Unsupervised Design of Explainable Meaning Representations

Mark Steedman

(*with* Javad Hosseini (ATI, Edinburgh), and many others)

July 14th 2020

Outline

- I: The Problem for NLP of Linguistic Form-dependence in Semantics.
- II: Explainability and Form-Independence in Semantics.
- III: Natural Language Inference (NLI) At Scale (Hosseini *et al.*, 2018, 2019; Hosseini, 2020; Hosseini *et al.*, submitted).
- IV: Work in Progress Towards Form-Independent Semantics

I: The Problem

- We have (somewhat) robust wide-coverage (supervised) parsers that work on the scale of Bn of words. They can read the web (and build logical forms) much faster than we can ourselves.
- So why can't we have them read the web for us, and answer questions, rather than giving us a bunch of snippets that may or may not allow us to do that ourselves?

Too Many Ways of Answering The Question

- The central problem of QA is that it involves inference as well as semantics, and (despite our best efforts), we have no idea of the logic involved.
- Your Question: *Does Verizon own Yahoo?* The Text:

1. Verizon purchased Yahoo. ("Yes")
2. Verizon's purchase of Yahoo ("Yes")
3. Verizon managed to buy Yahoo. ("Yes")
4. Verizon acquired every company. ("Yes")
5. Verizon sold Yahoo ("No")
6. Yahoo may be sold to Verizon. ("Maybe")
7. Verizon will buy either Yahoo or Yazoo. ("Maybe not")

- No chance of using sequence-to-sequence learning, since we don't have any labeled data.

The Problem of Content

- The problem is that, as adults, we have almost no access to the conceptual basis for verbs like “buy” or “run” that supports uses like the following:
  1. Mayor Koch runs this town.
  2. Koch ran for Mayor.
  3. It runs from Chicago to LA
  4. My luck ran out.
  5. *&c. &c.*
- Nevertheless, inference with these apparently “metaphorical” uses is effortless.
- Perhaps they are all atomic relations, typed in some equally obscure natural ontology, and connected by a graph of meaning postulates (Fodor, 1975).
- Constructing such a semantics by hand would be hard. Perhaps we can do it by machine.

II: The Approach

- Use semantic parsers to Machine-Read multiple relations over Named Entities in web text.
- Capture relations of entailment and paraphrase over relations between NEs of the same types (Lewis and Steedman, 2013a,b, 2014; Lewis, 2015).
  - If you read somewhere that a person—say, Biden—was elected to an office—say, President—than you are highly likely to also read somwhere that that person ran for that office—
  - —but not the other way round
- Entailment can be detected under the Distributional Inclusion Hypothesis (Geffet and Dagan, 2005) over pairs of vectors of typed named-entity tuples such as $\langle Biden, president \rangle$ for relations like *be elected* and *run*, using a directional similarity measure. (In what follows, we use BInc.)

Unsupervised Explainable Form-Independent Semantics

- Cliques in the entailment graph represent paraphrase clusters that can be collapsed to a single relation, represented by a single identifier.
- The parser semantics can then be redefined in terms of the entailment graph and form-independent paraphrase identifiers.
- The parser can then be used to create a knowledge graph from text for IR/QA.
- Such graphs (which used to be called "Semantic Networks") can be interrogated directly in natural language, using the same form-independent semantics.
- This form-independent semantics is explainable and immediately compatible with logical semantic devices like quantifiers and negation.

There is Another Approach . . .

- There is another widely-canvassed approach to meaning, based on vector embeddings and linear-algebraic composition.
- It is hard to define directional similarity measures for embeddings.
- It is also hard to combine them with logical operators.
- Embeddings are associative rather than semantic.
- They are by their nature opaque to explanation.
- I'll argue that their strength lies in disambiguation rather than meaning representation as such.
- We'll return to this point later.

Toy Example: Local Entailment Probabilities

- First, the typed named-entity technique is applied to (errorfully) estimate local probabilities of entailments:

a. $p(buy\,x\,y \Rightarrow acquire\,x\,y) = 0.9$
b. $p(acquire\,x\,y \Rightarrow own\,x\,y) = 0.8$
c. $p(acquisition\,(of\,x)\,(by\,y) \Rightarrow own\,x\,y) = 0.8$
d. $p(acquire\,x\,y \Rightarrow acquisition\,(of\,x)\,(by\,y)) = 0.7$
e. $p(acquisition\,(of\,x)\,(by\,y) \Rightarrow acquire\,x\,y) = 0.7$
f. $p(buy\,x\,y \Rightarrow own\,x\,y) = \mathbf{0.4}$
g. $p(buy\,x\,y \Rightarrow buyer\,(of\,x)\,y) = 0.7$
h. $p(buyer\,(of\,x)\,y \Rightarrow buy\,x\,y) = 0.7$
i. $p(inherit\,x\,y \Rightarrow own\,x\,y) = 0.7$
(etc.)

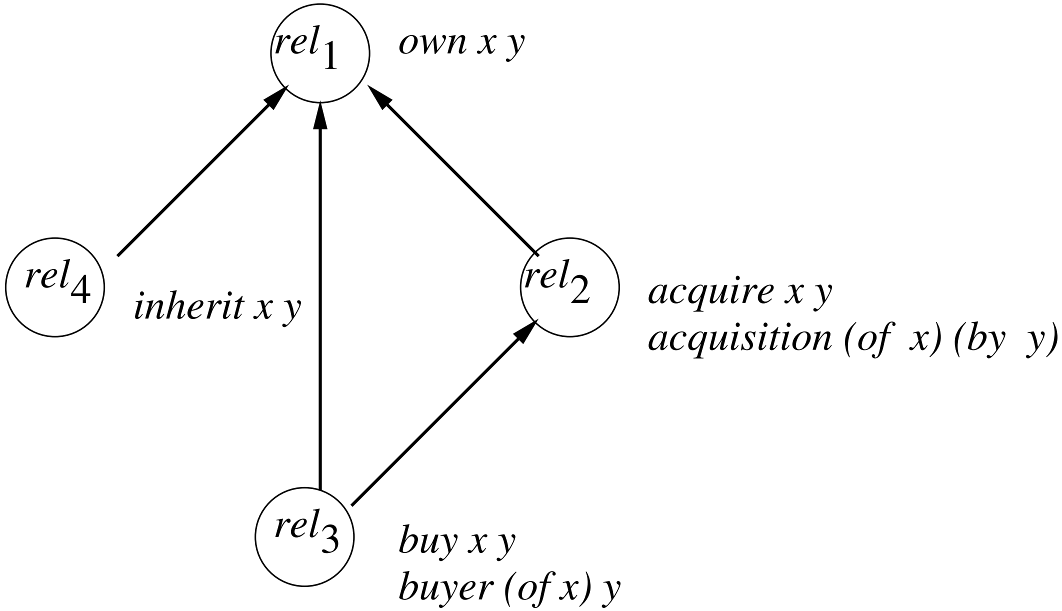
Toy Example: Global Entailments

- The local entailment probabilities are used to construct an entailment graph, with the global constraint that the graph can be closed under transitivity of entailment (Berant *et al.*, 2015).

