG1}(i, \mathcal{R}(\text{ID}))))))) \]
Pyongyang officials denied their involvement

Underspecified Logical Form

Model
Pyongyang officials denied their involvement

\[ A_1(\lambda d.\text{deny-01}(d)) \land \]
\[ \text{ARG0}(d, A_2(\lambda p.\text{person}(p))) \land \]
\[ \text{REL-of}(p, A_3(\lambda h.\text{have-org-role-91}(h)) \land \]
\[ \text{ARG1}(h, A_4(\lambda c.\text{city}(c)) \land \]
\[ \text{name}(c, A_5(\lambda n.\text{name}(n) \land \text{op1}(n, \text{PYONGYANG})))))) \land \]
\[ \text{REL}(h, A_6(\lambda o.\text{official}(o)))))) \land \]
\[ \text{ARG1}(d, A_7(\lambda i.\text{involve-01}(i) \land \text{ARG1}(i, \text{R(ID)))))))) \]
Pyongyang officials denied their involvement

Underspecified Logical Form

$\mathcal{R}(1) \quad \mathcal{R}(2) \quad \mathcal{R}(7)$
Pyongyang officials denied their involvement
Model Advantages

• Reason about non-compositional distant references, including:
  - Co-reference
  - Control structures (often compositional, but not distinguished)

• Defer certain compositional decisions from the difficult CCG parsing problem
Officials denied their involvement in Pyongyang.

Derivation:

\[
A_1(\lambda d.\text{deny-01}(d) \land \text{ARG0}(d, A_2(\lambda p.\text{person}(p) \land \text{ARG0-of}(p, A_3(\lambda h.\text{have-org-role-91}(h) \land \text{ARG1}(h, A_4(\lambda o.\text{official}(o)))))) \land \text{REL}(h, A_6(\lambda o.\text{official}(o))))) \land \text{ARG1}(d, A_7(\lambda i.\text{involve-01}(i) \land \text{ARG1}(i, R(ID))))))
\]
Log-linear Model

- Given a sentence $x \in \mathcal{X}$:
  - The probability of a logical form $z$ is:
    \[ p(z \mid x; \theta, \Lambda) = \sum_{d \in \mathcal{D}(z)} p(d \mid x; \theta, \Lambda) \]
  - The probability of a derivation $d \in \mathcal{D}$ is:
    \[ p(d \mid x; \theta, \Lambda) = \frac{e^{\theta \cdot \phi(x,d)}}{\sum_{d' \in \mathcal{D}} e^{\theta \cdot \phi(x,d')}} \]
Joint scoring

Inference

CCG Parse

CKY parsing

Factor graph

Constant Mapping

Pyongyang officials denied their involvement
Constant Mapping with a Factor Graph

Build a factor graph for each underspecified logical form

\[ A_1(\lambda d.\text{deny-01}(d)) \land \\
\text{ARG0}(d, A_2(\lambda p.\text{person}(p)) \land \\
\text{REL-of}(p, A_3(\lambda h.\text{have-org-role-91}(h)) \land \\
\text{ARG1}(h, A_4(\lambda c.\text{city}(c)) \land \\
\text{name}(c, A_5(\lambda n.\text{name}(n) \land \text{op}(n, \text{PYONGYANG}))) \land \\
\text{REL}(h, A_6(\lambda o.\text{official}(o)))))) \land \\
\text{ARG1}(d, A_7(\lambda i.\text{involve-01}(i)) \land \\
\text{ARG1}(i, R(\text{ID})))) \]
Each constant is a random variable

\[ A_1(\lambda d.\text{deny-01}(d)) \land \\
\quad \text{ARG0}(d, A_2(\lambda p.\text{person}(p)) \land \\
\quad \quad \text{REL-of}(p, A_3(\lambda h.\text{have-org-role-91}(h)) \land \\
\quad \quad \quad \text{ARG1}(h, A_4(\lambda c.\text{city}(c)) \land \\
\quad \quad \quad \quad \text{name}(c, A_5(\lambda n.\text{name}(n) \land \text{op}(n, \text{PYONGYANG})))))) \land \\
\quad \text{REL}(h, A_6(\lambda o.\text{official}(o)))))) \land \\
\quad \text{ARG1}(d, A_7(\lambda i.\text{involve-01}(i)) \land \\
\quad \quad \text{ARG1}(i, \mathcal{R}(\text{ID}))))} \]
Factor Graph

Potential mapping of placeholders defines assignments

\( A_1(\lambda d.\text{deny}-01(d)\land \) \\
\( \text{ARG0}(d, A_2(\lambda p.\text{person}(p)\land \) \\
\( \text{REL-of}(p, A_3(\lambda h.\text{have-org-role-91}(h)\land \) \\
\( \text{ARG1}(h, A_4(\lambda c.\text{city}(c)\land \) \\
\( \text{name}(c, A_5(\lambda n.\text{name}(n) \land \text{op}(n, \text{PYONGYANG}))))))\land \) \\
\( \text{REL}(h, A_6(\lambda o.\text{official}(o))))))\land \) \\
\( \text{ARG1}(d, A_7(\lambda i.\text{involve-01}(i)\land \) \\
\( \text{ARG1}(i, R(\text{ID}))))))))\)

unit, prep-with, frequency, prep-against, compared-to, employed-by, ARG2, ...

