Semantics for Semantic Parsing

Mark Steedman (*with* Mike Lewis, Siva Reddy, and Mirella Lapata)

26 June 2014

School of informatics

Semantic Parsing: The First Ten Years

- The term "Semantic Parsing" refers to two distinct programs:
  - Parsing directly coupled with compositional assembly of meaning representation or "logical form";
  - More recently, the induction of such parsers from data consisting of string-meaning pairs.

- I'll distinguish the latter as "semantic parser induction".

- I'm going to argue that there is still life in the older enterprise.

Outline

- I: Supervised Semantic Parser Induction
- II: Semisupervised Semantic Parser Induction with and without QA pairs
- III: Learning the Hidden Language of Logical Form
- IV: Semantics for Semantic Parsers

I: Supervised Semantic Parser Induction

- Thompson and Mooney (2003); Zettlemoyer and Collins (2005, 2007); Wong and Mooney (2007); Lu *et al.* (2008); Kwiatkowski *et al.* (2010, 2011); Börschinger *et al.* (2011) generalize the problem of inducing parsers from language-specific treebanks like WSJ to that of inducing parsers from paired sentences and unaligned language-independent logical forms.
  - The sentences can be in any language.
  - The logical forms might be database queries, dependency graphs, λ-terms, robot action primitives and PDDL state descriptions, etc.
- This is the way the child learns language, *pace* Montague 1970 (Kwiatkowski *et al.* 2012)
- However, the approach suffers from an acute shortage of suitable datasets.

II: Semisupervised Semantic Parsing

- Question-answer pairs are abundantly available for large databases. So, learn from them.
- Clarke *et al.* (2010); Liang *et al.* (2011); Cai and Yates (2013a,b); Kwiatkowski *et al.* (2013); Berant *et al.* (2013)
- "Given my dataset, to what questions is 42 the answer?"
- Not that many—very few with the same content words

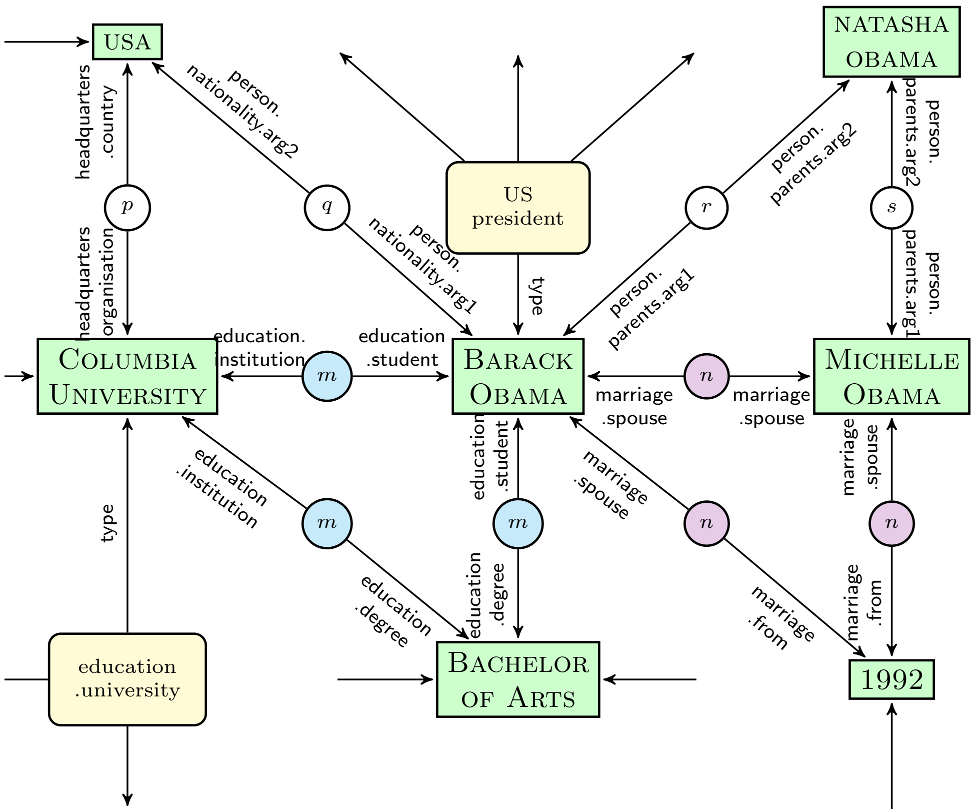
Semantic Parsing with Freebase without QA pairs