- Thus, local entailment (f) is supported by transitivity despite low observed frequency, while unsupported spurious low frequency local entailments can be excluded.

- Cliques within the entailment graphs can be collapsed to a single paraphase cluster relation identifier.

Toy Example: Entailment graph



- A simplified entailment graph for relations between people and property.
- Should be thought of as a structure of typed Meaning Postulates (Carnap, 1952), rather than entailment in the logician's sense.

Toy example: Form-Independent Lexicon

- The new semantics obtained from the entailment graph replaces form-dependent relations like *acquire* with paraphrase cluster identifiers like rel_2

own	:=	$(S\backslash NP)/NP$	$: \lambda x \lambda y.rel_1 x y$
inherit	:=	$(S\backslash NP)/NP$	$: \lambda x \lambda y.rel_4 x y$
acquire	:=	$(S\backslash NP)/NP$	$: \lambda x \lambda y.rel_2 x y$
's acquisition of	:=	$(N\backslash NP)/NP$	$: \lambda x \lambda y.rel_2 x y$
buy	:=	$(S\backslash NP)/NP$	$: \lambda x \lambda y.rel_3 x y$
buyer of	:=	N/PP_{of}	$: \lambda x \lambda y.rel_3 x y$
etc.			

- These logical forms support correct inference under negation, such as that *Verizon bought Yahoo* entails *Verizon acquired Yahoo* and *Verizon doesn't own Yahoo* entails *Verizon didn't buy Yahoo*

Potential Applications

1. Question Answering.

2. Reranking machine Summarization.

3. Building Knowledge Graphs from text.

- There is an acute lack of suitable test data.

III: Natural Language Inference (NLI) At Scale

- We have trained an entailment graph on the NewsSpike corpus (Hosseini *et al.*, 2018)
  - 0.5M multiply-sourced news articles over 2 months, 20M sentences.
  - 29M binary relation tokens extracted using the CCG parser.
- We have built a 19GB working typed global entailment graph:
  - 101K relation types
  - 346 local typed entailment subgraphs
  - 23 subgraphs with more than 1K nodes e.g. Person$\times$Location, Location$\times$Thing, Organisation$\times$Organisation, etc.
  - 7 subgraphs with more than 10K nodes.

Idioms, Metaphors, Presuppositions, etc.

- Idioms are found just like any other typed entailment:
  - $keep\_tabs\_on\,(\#government\_agency, \#thing) \models' s\_surveillance\_of\,(\#government\_agency$
- So are metaphors:
  - $take\_shot\_at\,(\#person, \#person) \models slam\,(\#person, \#person)$
- Likewise light verbs, particle verbs, etc.:
  - $call\_up(\#person, \#thing) \models work\_with\,(\#person, \#thing)$
- Presuppositions are relations entailed by another relation and its negation:
  - $manage\_to(\#person, \#event) \models try\_to\,(\#person, \#event)$
  - $\neg manage\_to(\#person, \#event) \models try\_to\,(\#person, \#event)$

Intrinsic Evaluation Datasets

- We evaluate on Levy/Holt's (Levy and Dagan, 2016) crowd-annotated entailment dataset
  - Improved by (Holt, 2018), adding inverse pairs and redoing the crowd annotation.
  - This move is essential to eliminate the learnable artefacts that plague NLI datasets (Gulordava *et al.*, 2018).
  - 18407 entailment pairs (3916 positively entailing, 14491 nonentailing).
- We also evaluate on Berant's dataset (Berant *et al.*, 2011), obtained by hand-building a gold-standard entailment graph for all parsed relations in their dataset for 10 frequent n-tuples of types, then comparing the extracted graph with this gold-standard.
  - 39012 entailment pairs (3472 positively entailing, 35585 nonentailing).

Refining the Entailment Graph

- Major problem with existing entailment graph learners:
  - Many correct edges are missing because of data sparsity
- Berant *et al.* (2011) used Integer Linear Programming to globalize local entailment graphs, using transitivity closure on the entailments as the objective function: $P \to Q$ and $Q \to R$ implies that $P \to R$.
- ILP does not scale to graphs with more than 100 nodes.
- Berant *et al.* (2015) propose an approximation, removing entailment links to make the graph "Forest-Reducible".
- FRG loses many valid entailments.

Global Learning of Typed Entailment Graphs

- Instead we propose a scalable method that does not directly depend on transitivity, but instead uses two global soft constraints.
  - Our method scales to more than 100K nodes.