1, 2, 3, 4, 5, 6, 7
Features define factors to resolve underspecified constants

\[ \mathcal{A}_1(\lambda d.\text{deny-01}(d) \land \mathcal{A}_2(\lambda p.\text{person}(p) \land \mathcal{A}_3(\lambda h.\text{have-org-role-91}(h) \land \mathcal{A}_4(\lambda c.\text{city}(c) \land \mathcal{A}_5(\lambda n.\text{name}(n) \land \mathcal{A}_6(\lambda o.\text{official}(o))) \land \mathcal{A}_7(\lambda i.\text{involve-01}(i) \land \mathcal{A}_8(\lambda R(\text{ID})))))) \]

Selectional preference features to specify REL to one of 67 active relations

Features for resolving ID to 3
Approach

- Model:
  - Two-stage model for compositional semantics and non-compositional distant references

- Learning:
  - Lexicon induction
  - Parameter estimation
Learning Algorithm Sketch

For $T$ iterations:

- For each training sample:
  - Two-pass generation of new lexical entries
- Update the model lexicon

- For each mini-batch of size $M$
  - Compute gradient with early updates
  - Apply update with AdaGrad

Lexicon Induction

Parameter Estimation
Learning Algorithm Sketch

For T iterations:

• For each training sample:
  - Two-pass generation of new lexical entries

• Update the model lexicon

• For each mini-batch of size M
  - Compute gradient with early updates
  - Apply update with AdaGrad
Two-pass Lexical Generation

- Bottom-up: over-generate new entries and parse
- Top-down: recursive splitting to complete partial derivations
Bottom-up Pass

Underspecified Logical Form

Templates

Generated Entries

Pyongyang officials denied their involvement
Bottom-up Pass

CCG Parsing

Generated Entries

Pyongyang officials denied their involvement

Underspecified logical form
Bottom-up Pass

Select lexical entries from max **scoring** correct derivation

Generated Entries

Pyongyang officials denied their involvement

Underspecified logical form
Common Failure

- High syntactic variation
- Missing templates
- No complete correct derivation created

Need to learn new templates

Generated Entries

Pyongyang officials denied their involvement
Splitting CCG Categories

- Introduced by Kwiatkowski et al. 2010
- Approximately reverses CCG parsing operations
- Explore new syntactic structures, learn new templates
Splitting CCG Categories

Given a CCG category $C : h$:

1. Split logical form $h$ to $f$ and $g$ s.t.:

   $f(g) = h$ or $\lambda x. f(g(x)) = h$

$NP_{[nb]} : \lambda i. \text{involve-01}(i) \wedge \text{ARG1}(i, \mathcal{R}(ID))$

$\lambda f. \lambda i. f(i) \wedge \text{ARG1}(i, \mathcal{R}(ID))$

$\lambda i. \text{involve-01}(i)$

$\mathcal{R}(ID)$

$\lambda x. \lambda i. \text{involve-01}(i) \wedge \text{ARG1}(i, x)$
Splitting CCG Categories

Given a CCG category $C : h$:

1. Split logical form $h$ to $f$ and $g$ s.t.:
   \[ f(g) = h \text{ or } \lambda x. f(g(x)) = h \]

2. Infer syntax from logical form type

\[
\begin{align*}
\text{NP}_{[nb]} & : \lambda i.\text{involve-01}(i) \land \\
\text{ARG1}(i, \mathcal{R}(\text{ID})) \\
\text{NP}_{[pl]} & : \mathcal{R}(\text{ID}) \\
\text{NP}_{[nb]} \setminus \text{NP} & : \lambda x.\lambda i.\text{involve-01}(i) \land \text{ARG1}(i, x)
\end{align*}
\]
Top-down Pass

• Given a packed chart without a correct parse

Generated Entries

Pyongyang officials denied their involvement
Top-down Pass

• Starting from **correct logical form**
• **Recursively** split to create a complete tree

Generated Entries

Pyongyang officials denied their involvement
Top-down Pass

- Each split combines a **new category** with an existing partial derivation

![Diagram](image.png)

- Generated Entries: Pyongyang officials denied their involvement

- Underspecified logical form
Splitting for CCG Induction

- Kwiatkowski et al. 2010:
  - No restriction on result categories
  - Applied up to depth one

- Our approach:
  - Combined with bottom-up template approach
  - Must connect to an existing partial derivation
  - Applied recursively
Learning Algorithm Sketch