- Reddy (2014):
  - Rather than inducing a parser from questions and answers. . .
  - Take a parser that already builds logical forms and learn the relation between those logical forms and the knowledge graph,

- Specifically:
  - First turn the logical forms into graphs of the same type as the knowledge graph
  - Then learn the mapping between the elements of the semantic and knowledge-base graphs.

The Knowledge Graph

- Freebase is what used to be called a Semantic Net
- Cliques represent facts.
- Clique q represents the fact that *Obama's nationality is American*
- Clique m represents the fact that *Obama did his BA at Columbia*



Parsing to Logical Form using CCG

- Cameron directed *Titanic* in 1997.

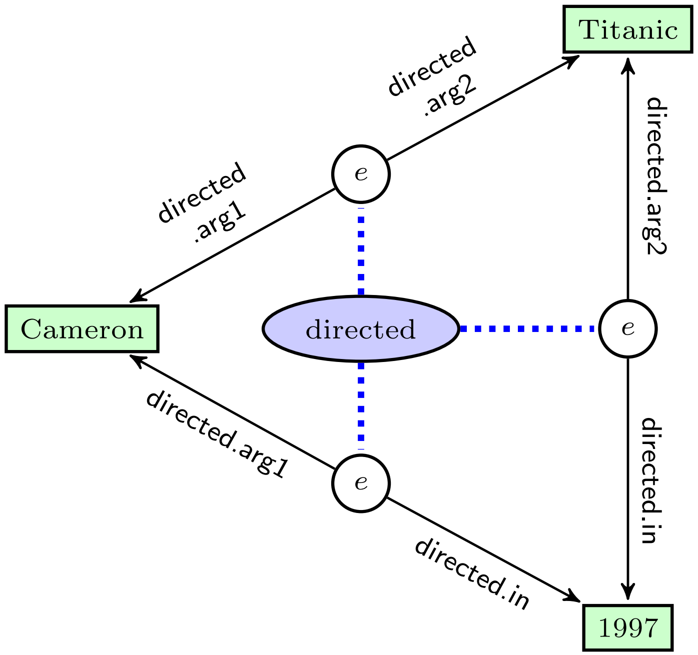
$$
\begin{array}{ccccc}
\text{Cameron} & \text{directed} & \textit{Titanic} & \text{in} & \text{1997} \\
\hline
NP & S\backslash NP/PP_{in}/NP & NP & PP_{in}/NP & NP \\
cameron & \lambda w \lambda x \lambda y.directed.arg1(E,y) \wedge directed.arg2(F,w) \wedge directed.in(G,x) & titanic & \lambda x.x & 1997
\end{array}
$$

$$
\begin{array}{cc}
\overline{S\backslash NP/PP_{in}/NP \quad NP}^{\;>} & \overline{PP_{in}/NP \quad NP}^{\;>} \\
S\backslash NP/PP & \\
\lambda x \lambda y.directed.arg1(E,y) \wedge directed.arg2(F,titanic) \wedge directed.in(G,x) & 1997
\end{array}
$$

$$
\xrightarrow{\quad > \quad} \quad S\backslash NP : \lambda y.directed.arg1(E,y) \wedge directed.arg2(F,titanic) \wedge directed.in(G,1997)
$$