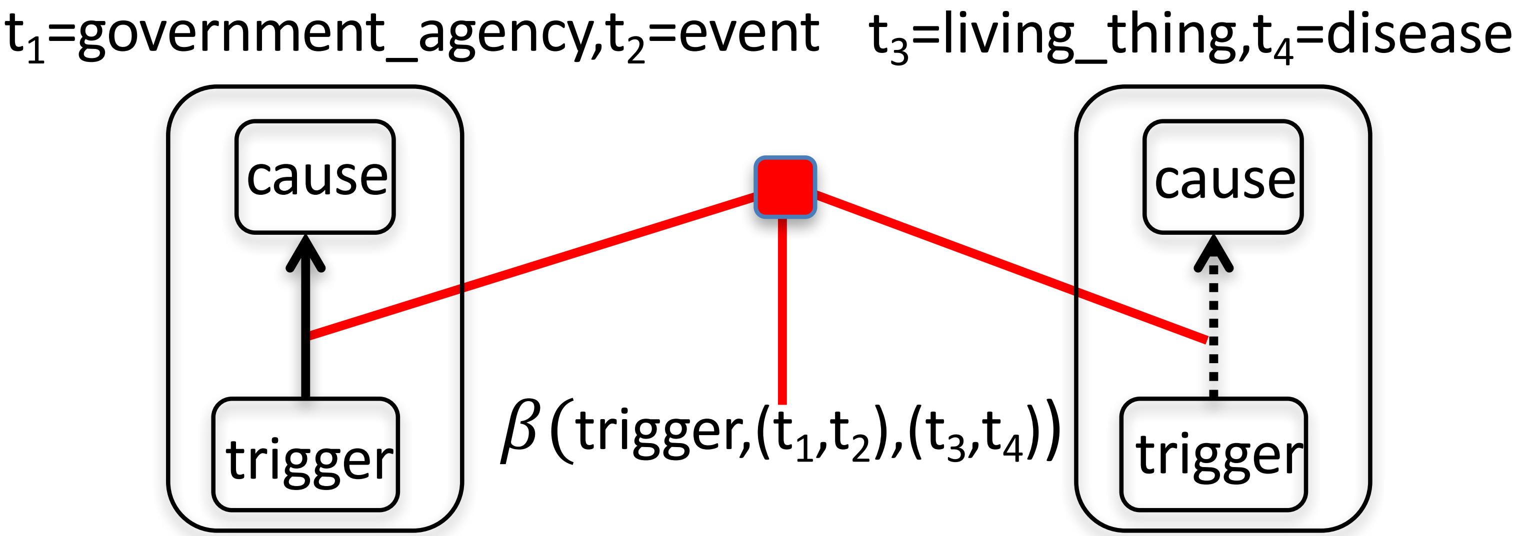
Global Soft Constraint 1: Cross Graph Transfer

- It is standard to learn a separate typed entailment graph for each (plausible) type-pair Berant *et al.* (2011, 2012); Lewis and Steedman (2013a,b); Berant *et al.* (2015).

- However, many entailment relations for which we have direct evidence only in a *few* subgraphs may apply over *many* others.

- This is a form of Domain Transfer.

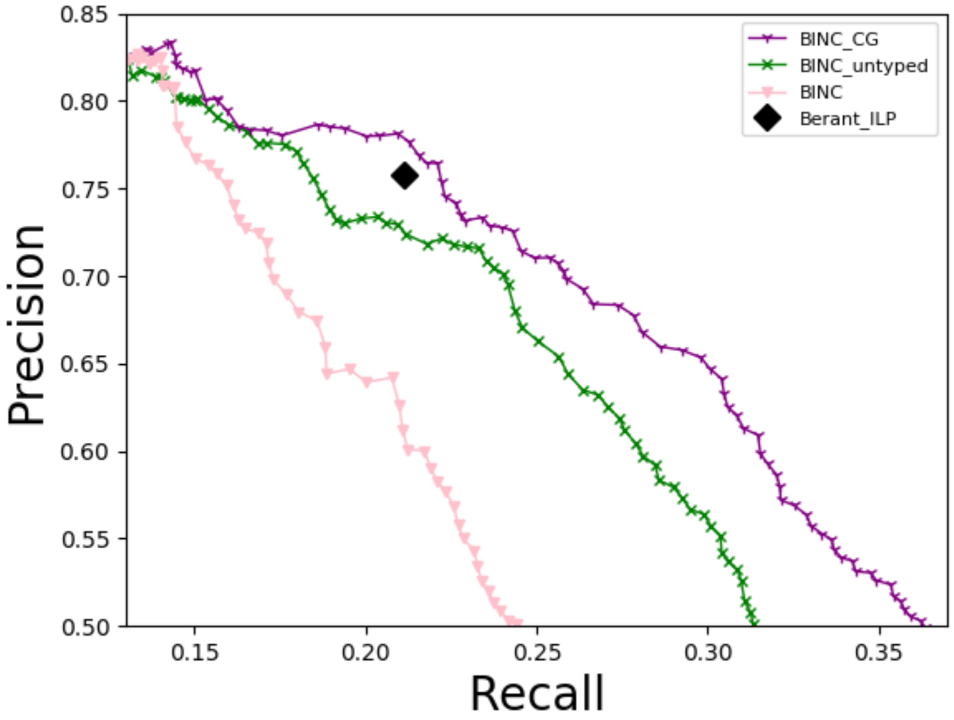
Global Soft Constraint 1: Cross Graph



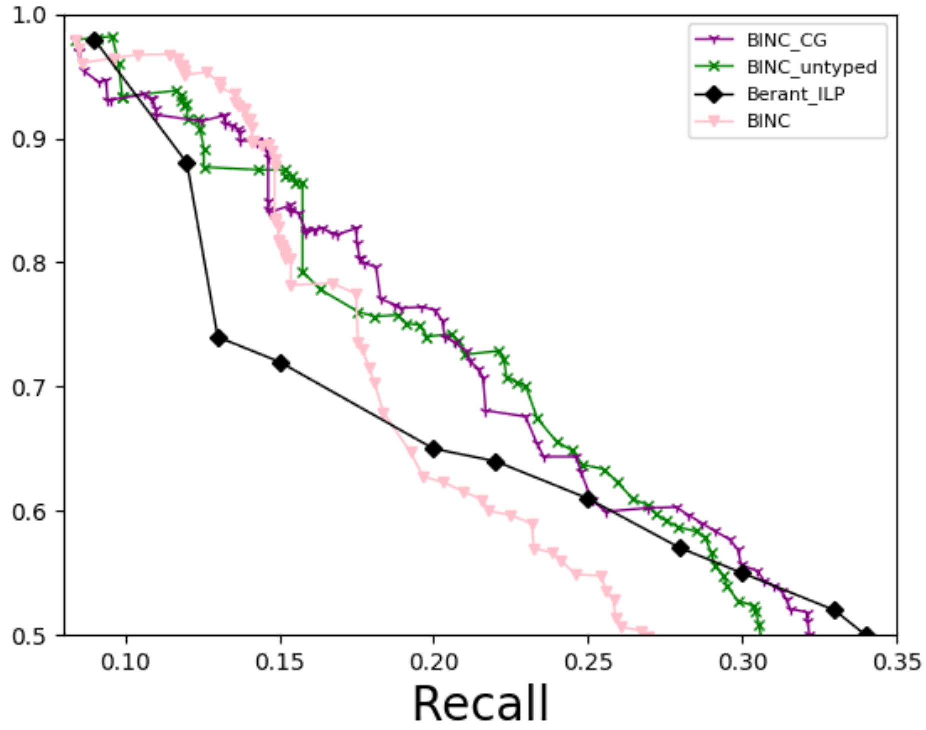
- $0 \leq \beta(.) \leq 1$ determines how similar differently typed graphs are, and is learned jointly.

