Linguistic Semantics and Contemporary NLI

Mark Steedman

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

July 14th 2020

School of informatics

Outline

- I: The Problem of Linguistic Form-dependence in Semantics for NLP.
- II: The Proposal for a Form-Independent Semantics.
- III: Results So Far (Hosseini *et al.*, 2018, 2019).
- IV: Work in Progress Towards Form-Independence

I: The Problem of Content

- 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, so that we can ask them questions like "What are recordings by Bill Evans without Fender-Rhodes piano", and get a more helpful answer than the following?

Recordings by Bill Evans that do not use Fender-Rhodes piano

All | Shopping | Images | Videos | News | More | Settings | Tools

About 1,460,000 results (0.66 seconds)

www.billevanswebpages.com › rhodespiece

Bill Evans Webpages: Evans and the Fender-Rhodes

Years later, Ray Charles **use** the Wurlitzer **electric piano** on "What I Say". ... The 88-note models **were not** released until 1970, the same year as the Stage and the Suitcase **Pianos** ... I was there primarily to hear some of his latest **recordings**.

forums.stevehoffman.tv › ... › Music Corner

On what recording did Bill Evans first play the Fender Rhodes ...

22 Sep 2008 - He plays it on a few of the tracks on the Intuition album **recorded** with Eddie ... My favorite example of Evans' **electric piano** playing is on The **Bill Evans** ... They gave keyboard players new tone colors **without** making you ...

ep-forum.com › smf

Fender Rhodes Odyssey: Bill Evans' From Left To Right 1970

4 May 2006 - In the '60s the jazz **pianist Bill Evans** would occasionally record an ... instrumental **albums** with arranger Claus Ogerman, even **without** those ...

books.google.co.uk › books

Essential Jazz - Google Books Result

The instrument was a **Fender Rhodes** electric **piano**. Davis's first **recording** with **Fender Rhodes** was the 1968 release Miles in the Sky, and his continued **use** ... Here we look at **Bill Evans**, Herbie Hancock, Chick Corea, and Keith Jarrett. ... Now, I **don't** know how obvious that

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: *Has Verizon bought 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 doesn't own Yahoo ("No")
6. Yahoo may be sold to Verizon. ("Maybe")
7. Verizon will buy Yahoo or Yazoo. ("Maybe not")

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

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, Obama—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
- Redefine the parser semantics in terms of entailments and paraphrases, and reparse and index the entire text for IR.

(There is another approach, based on vectors and linear algebraic composition.)

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.)