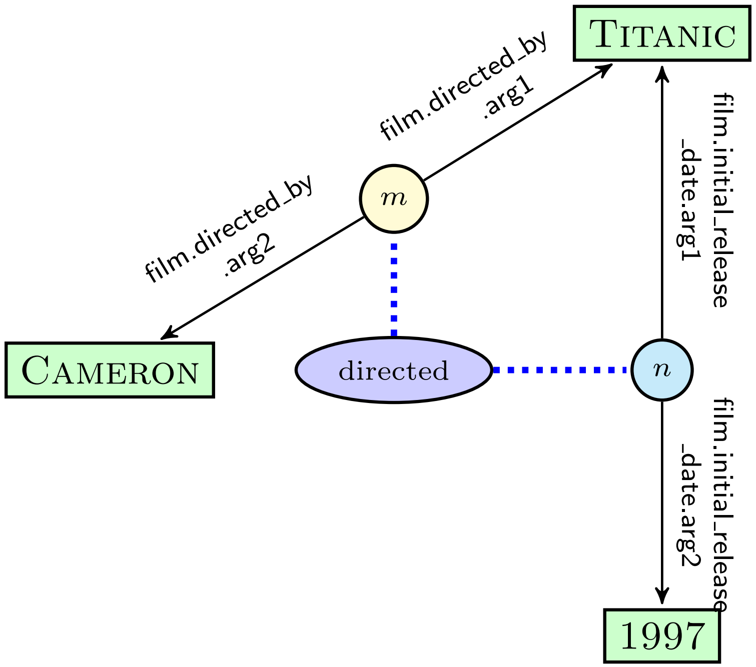
$$
\xrightarrow{\quad < \quad} \quad S : directed.arg1(E,cameron) \wedge directed.arg2(F,titanic) \wedge directed.in(G,1997)
$$

Map Logical Form to LF graph



$$\text{directed.arg1}(e, \text{Cameron}) \wedge$$
$$\text{directed.arg2}(e, \text{Titanic}) \wedge \text{directed.in}(e, 1997)$$

Map LF graph to Knowledge graph



$$\begin{array}{c}\text{film.directed\_by.arg2}(m, \textsc{Cameron}) \land \\ \text{film.directed\_by.arg1}(m, \textsc{Titanic}) \land \\ \text{film.initial\_release\_date.arg1}(n, \textsc{Titanic}) \land \\ \text{film.initial\_release\_date.arg2}(n, 1997)\end{array}$$

The Nature of the Mapping

- In the ungrounded graph, we need to replace
  - Entity variables with Freebase entities (e.g. *Cameron* with CAMERON)
  - Edge labels with Freebase relations (e.g. directed.arg1 with film.directed_by.arg2)
  - Event variables with factual variables (e.g. E becomes m and F becomes n)

- But there are $O(k+1)^n$ grounded graphs possible for each logical form (including no edges)

Learning from Denotations

- Learning proceeds by creating question-like logical forms by replacing named entities in logical forms mined from web text with a variable to produce property-denoting graphs, such as the one corresponding to:
  $\lambda x.directed.arg1(E, cameron) \wedge directed.arg2(F, x) \wedge directed.in(G, 1997)$
- The learner then finds the denotation of this property from other similar sentences in the mined logical forms—in this case, other films directed by Cameron.
- It then tries to find the subgraph of the knowledge graph with the the most similar denotation—in this case, the subgraph composed of relations m and n.
- The mapping of terms from logical forms to Freebase is determined by such pairings.

Choosing a Knowledge Base Subgraph

- A number of heuristics exploit similarities between the two graphs (cf. Kwiatkowski *et al.* 2013).
- Learning is by Averaged Perceptron (Collins, 2002).
- Features classes are:
  - subsumption relations between semantic graph and knowledge base subgraph;
  - Lexical similarity of edge labels in semantic graph and knowledge base subgraph;
  - Multiple knowledge base edge labels with the same stem;
  - Multiple knowledge base edges with the same mediating fact label;
- There are also a number of heuristic constraints on the answer term, such as definiteness/uniqueness.