Adding Cross-Graph Transfer Soft Constraints

Levy/Holt's dataset

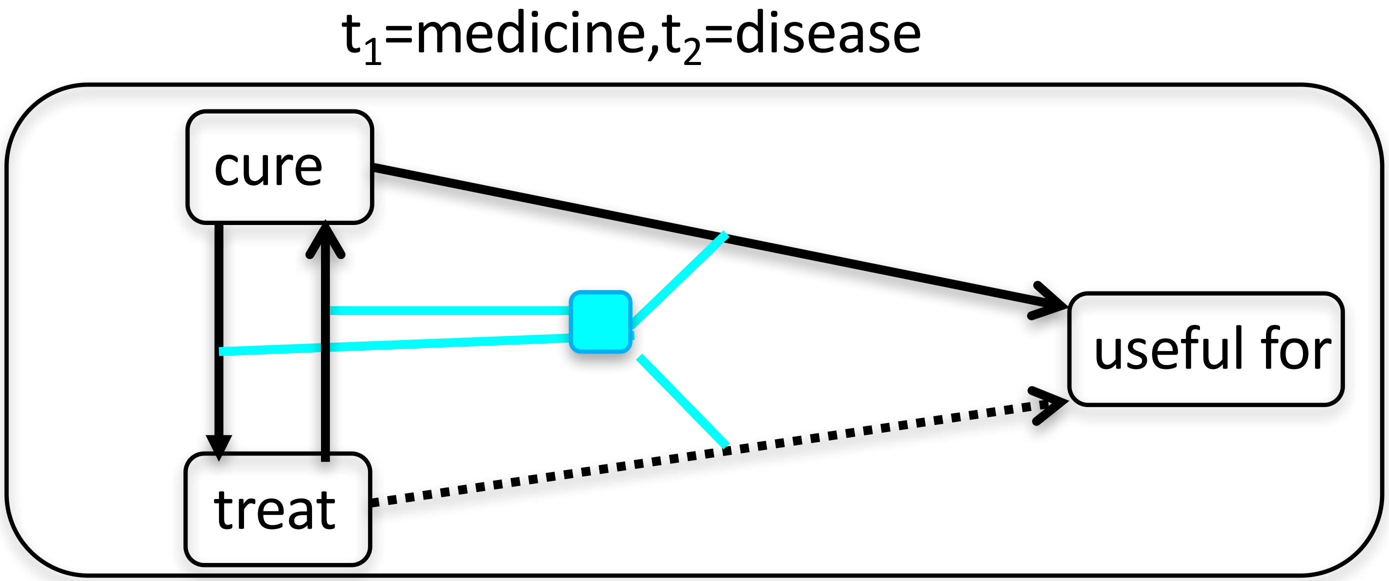


Berant's dataset



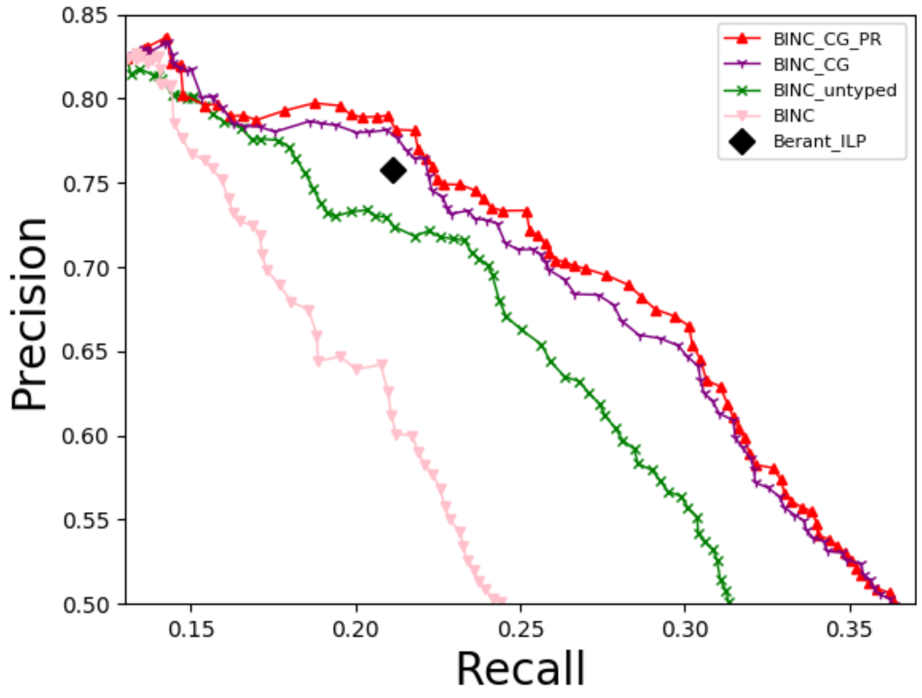
Global Soft Constraint 2: Paraphrase Resolution

- We encourage paraphrase predicates (where $i \rightarrow j$ and $j \rightarrow i$) to have the same patterns of entailment
  - i.e. to entail and be entailed by the same predicates