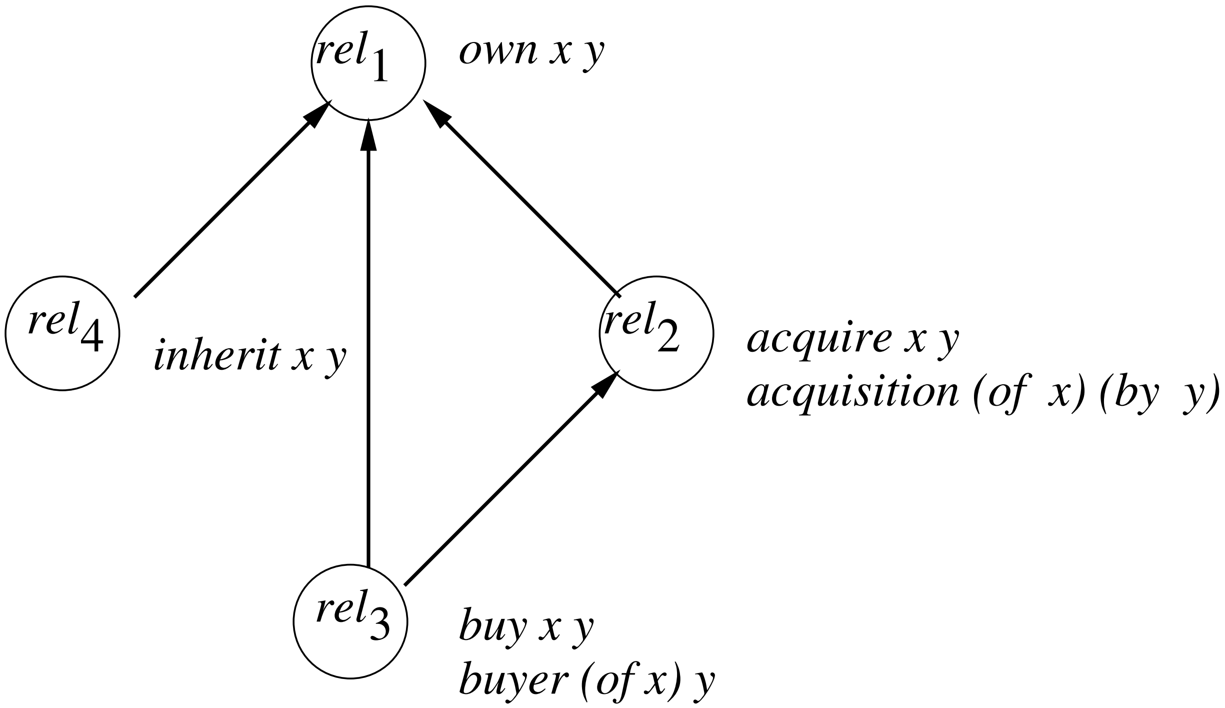
Global Entailments

- The local entailment probabilities are used to construct an entailment graph, with the global constraint that the graph must be closed under transitivity (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.

Entailment graph



- A simplified entailment graph for relations between people and property.

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	:=	$(S\backslash NP)/NP$	$: \lambda x \lambda y.rel_2 x y$
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*

Applications

1. Question Answering.

2. Reranking machine Summarization.

3. Building Knowledge Graphs from text.

III: Progress So Far (Hosseini *et al.*, 2018)

- We have trained an entailment graph on the NewsSpike corpus
  - 0.5M multiply-sourced news articles over 2 months, 20M sentences.
  - 29M binary relation tokens extracted using the CCG parser.
- We have built a working typed global entailment graph, collapsing paraphrase cliques
  - 101K relation types
  - 346 local typed entailment subgraphs
  - 23 subgraphs with more than 1K nodes e.g. Person$\times$Location, Location$\times$Thing, Org$\times$Org, etc.
  - 7 subgraphs with more than 10K nodes
- We redefined the semantics and have built a scalable knowledge graph

Idioms, Metaphors, and Presuppositions

- Idioms are found just like any other typed entailment:
  - $keep\_tabs\_on\,(\#government\_agency, \#thing) \models' s\_surveillance\_of\,(\#government\_agenc$
- 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:
  - $manage\_to(\#person, \#event) \models try\_to\,(\#person, \#event)$
  - $\neg manage\_to(\#person, \#event) \models try\_to\,(\#person, \#event)$

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 learn 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 depend on transitivity, but instead uses two global soft constraints.
  - Our method scales to more than 100K nodes.

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.
  - 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).