Experiments

- Training Data: ClueWeb09, a snapshot of Web in 2009
  - 503.9 million webpages
  - Automatically annotated with Freebase entities
  - Select sentences containing at least two entities in relation in Freebase
  - Noisy lexicon for lexical alignments initialisation
- Test Datasets: Free917 and WebQuestions

Freebase Domains

- Target Domains: Business, Film, People
  - Largest domains of Freebase
- 5-10 million denotation queries for 10-20 iterations
  - Virtuoso RDF/SQL server
  - Slow in dealing with millions of queries
  - So we currently work with limited domains

Results

Dataset	System	P	R	F
Free917	MWG	52.6	49.1	50.8
	KCAZ13	72.6	66.1	69.2
	GRAPHPARSER	**81.9**	**76.6**	**79.2**
WebQuestions	MWG	39.4	34.0	36.5
	PARASEMPRE			37.5
	GRAPHPARSER	**41.9**	**37.0**	**39.3**

- MWG: Greedy Maximum Weighted Graph; KCAZ13: Kwiatkowski *et al.* (2013) supervised model; PARASEMPRE: Berant and Liang (2014) supervised model along with paraphrasing; GRAPHPARSER: Our model

Error Analysis on Free917

- **Syntactic Parser**: 25% e.g. When Gatorade was first developed?

- **Freebase inconsistencies**: 19% e.g. How many stores are in Nittany_mall?

- **Structural Mismatch**: 15% (Interesting category)

  – *president* as type in language
  – employment.job.title as relation in Freebase

- **Misc**: Ambiguity e.g. What are some films on Antarctica?

Error Analysis on WebQuestions

- $>$15% structural mismatch between language and Freebase
  - What did Charles Darwin do? (Charles Darwin does Biologist)
  - Where did Charles Darwin come from? (UK vs The Mount)
  - Who is the grandmother of Prince William? (Freebase does not express grandmother relation directly.)

Error Analysis on WebQuestions

- Reddy adds two paraphrase rules which convert *do* $\Rightarrow$ *profession*, and *come from* $\Rightarrow$ *birthplace*.

Dataset	System	P	R	F
WebQuestions	MWG	39.4	34.0	36.5
	PARASEMPRE			37.5
	GRAPHPARSER	41.9	37.0	39.3
	GRAPHPARSER+PARA	44.7	38.4	41.3

Interim Summary

- Scalable Semantic Parsing without Question-Answer pairs
- Semantic Parsing as a Graph Matching Problem
- Denotation-based weak supervision
- Improves over the state of the art

III: Another Solution

- Treat the knowledge graphs as a case for machine translation.
- Find the part of the knowledge graph that looks like the best translation of the question graph into knowledge graphics.
- To do this we need logical forms for NL that are less tied to the form of specific sentences in specific languages, embodying notions of paraphrase and entailment as a founding principle, rather than as an add-on.
- A different aproach to graded semantics from the vector-based approaches of Garrette *et al.* (2011); Beltagy *et al.* (2013) and Riedel *et al.* (2013)

Learning the Hidden Language of Logical Form

- The problem is that there are too many ways of asking and answering questions, and we have no idea of the semantics that relates them.
- *Did Google buy YouTube?*

  1. Google purchased YouTube.
  2. Google's purchase of YouTube
  3. Google acquired every company.
  4. YouTube may be sold to Google.
  5. Google will buy YouTube or Microsoft.
  6. Google didn't take over YouTube.

- Major motivation for PropBank/VerbNet annotation (Banarescu *et al.*, 2012)
- Can we do it automatically?

Answering questions by Parsing Text

- Question answering with traditional logical forms and theorem proving augmented with resources like WordNet, gazeteers, etc. has precision around 75%–but recall is around 4% (Bos and Markert 2005).