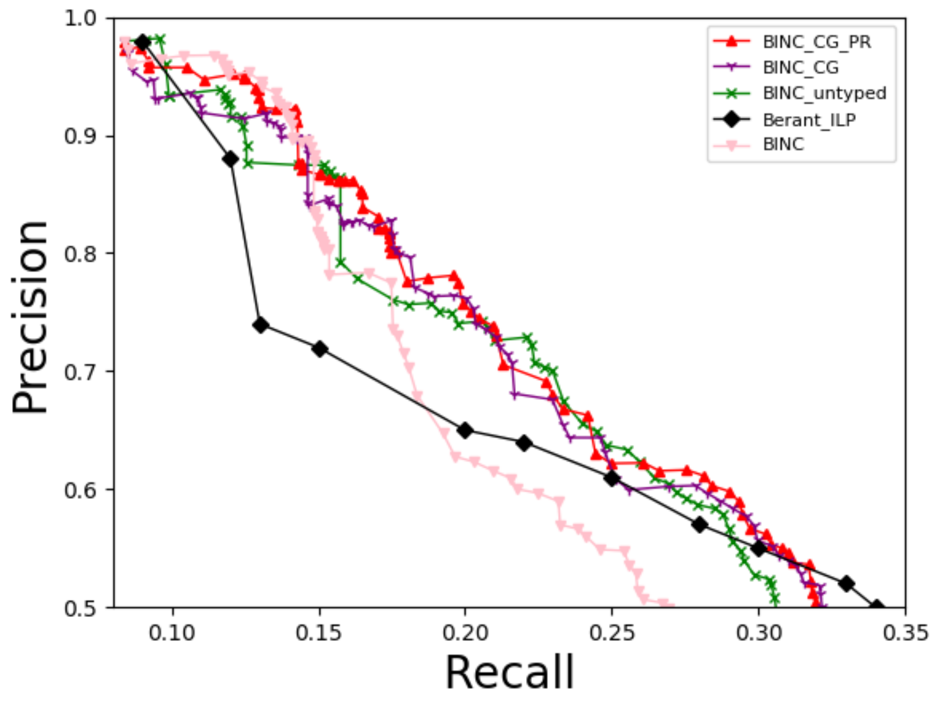


Adding Paraphrase Resolution Soft Constraints

Levy/Holt's dataset



Berant's dataset



Example Subgraph after CG and PR

Premise	Entails	Consequents
location suffers from *thing*	$\rightarrow$	*thing* killing in *location* *location* has *thing* *location* 's price for *thing* *location* suffers *thing* *location* diagnosed with *thing* destroyed during *thing* in *location* *thing* affects *location* *thing* 's image in *location* *location* recovers *thing* *location* 's *thing* *location* experiences *thing* took across *location* in *thing*

Test: Africa suffers from droughts $\rightarrow$ Africa experienced a drought Correct

Error Analysis

Error type	Example
	False Positive
High correlation (57%)	Microsoft **released** Internet Explorer → Internet Explorer **was developed by** Microsoft
Relation normalization (31%)	The pain **may be** relieved by aspirin → The pain can be treated with aspirin
Lemma baseline & parsing (12%)	President Kennedy **came** to Texas → President Kennedy **came** from Texas
	False Negative
Sparsity (93%)	Cape town **lies at the foot of** mountains → Cape town is located near mountains
Wrong label & parsing (7%)	Horses are imported from Australia → Horses are native to Australia

Using Embeddings in Entailment Graph Induction

- Rather than guessing entailment relations based on directional similarity of vectors of named-entity pairs, our colleagues frequently urge us to try the "alternative approach", representing relations as embeddings, and applying a directional distributional inclusion similarity measure.
- We keep trying this. It hasn't worked yet.
- However, Hosseini *et al.* (2019) map the relation-oriented entailment graph onto an equivalent Knowledge Graph, in which entities are the nodes.
- Applying embeddings-based methods for link-prediction (Riedel *et al.*, 2013; Dettmers *et al.*, 2018) to this knowledge graph improves the entailment graph.
- Conversely, the entailment graph improves link-prediction for low-scoring KG triples which are entailed by high-scoring paraphrases generated using the entailment graph.

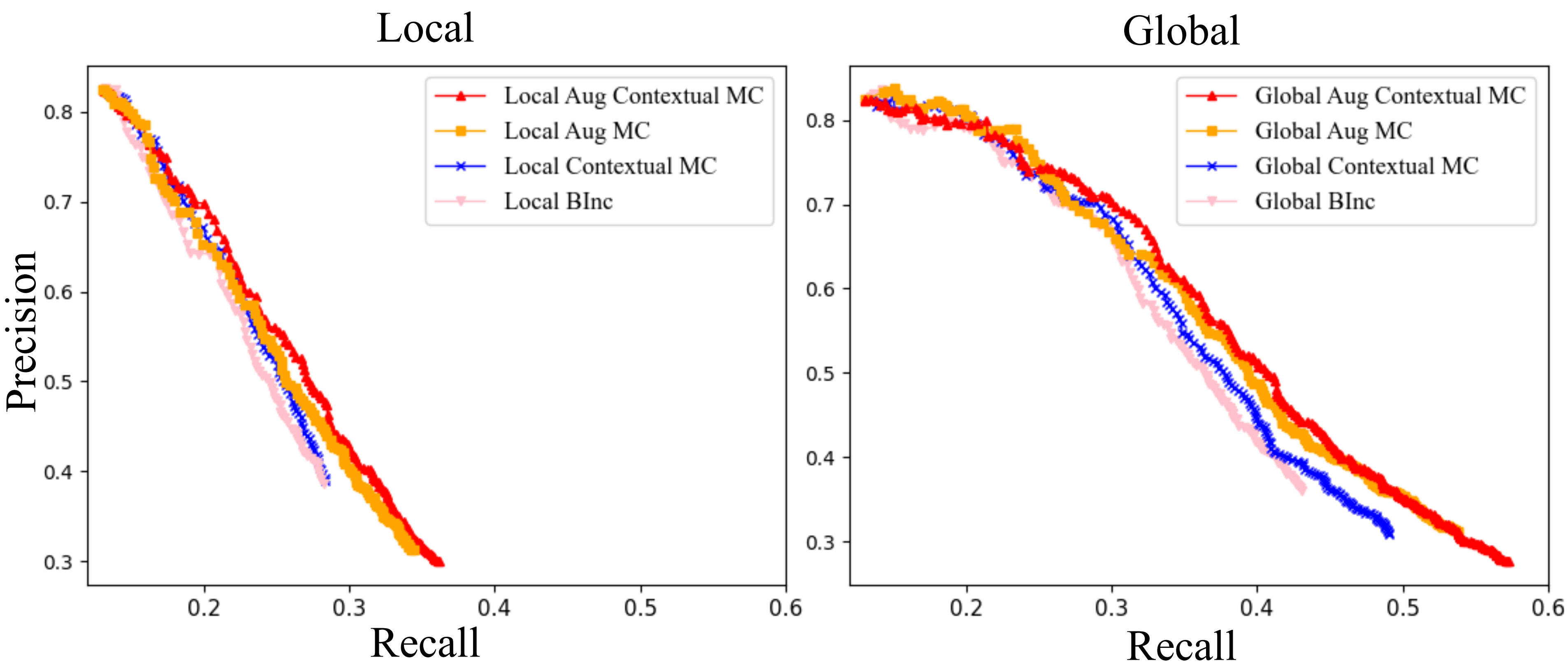
Duality of Embeddings and Entailment Graphs