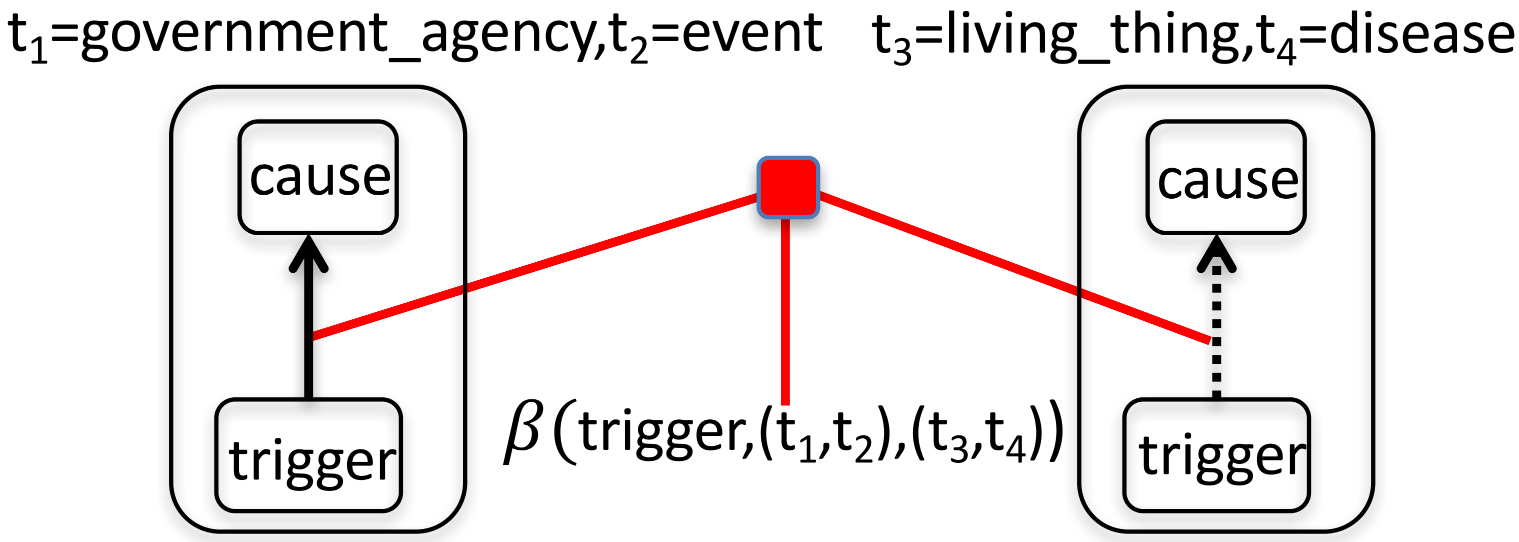
Global Soft Constraint 1: Cross Graph Transfer

- 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 Tramsfer.