- This is worse than finite-state string matching

- Instead, Lewis and Steedman 2013a parse text errorfully to mine clusters of paraphrases for relations between named entities like "Shakespeare" and "Macbeth", "Google" and "YouTube" (cf. Riedel *et al.* 2013).

- Typing the relation clusters probabilistically eliminates ambiguity of the *Born in Hawai'i/born in 1961* kind.

- Logical operators such as negation and modality are handled by a (more or less) traditional Montagovian semantics (Steedman 2012)

Results: PASCAL Challenge dataset

- Examples:

Question	Answer	Sentence
What did Delta merge with?	Northwest	The 747 freighters came with Delta's acquisition of Northwest
What spoke with Hu Jintao?	Obama	Obama conveyed his respect for the Dalai Lama to China's president Hu Jintao during their first meeting
What arrived in Colorado?	Zazi	Zazi flew back to Colorado. . .
What ran for Congress?	Young	. . . Young was elected to Congress in 1972

- Full results in Lewis and Steedman (2013a)

Clustering Cross-linguistically

- We apply the method to answer questions in language A from text in language B, using standard MOSES MT as a baseline (Lewis and Steedman 2013b).
  - Find the answer to: *Who wrote Measure for Measure?*
  - From e.g. French—e.g.: *Shakespeare est l'auteur de* Mesure pour mesure.

- We also use cross-linguistic clusters to re-rank Moses n-best lists to promote translations that preserve the cluster-based meaning representation from source to target.

Example

Source: Le Princess Elizabeth arrive à Dunkerque le 3 août 1999
SMT 1-best: The Princess Elizabeth is to manage to Dunkirk on 3 August 1999.
Reranked 1-best: The Princess Elizabeth arrives at Dunkirk on 3 August 1999.

Experiment II: Reranking Moses

	Percentage of Translations preferred
1-best Moses	5%
Reranked best	39%
No preference	56%

- Many cases of "no preference" were where Moses and the prefered translation were similar strings but differed in attachment decisons invisible to the human judges.

- These results are obtained without the use of parallel text.

- Full results in Lewis and Steedman (2013b).

Directional Entailments (Lewis, 2014)

- The above approach engenders overinclusive paraphrase clusters.

- X_{person} *elected to* Y_{office} entails X_{person} *ran for* Y_{office} but not *vice versa*.

- The paraphrase relation depends on more global properties of the named entity relation graph.

- Lewis (2014); Lewis and Steedman (2014) apply the entailment graphs of Berant *et al.* (2012) to generate more articulated entailment structures.

Local Entailment Probabilities

- The typed named-entity technique is applied to estimate local probabilites of entailments:

a. $p(conquer\,x\,y \Rightarrow invade\,x\,y) = 0.9$

b. $p(invade\,x\,y \Rightarrow attack\,x\,y) = 0.8$

c. $p(conquer\,x\,y \Rightarrow attack\,x\,y) = 0.4$

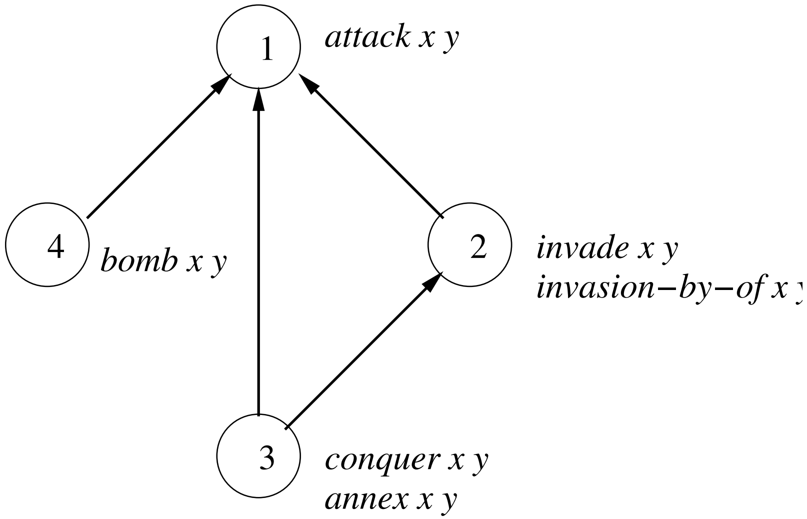
d. $p(bomb\,x\,y \Rightarrow attack\,x\,y) = 0.7$

(etc.)