- Hosseini *et al.* (submitted) show that these bidirectional effects are even stronger if the original string context for the entities and relations are included in the Knowledge Graph, and pretrained context-dependent BERT embeddings are used, fine tuned with a contextualized link-prediction objective
- Embeddings therefore seem to learn information that is complementary to machine-reading.
- This seems to reflect the fundamental distinction between the semantic nature of entailment and the associative nature of embeddings.
- We conjecture that the embeddings work as a latent fine-grain type classifier that mitigates the shortcomings of the Figer types used in relation extraction
- . . . while the directionality of entailment mitigates the non-directionality of embeddings.

Examples

Contextual Link Prediction Improves Entailment Graphs		
Triple Predictions	Microsoft **is committed to** success	
	Microsoft **builds** success	↓
	Microsoft **switches to** success	↓
	Microsoft **’s** success	↑
	Microsoft **achieves** success	↑
	Microsoft **hopes for** success	↑
Entailment Graphs Improve Contextual Link Prediction		
Triple Predictions	Apple **is working on** watch	
	Watch **falls on** Apple	↓
	Apple **’s** watch	↑
	Apple **has** watch	↑
	Apple **launches** watch	↑
	Apple **tests** watch	↑

Precision-Recall Curves



Recent Extensions

- We have shown that the method scales to a larger Entailment Graph (35GB) using an order of magnitude more data from our NewsCrawl data-set, built from Common Crawl for English (5.4M articles from 62 sources) and German (13M from 261 sources).

- We have also extended the method to entailments over unary relations, such as that if *A kills B*, *B is dead, is deceased* etc. (McKenna *et al.*, submitted)

IV: Current Work

- Refine method for building entailment graphs to take account of Temporality.
- Define Form-independent clustered entailment-based Semantic Parser.
- Use it to build a Large Knowledge Graph using form-independent semantic representations from text.
- Port the same form-independent semantics to other languages.

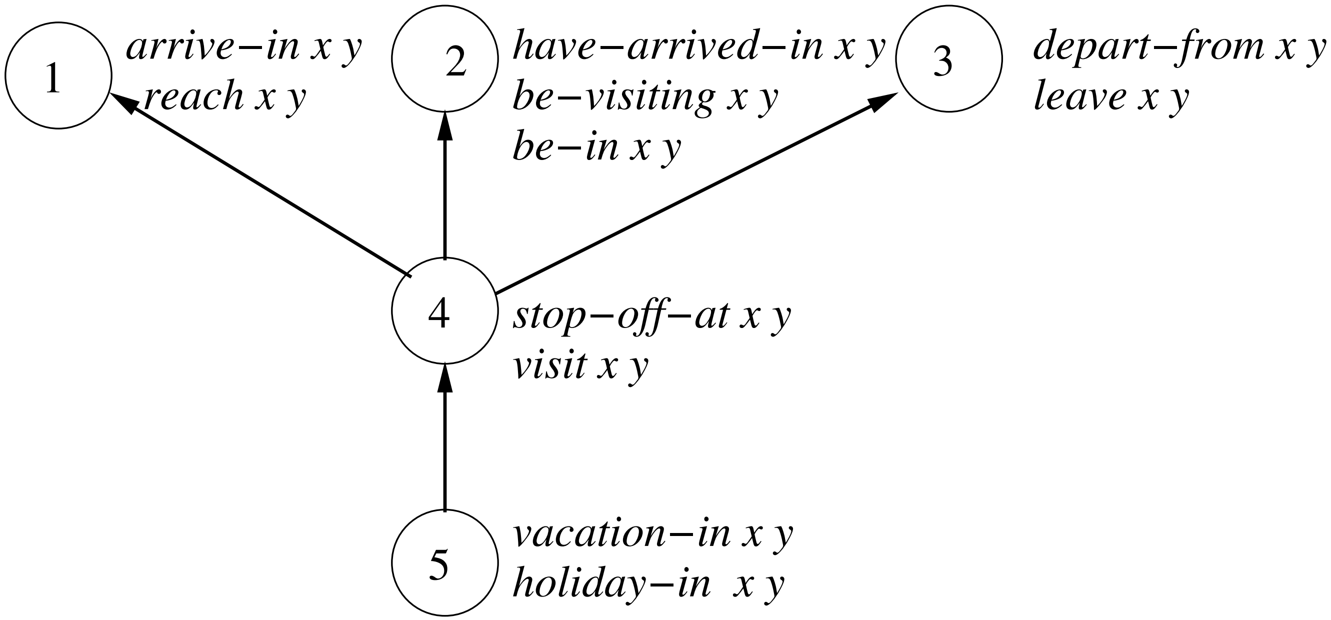
We Need to Take Account of Tense (etc.)

- A fallible assumption in what we have seen up till now is that all sentences about the same n-tuple of entities refer to the same event-complex.
- There are some exceptions. In the domain of sports, there are multiple occurrences of e.g. *Watford playing Ipswich*.
- Since in some of them Watford wins, and in others, Ipswich, our procedure concludes that *A winning against B* entails *A losing against B*.
- Segregating text by dateline, tense, and temporal modifiers helps a bit.
- But the window for ramifications depends on the core event.
- For sports, around a week, but for elections more than a year.

Modality, Aspect, Conditionals, Negation . . .