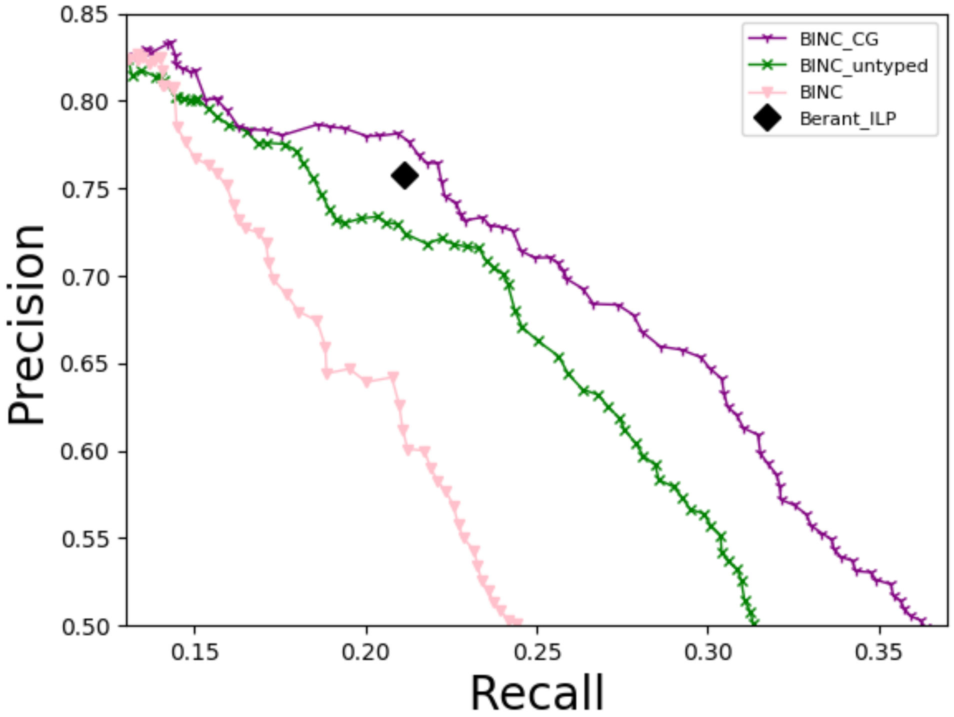
Global Soft Constraint 1: Cross Graph



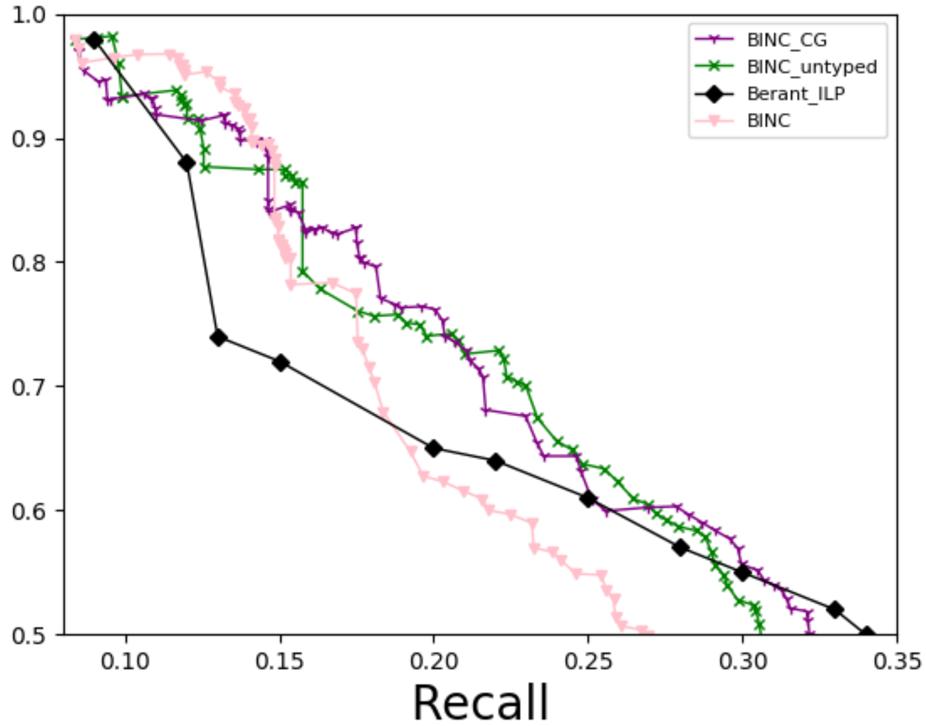
- $0 \leq \beta(.) \leq 1$ determines how much different graphs are related and will be learned jointly.