Local Entailment Probabilities

- These are used to construct an entailment graph using integer linear programming with $\pm$ weights around $p = 0.5$ with the global constraint that the graph must be closed under transitivity.
- Thus, (c) will be included despite low observed frequency.
- Cliques within the entailment graphs are collapsed to a single cluster relation identifier, as in the previous approach.

Entailment graph



- A simple entailment graph for relations between countries.

Lexicon

- The lexicon obtained from the entailment graph

  attack := $(S\backslash NP)/NP : \lambda x\lambda y\lambda e.rel_1\, x\, y\, e$
  bomb := $(S\backslash NP)/NP : \lambda x\lambda y\lambda e.rel_1\, x\, y\, e \wedge rel_4\, x\, y\, e$
  invade := $(S\backslash NP)/NP : \lambda x\lambda y\lambda e.rel_1\, x\, y\, e \wedge rel_2\, x\, y\, e$
  conquer := $(S\backslash NP)/NP : \lambda x\lambda y\lambda e.rel_1\, x\, y\, e \wedge rel_2\, x\, y\, e \wedge rel_3\, x\, y\, e$
  annex := $(S\backslash NP)/NP : \lambda x\lambda y\lambda e.rel_1\, x\, y\, e \wedge rel_2\, x\, y\, e \wedge rel_3\, x\, y\, e$

- These logical forms support correct entailment inferences, such as that *conquered* entails *attacked* and *didn't invade* entails *didn't conquer*
- Primitives like rel_3 correspond to "hidden" semantic primitives that distinguish these concepts.
- If we do this cross-linguistically we may see that some of them correspond to universal elements like evidentiality that are masked in English.

Experiment

- Baselines are Majority Class (don't know) and Berant *et al.* 2011 Non Compositional direct entailment between reverb patterns.
- We also compare with Additive and Multiplicative Vector-based distributional semantics (SCS) using a logistic regression classifier.
- The Zeichner entailments, unlike RTE, rely predominantly on lexical entailment.

This dataset does not otherwise play to the syntactic and logical strengths of CCG, and includes many non-compositional idioms (eg light verb construction) quite favorable to e.g. vector composition.

Results

System	Accuracy (all)	AUC (all)	Accuracy (subset)	AUC (subset)
Majority Class	56.8%	0.46	51.5%	0.46
Non Compositional	57.4%	0.48	52.8%	0.48
CCG Baseline	57.8%	0.46	55.8%	0.53
CCG ChineseWhispers	58.0%	0.50	57.5%	0.56
VectorMultiplicative	61.3%	0.51	59.3%	0.56
VectorAdditive	63.5%	0.57	61.2%	0.61
CCG Entailment Graphs	64.0%	0.58	65.0%	0.65

- Subset Accuracy is on the set where we make a prediction.
- AUC is area under Precision-Recall curve, computed with a trapezoid approximation, as a measure of reliability of confidence estimates.

Conclusion: Lexical Semantics as Entailment

- The typed named entity-based technique allows us to construct the logical forms for content words as conjunctions of entailments.
- Under more traditional semantic theories employing eliminative definitions these entailments would have been thought of as belonging to the domain of inference rather than semantics, either as meaning postulates relating logical forms or as “encyclopædic” general knowledge.
- These conjunctive terms of this logical language are very close to the language-specific grammar, and support fast inference of entailment.