- It's worse than that. For the same game, one text may say that *Watford may win*, while another says *If Watford loses, ...*
- We currently ignore modal, aspectual, conditional etc. modifiers to counter sparsity, so again we conclude that winning entails losing.
- The problem here is that playing a match entails a disjunction: either you win or you lose.
- Most entailments like $buy \models own$, $be\ elected \models run$, aren't like that.
- One way to handle the modal $win \models lose$ problem may be to reject any entailment $P \models Q$ where P and Q are identified as antonyms by e.g. WordNet.
- However, WordNet is notoriously incomplete (e.g. "draw/tie" are missing as antonyms of win/lose).
- Embeddings might help with this (Ono *et al.*, 2015).

Form-Independent Temporal Semantics



- A simplified entailment graph for relations over events does not capture relations of causation and temporal sequence (Moens and Steedman, 1988; Pustejovsky, 1991; Lewis and Steedman, 2014).

Learning from Timestamped Data

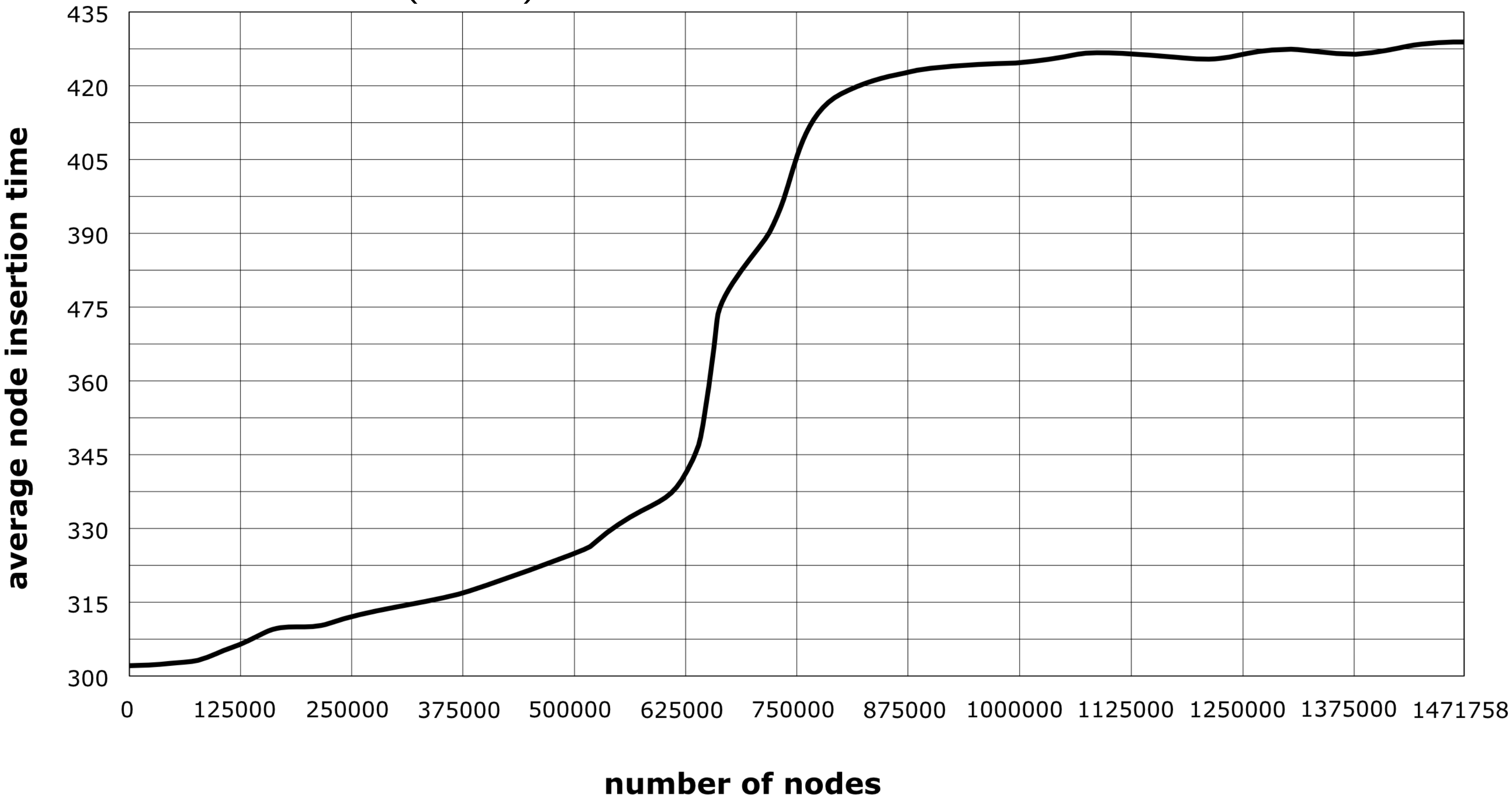
- One source of information concerning these hidden relations is timestamped news, of the kind available in the University of Washington NewsSpike corpus of 0.5M newswire articles (Zhang and Weld, 2013), and our own NewsCrawl.
- In such data, we find that statements that so-and-so *is visiting*, *is in* and the perfect *has arrived in* such and such a place, occur in stories with the same datestamp, whereas *is arriving*, *is on her way to*, occur in preceding stories, while *has left, is on her way back from, returned*, etc. occur in later ones.
- This information provides a basis for inference that *visiting* entails *being in*, that the latter is the consequent state of *arriving*, and that *arrival* and *departure* coincide with the beginning and end of the progressive state of *visiting*.

Needs new datasets for evaluation.

Building Knowledge Graphs from Text

- We would like to interrogate huge databases such as the Google knowledge graphs, a.k.a. Semantic Nets (Reddy *et al.*, 2014)
- There is a mismatch between the semantics delivered by parsers and the language of the knowledge graph.
- So let's build our own knowledge graph using the clustered entailment semantics of the parser, so that we can query it directly in natural language.
- This is a potentially a much bigger graph than the Entailment Graph.
- We will need techniques to limit exponential growth in the costs of loading and interrogating this graph.
- Pilot experiments by Harrington and Clark (2009); Lao *et al.* (2012); Szubert and Steedman (2019) suggest this can be done by spreading activation (Collins and Loftus, 1975).