Adding Cross-Graph Transfer Soft Constraints

Levy/Holt's dataset

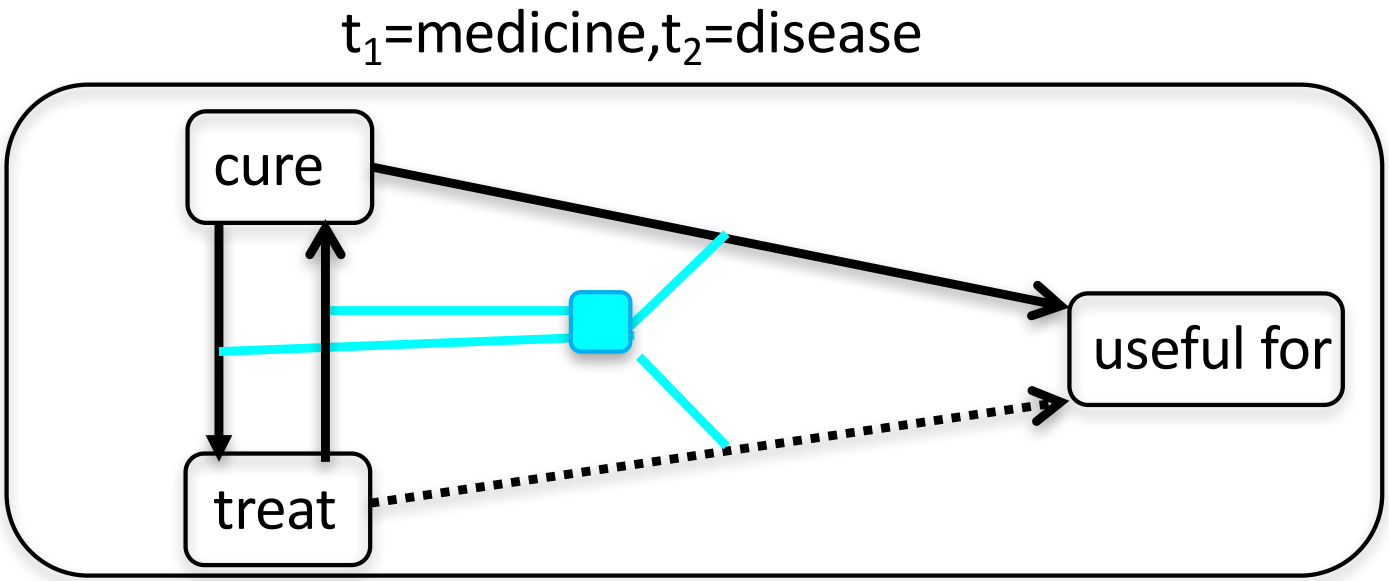


Berant's dataset

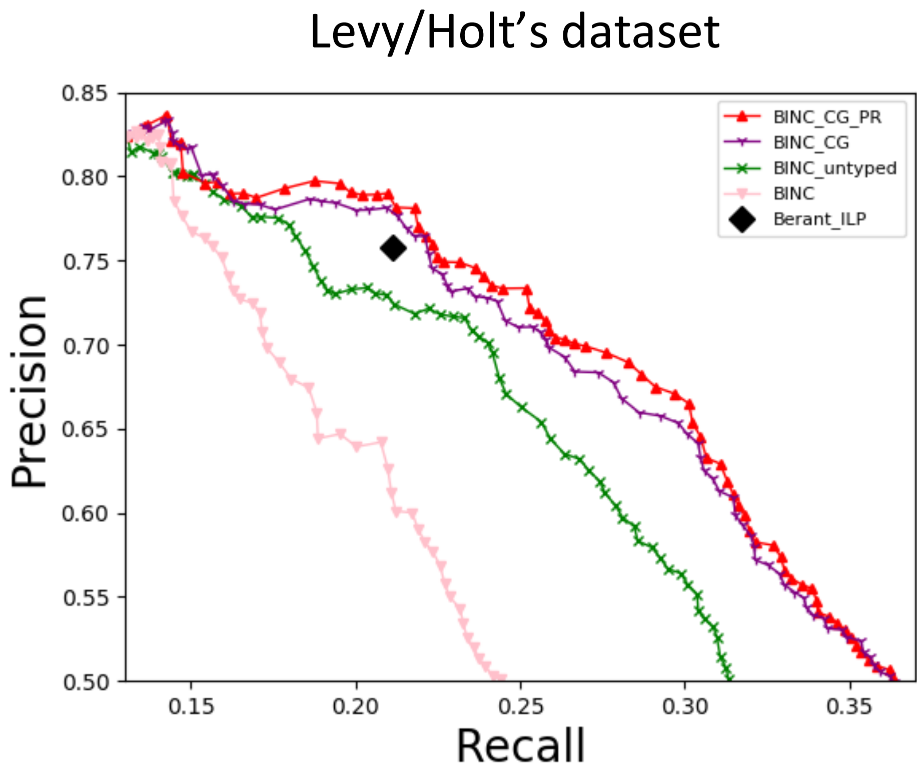


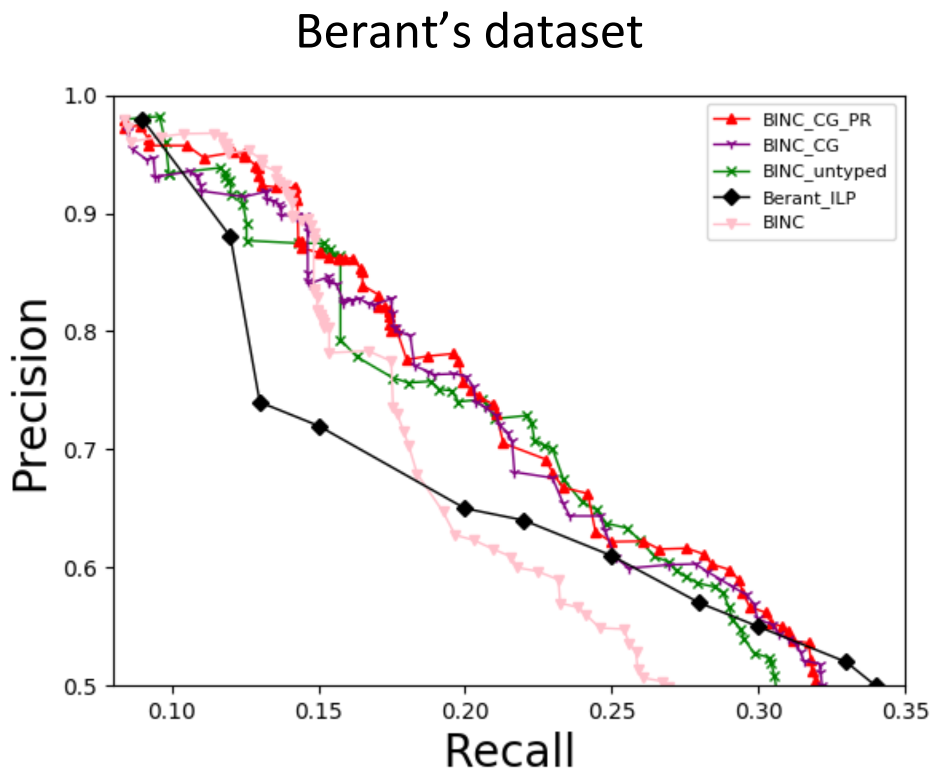
Global Soft Constraint 2: Paraphrase Resolution

- 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





Results for Various Similarity Measures

- Area under precision-recall curve (precision $> .5$) for different variants of distributional similarities
  - Boldfaced results are statistically significant

	local	untyped	CG	CG_PR
	Levy/Holt's dataset			
BInc	.076	.127	.162	**.165**
Lin	.074	.120	**.151**	**.149**
Weed	.073	.115	**.149**	**.147**
	Berant's dataset			
BInc	.138	.167	**.177**	**.179**
Lin	.147	.158	.186	**.189**
Weed	.146	.154	.184	**.187**

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

Extrinsic Evaluation

- We have carried out a limited extrinsic evaluation on an answer selection task on the NewsQA test set of text-questions (Trischler *et al.*, 2017), achieving a 1-2% increase in performance over a baseline inverse sentence frequency (ISF) measure (cf. Narayan *et al.*, 2018).

	ACC	MRR	MAP
ISF	.3618	.4899	.4857
ISF+ENT	**.3761**	**.5006**	**.4963**

Table 1: Answer selection on NewsQA

- NewsQA example:

  Question: Who praised Mitt Romney's credentials?
  Selected sentence: The board hailed Romney for his solid credentials

Does BERTology help? (Hosseini *et al.* (2019))

- Rather than guessing entailment relations based on directional similarity of vectors of named-entity pairs, our colleagues frequently ask us, why not 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) show that embeddings-based methods for link-prediction in existing knowledge graphs (Riedel *et al.*, 2013) do improve the entailment graph.
- And vice versa—access to the entailment graph improves link-prediction.
- Embeddings seem to learn information that is complementary to machine-reading.