Conclusion: Lexical Semantics as Entailment

- We can think of the cliques or clusters in the graph as related to the hidden primitives of the Language of Mind.
- However, very few terms in the adult logical form correspond directly to primitives of the Language of Mind. (*red* and maybe *attack* might be exceptions.)
- Even those terms that are cognitively primitive like color terms will not be unambiguously lexicalized in all languages
- Perhaps this can be developed cross-linguistically into a full hidden interlingua for MT.

References

Banarescu, Laura, Bonial, Claire, Cai, Shu, Georgescu, Madalina, Griffitt, Kira, Hermjakob, Ulf, Knight, Kevin, Koehn, Philipp, Palmer, Martha, and Schneider, Nathan, 2012. "Abstract Meaning Representation (AMR) 1.0 Specification."

Beltagy, Islam, Chau, Cuong, Boleda, Gemma, Garrette, Dan, Erk, Katrin, and Mooney, Raymond, 2013. "Montague Meets Markov: Deep Semantics with Probabilistic Logical Form." In *2nd Joint Conference on Lexical and Computational Semantics (*SEM): Proceeding of the Main Conference and the Shared Task*. 11–21.

Berant, Jonathan, Chou, Andrew, Frostig, Roy, and Liang, Percy, 2013. "Semantic

Parsing on Freebase from Question-Answer Pairs." In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. Seattle: ACL, 1533–1544.

Berant, Jonathan, Dagan, Ido, Adler, Meni, and Goldberger, Jacob, 2012. "Efficient Tree-Based Approximation for Entailment Graph Learning." In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1*. Association for Computational Linguistics, 117–125.

Berant, Jonathan, Goldberger, Jacob, and Dagan, Ido, 2011. "Global Learning of Typed Entailment Rules." In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*. Portland, OR, 610–619.

Berant, Jonathan and Liang, Percy, 2014. "Semantic Parsing via Paraphrasing." In

Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Baltimore, MD, 1415–1425.

Börschinger, Benjamin, Jones, Bevan K., and Johnson, Mark, 2011. "Reducing Grounded Learning Tasks To Grammatical Inference." In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*. Edinburgh: ACL, 1416–1425.

Bos, Johan and Markert, Katja, 2005. "Combining Shallow and Deep NLP Methods for Recognizing Textual Entailment." In *Proceedings of the First PASCAL Challenge Workshop on Recognizing Textual Entailment*. `http://www.pascal-network.org/Challenges/RTE/`: Pascal, 65–68.

Cai, Qingqing and Yates, Alexander, 2013a. "Large-scale Semantic Parsing via Schema Matching and Lexicon Extension." In *Proceedings of the 51st Annual*

Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Sofia, Bulgaria: ACL, 423–433.

Cai, Qingqing and Yates, Alexander, 2013b. "Semantic Parsing Freebase: Towards Open-domain Semantic Parsing." In *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 1: Proceedings of the Main Conference and the Shared Task: Semantic Textual Similarity*. Atlanta: ACL, 328–338.

Clarke, James, Goldwasser, Dan, Chang, Ming-Wei, and Roth, Dan, 2010. "Driving Semantic Parsing from the World's Response." In *Proceedings of the Fourteenth Conference on Computational Natural Language Learning*. Uppsala: ACL, 18–27.

Collins, Michael, 2002. "Discriminative Training Methods for Hidden Markov Models: Theory and Experiments with Perceptron Algorithms." In *Proceedings* *of the 10th Conference on Empirical Methods in Natural Language Processing*. ACL, 1–8.

Garrette, Dan, Erk, Katrin, and Mooney, Raymond, 2011. "Integrating Logical Representations with Probabilistic Information using Markov Logic." In *Proceedings of the Ninth International Conference on Computational Semantics*. ACL, 105–114.

Kwiatkowski, Tom, Choi, Eunsol, Artzi, Yoav, and Zettlemoyer, Luke, 2013. "Scaling Semantic Parsers with On-the-Fly Ontology Matching." In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. Seattle: ACL, 1545–1556.