For T iterations:

- For each training sample:
  - Two-pass generation of new lexical entries
- Update the model lexicon
- For each mini-batch of size M
  - Compute gradient with early updates
  - Apply update with AdaGrad
Gradient Computation

• If a correct derivation exists:
  - Compute gradient with inside-outside
  - Re-normalize with constant mapping features
Common Failure

• No correct derivation exists, ~40% of training data

• Previous work assumed that all (or at least most) corpus can be parsed

• Instead: early updates

Pyongyang officials denied their
Early Updates

• Collins and Roark (2004):
  - Given fully labeled parse trees
  - Update with partial derivations

• Challenge: derivation is latent
Early Update with Latent Structures

- Extract sub-expression from underspecified logical form

- For each sub-expression:
  - Identify largest max-scoring partial derivation
  - Compute gradient

Pyongyang officials denied their
Early Update with Latent Structures

Underspecified Logical Form

\[ \mathcal{A}_2(\lambda p. \text{person}(p)) \land \\
   \text{REL-of}(p, \mathcal{A}_3(\lambda h. \text{have-org-role-91}(h)) \land \\
   \text{ARG1}(h, \mathcal{A}_4(\lambda c. \text{city}(c)) \land \\
   \text{name}(c, \mathcal{A}_5(\lambda n. \text{name}(n) \land \text{op}(n, \text{PYONGYANG})))) \land \\
   \text{REL}(h, \mathcal{A}_6(\lambda o. \text{official}(o)))) \]

\[ \mathcal{A}_4(\lambda c. \text{city}(c)) \land \\
   \text{name}(c, \mathcal{A}_5(\lambda n. \text{name}(n) \land \text{op}(n, \text{PYONGYANG})))) \]

Pyongyang officials denied their
Early Update with Latent Structures

Underspecified Logical Form

\[ A_2(\lambda p. \text{person}(p)) \land \\
\text{REL-of}(p, A_3(\lambda h. \text{have-org-role-91}(h)) \land \\
\text{ARG1}(h, A_4(\lambda c. \text{city}(c)) \land \\
\text{name}(c, A_5(\lambda n. \text{name}(n) \land \text{op}(n, \text{PYONGYANG})))))) \\
\text{REL}(h, A_6(\lambda o. \text{official}(o)))))) \]

\[ A_4(\lambda c. \text{city}(c)) \land \\
\text{name}(c, A_5(\lambda n. \text{name}(n) \land \text{op}(n, \text{PYONGYANG})))) \]
## Related Work

<table>
<thead>
<tr>
<th>Area</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCG Semantic Parsing</td>
<td>Zettlemoyer and Collins 2005, 2007;</td>
</tr>
<tr>
<td></td>
<td>Kwiatkowski et al. 2010, 2011;</td>
</tr>
<tr>
<td></td>
<td>Artzi and Zettlemoyer 2013</td>
</tr>
<tr>
<td>Skolem Terms for CCG</td>
<td>Steedman 2011</td>
</tr>
<tr>
<td>AMR Evaluation</td>
<td>Cai and Knight 2013</td>
</tr>
<tr>
<td>Graph-based Parsing for AMR</td>
<td>Flanigan et al. 2014</td>
</tr>
<tr>
<td>Dependency Structure Transformation for AMR</td>
<td>Wang et al. 2015a, 2015b</td>
</tr>
<tr>
<td>Syntax-based MT for AMR</td>
<td>Pust et al. 2015</td>
</tr>
<tr>
<td>Rule-based Parsing for AMR</td>
<td>Vanderwende et al. 2015</td>
</tr>
<tr>
<td>AMR Applications</td>
<td>Pan et al. 2015; Lin et al. 2015</td>
</tr>
</tbody>
</table>
Experimental Setup

• AMR Bank release 1.0, proxy report portion

• Evaluation metric: SMATCH [Cai and Knight 2013]

• Features: lexical features, parsing operations, parsing attachment, selectional preferences, control structures

• Seed lexicon and templates:
  - 50 annotated sentences
  - Heuristic alignment from JAMR [Flanigan et al. 2014]
Without early updates we fail to learn effectively from much of the data.

Poor performance without heuristics demonstrates need for future work.

- Full system
- w/o unrestricted lexical generation
- w/o early updates
- w/o surface-form similarity
Results

SMATCH F1

- JAMR (fixed)
- Werling et al. 2015
- Pust et al. 2015
- Our Approach
- Wang et al. 2015b

- AMR is getting a lot of attention! … and will: SemEval 2016
- Using solutions sub-problem solution is a promising complimentary direction
Contributions

• Joint model for compositional and non-compositional semantics

• Scalable CCG induction for semantic parsing

• First CCG approach to AMR

• Code and models available in Cornell SPF: http://yoavarzti.com/spf
[fin]