- Szubert and Steedman (2019):



From Entailment Graph to Knowledge Graph

- We have replicated the spreading activation method of Harrington's AskNET and evaluated in comparison with graph convolution.
- We have identified improved methods for node identification in growing the Knowledge Graph, using both graph embeddings (GraphSAGE) and word embeddings (ELMo).
- We have shown a 50% reduction in errors both from wrong mergers of nodes and failure to make correct mergers over the AskNET Baseline (Szubert and Steedman, 2019).
- We are currently conducting experiments to show that building and interrogating the graph using entailment-based form-independent paraphrase-cluster semantics improves question answering over AskNET's form-dependent DRS semantics.

Generalizing to Other Languages

- Since our semantics is form-independent, it is also potentially language independent.
- We can therefore integrate relations and entailments mined from text in other languages into the same entailment graph to improve QA and SMT.
- In parallel, we are developing a similar pipeline for German using the Stanford Universal Dependencies (UD) Parser.
- A pilot study (Lewis and Steedman, 2013b) shows that this should be done by first building monolingual entailment graphs, and then aligning and merging nodes.
- We are interested in generalizing this to other languages with UD corpora.

Thanks!

- To: ERC Advanced Fellowship SEMANTAX; ARC Discovery grant DP160102156; Huawei/Edinburgh Research; Google Faculty Award; Bloomberg L.P. Gift Award.

References

Berant, Jonathan, Alon, Noga, Dagan, Ido, and Goldberger, Jacob, 2015. "Efficient Global Learning of Entailment Graphs." *Computational Linguistics* 42:221–263.

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.

Carnap, Rudolf, 1952. "Meaning Postulates." *Philosophical Studies* 3:65–73. reprinted as Carnap, 1956:222-229.

Carnap, Rudolf (ed.), 1956. *Meaning and Necessity*. Chicago: University of Chicago Press, second edition.

Collins, Allan and Loftus, Elizabeth, 1975. "A Spreading Activation Theory of Semantic Processing." *Psychological Review* 82:407–428.

Dettmers, Tim, Minervini, Pasquale, Stenetorp, Pontus, and Riedel, Sebastian, 2018. "Convolutional 2D Knowledge Graph Embeddings." In *32nd National Conference on Artificial Intelligence*. AAAI Publications, 1811–1818.

Fodor, Jerry, 1975. *The Language of Thought*. Cambridge, MA: Harvard.

Geffet, Maayan and Dagan, Ido, 2005. "The Distributional Inclusion Hypotheses and Lexical Entailment." In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*. ACL, 107–114.

Gulordava, Kristina, Bojanowski, Piotr, Grave, Edouard, Linzen, Tal, and Baroni, Marco, 2018. "Colorless Green Recurrent Networks Dream Hierarchically." In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*. 1195–1205.

Harrington, Brian and Clark, Stephen, 2009. "ASKNet: Creating and Evaluating Large Scale Integrated Semantic Networks." *International Journal of Semantic Computing* 2:343–364.

Holt, Xavier, 2018. *Probabilistic models of relational implication*. Master's thesis, Macquarie University, Sydney.

Hosseini, Javad, 2020. *Learning Typed Entailment Graphs from Text*. Ph.D. thesis, University of Edinburgh.

Hosseini, Javad, Chambers, Nathaniel, Reddy, Siva, Ricketts-Holt, Xavier, Cohen, Shay, Johnson, Mark, and Steedman, Mark, 2018. "Learning Typed Entailment Graphs with Global Soft Constraints." *Transactions of the Association for Computational Linguistics* 6:703–718.

Hosseini, Javad, Cohen, Shay, Johnson, Mark, and Steedman, Mark, 2019. "Duality of Link Prediction and Entailment Graph Induction." In *Proceedings of the 57th Annual Conference of the Association for Computational Linguistics (long papers)*. ACL, 4736–4746.

Hosseini, Javad, Cohen, Shay, Johnson, Mark, and Steedman, Mark, submitted. "Contextual Link Prediction to Learn Relational Entailment Graphs." In *2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. ACL, submitted.

Lao, Ni, Subramanya, Amarnag, Pereira, Fernando, and Cohen, William, 2012. "Reading the Web with Learned Syntactic-Semantic Inference Rules." In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*. ACL, 1017–1026.

Levy, Omer and Dagan, Ido, 2016. "Annotating Relation Inference in Context via Question Answering." In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. ACL, 249–255.

Lewis, Mike, 2015. *Combined Distributional and Logical Semantics*. Ph.D. thesis, University of Edinburgh.

Lewis, Mike 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 Conference on Empirical Methods in Natural Language Processing*. ACL, 681–692.

Lewis, Mike and Steedman, Mark, 2014. “Combining Formal and Distributional Models of Temporal and Intensional Semantics.” In *Proceedings of the ACL Workshop on Semantic Parsing*. Baltimore, MD: ACL, 28–32. Google Exceptional Submission Award.

McKenna, Nick, Guillou, Liane, Hosseini, Mohammad Javad, Bijl de Vroe, Sander, Johnson, Mark, and Steedman, Mark, submitted. "Multivalent Entailment Graphs for Question Answering." In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Mexico City: ACL, submitted.

Moens, Marc and Steedman, Mark, 1988. "Temporal Ontology and Temporal Reference." *Computational Linguistics* 14:15–28. reprinted in Inderjeet Mani, James Pustejovsky, and Robert Gaizauskas (eds.) *The Language of Time: A Reader*. Oxford University Press, 93-114.

Ono, Masataka, Miwa, Makoto, and Sasaki, Yutaka, 2015. "Word Embedding-Based Antonym Detection using Thesauri and Distributional Information." In *Proceedings of the 2015 Conference of the North American Chapter of the*

Association for Computational Linguistics: Human Language Technologies. 984–989.

Pustejovsky, James, 1991. "The Generative Lexicon." *Computational Linguistics* 17:409–441.

Reddy, Siva, Lapata, Mirella, and Steedman, Mark, 2014. "Large-scale Semantic Parsing without Question-Answer Pairs." *Transactions of the Association for Computational Linguistics* 2:377–392.

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.

Szubert, Ida and Steedman, Mark, 2019. "Node Embeddings for Graph Merging: Case of Knowledge-Graph Construction." In *Proceedings of the Thirteenth Workshop on Graph-Based Methods for Natural Language Processing (TextGraphs-13)*. 172–176.

Zhang, Congle and Weld, Daniel, 2013. "Harvesting Parallel News Streams to Generate Paraphrases of Event Relations." In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Seattle: ACL, 1776–1786.