Kwiatkowski, Tom, Goldwater, Sharon, Zettlemoyer, Luke, and Steedman, Mark, 2012. "A Probabilistic Model of Syntactic and Semantic Acquisition from

Child-Directed Utterances and their Meanings." In *Proceedings of the 13th Conference of the European Chapter of the ACL (EACL 2012)*. Avignon: ACL, 234–244.

Kwiatkowski, Tom, Zettlemoyer, Luke, Goldwater, Sharon, and Steedman, Mark, 2010. "Inducing Probabilistic CCG Grammars from Logical Form with Higher-Order Unification." In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Cambridge, MA: ACL, 1223–1233.

Kwiatkowski, Tom, Zettlemoyer, Luke, Goldwater, Sharon, and Steedman, Mark, 2011. "Lexical Generalization in CCG Grammar Induction for Semantic Parsing." In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Edinburgh: ACL, 1512–1523.

Lewis, Michael, 2014. *Natural Semantics for Wide Coverage CCG Parsers*. Ph.D. thesis, University of Edinburgh.

Lewis, Michael and Steedman, Mark, 2013a. "Combined Distributional and Logical Semantics." *Transactions of the Association for Computational Linguistics* 1:179–192.

Lewis, Mike and Steedman, Mark, 2013b. "Unsupervised Induction of Cross-Lingual Semantic Relations." In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. ACL, 681–692.

Lewis, Mike and Steedman, Mark, 2014. "Combining Formal and Distributional Models of Temporal Semantics." In *Proceedings of the ACL Conference Workshop on Semantic Parsing*. 28–32.

Liang, Percy, Jordan, Michael, and Klein, Dan, 2011. "Learning Dependency-Based Compositional Semantics." In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*. Portland, OR: ACL, 590–599.

Lu, Wei, Ng, Hwee Tou, Lee, Wee Sun, and Zettlemoyer, Luke S., 2008. "A Generative Model for Parsing Natural Language to Meaning Representations." In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*. Honolulu, Hawaii: ACL, 783–792.

Montague, Richard, 1970. "English as a Formal Language." In Bruno Visentini (ed.), *Linguaggi nella Società e nella Technica*, Milan: Edizioni di Communità. 189–224. Reprinted as Thomason 1974:188-221.

Reddy, Siva, 2014. "Large-scale Semantic Parsing without Question-Answer Pairs." *TACL, subject to revisions* .

Riedel, Sebastian, Yao, Limin, McCallum, Andrew, and Marlin, Benjamin, 2013. "Relation Extraction with Matrix Factorization and Universal Schemas." In *Proceedings of the 2013 Conference of the North American Chapter of the*

Association for Computational Linguistics: Human Language Technologies. Atlanta: ACL, 74–84.

Steedman, Mark, 2000. *The Syntactic Process*. Cambridge, MA: MIT Press.

Steedman, Mark, 2012. *Taking Scope: The Natural Semantics of Quantifiers*. Cambridge, MA: MIT Press.

Thomason, Richmond (ed.), 1974. *Formal Philosophy: Papers of Richard Montague*. New Haven, CT: Yale University Press.

Thompson, Cynthia and Mooney, Raymond, 2003. "Acquiring Word-Meaning Mappings for Natural Language Interfaces." *Journal of Artificial Intelligence Research* 18:1–44.

Wong, Yuk Wah and Mooney, Raymond, 2007. "Learning Synchronous Grammars for Semantic Parsing with Lambda Calculus." In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*. ACL, 960–967.

Zettlemoyer, Luke and Collins, Michael, 2005. "Learning to Map Sentences to Logical Form: Structured Classification with Probabilistic Categorial Grammars." In *Proceedings of the 21st Conference on Uncertainty in AI (UAI)*. Edinburgh: AAAI, 658–666.

Zettlemoyer, Luke and Collins, Michael, 2007. "Online Learning of Relaxed CCG Grammars for Parsing to Logical Form." In *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP/CoNLL)*. Prague: ACL, 678–687.