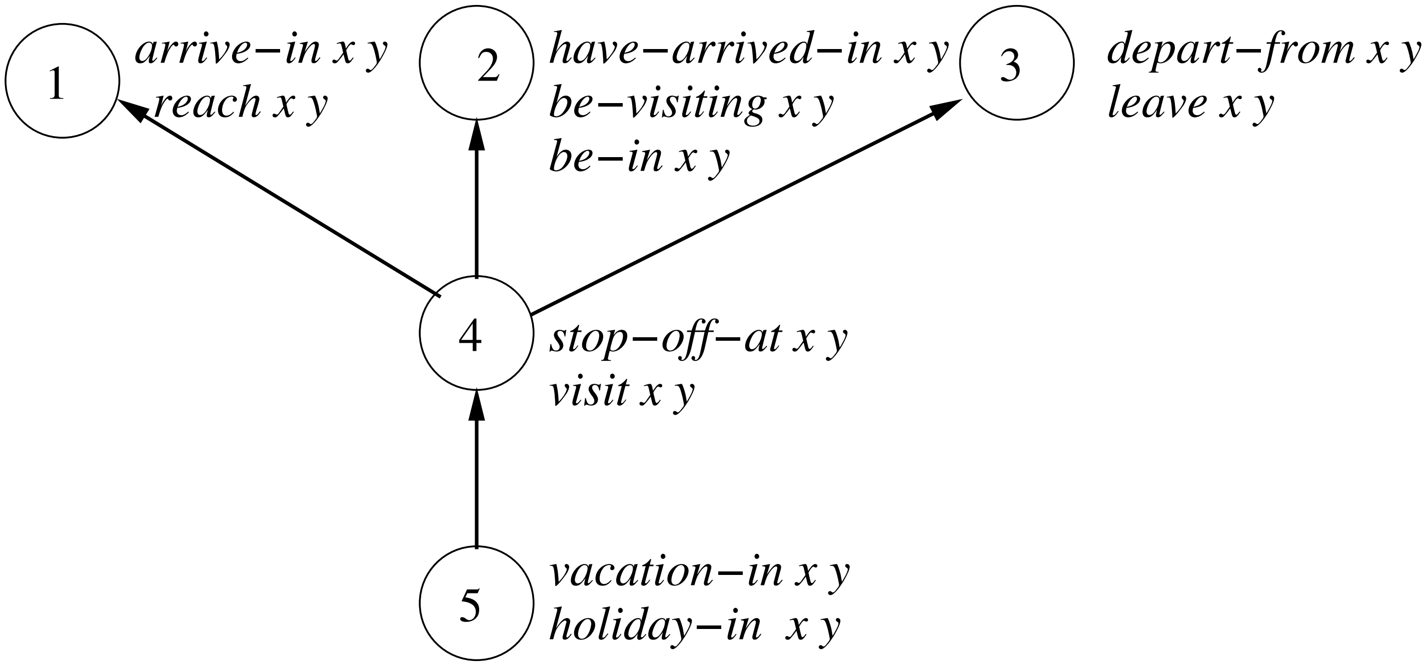
IV: Future Work

- Improve relation extraction from text to include negation, auxiliary verbs, implicative verbs, temporal relations, etc.
- Refine method for building entailment graphs.
- Define Form-independent clustered entailment-based Semantic Parser.
- Use it to build a Large Knowledge Graph using form-independent semantic representations from text.

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.

Form Independent Temporal Semantics



- A simplified entailment graph for relations over events does not capture relations of causation and temporal sequence.

Learning from Timestamped Data

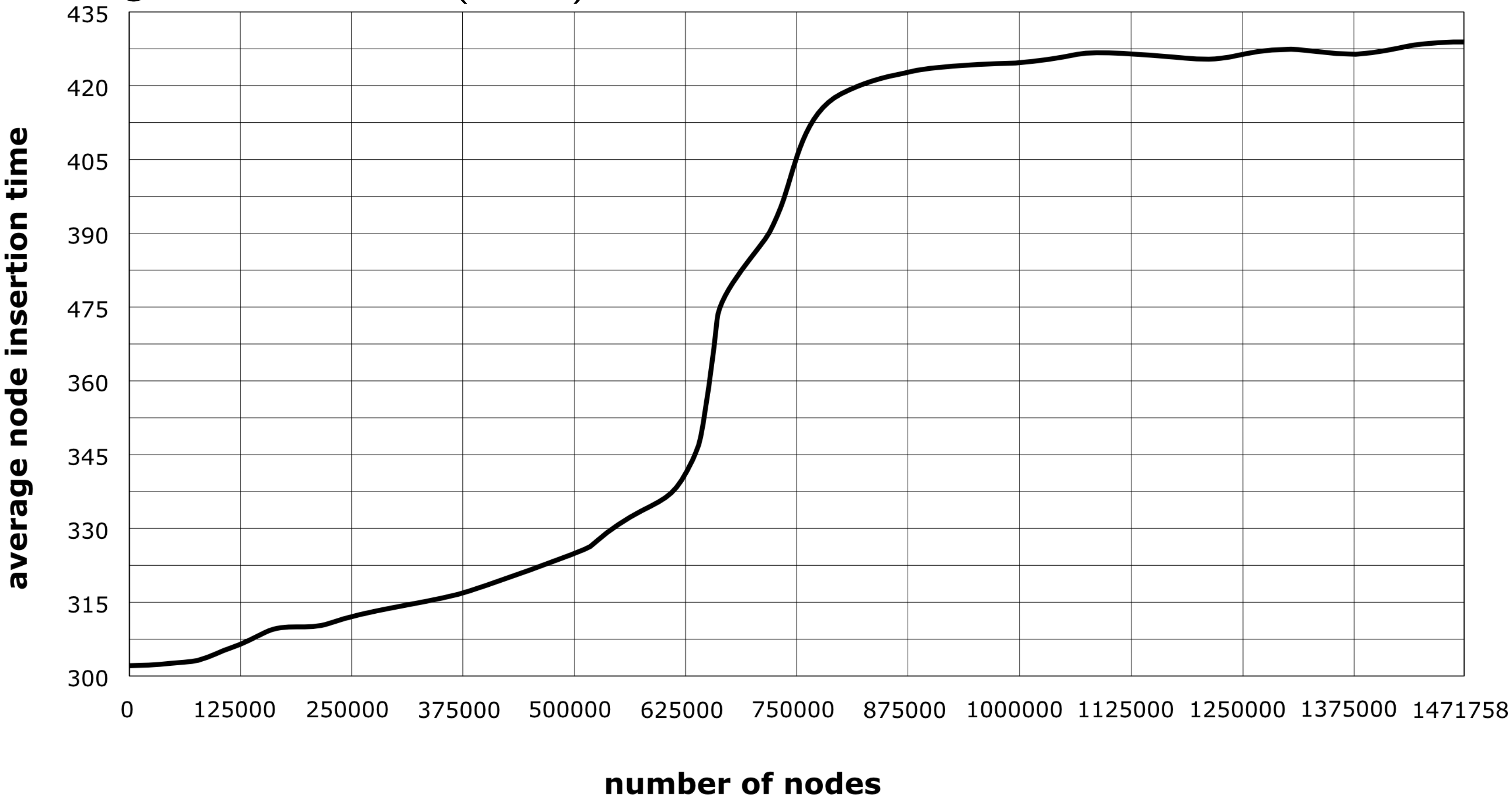
- One source of information concerning these hidden relations is timestamped news, of the kind available in the University of Washington NewsSpike corpus of 0.5M newswire articles (Zhang and Weld, 2013).
- In such data, we find that statements that so-and-so *isvisiting*, *is in* and the perfect *has 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 Knowledge 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) suggest this can be done by spreading activation (Collins and Loftus, 1975).

- Cf. Harrington and Clark (2009):



From Entailment Graph to Knowledge Graph

- We have replicated the spreading activation method of Harrington's AskNET and evaluated in comparison with graph convilution.
- 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.

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.

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

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, 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.

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.

Narayan, Shashi, Cardenas, Ronald, Papasarantopoulos, Nikos, Cohen, Shay, Lapata, Mirella, Yu, Jiangsheng, and Chang, Yi, 2018. "Document Modeling with External Attention for Sentence Extraction." In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. ACL, 2020–2030.

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.

Trischler, Adam, Wang, Tong, Yuan, Xingdi, Harris, Justin, Sordoni, Alessandro, Bachman, Philip, and Suleman, Kaheer, 2017. "NewsQA: A Machine Comprehension Dataset." In *Proceedings of the 2nd Workshop on Representation Learning for NLP*. ACL, 191–200.

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.