

# A Structured Distributional Semantic Model for Event Co-reference

Kartik Goyal\*    Sujay Kumar Jauhar\*    Huiying Li\*  
Mrinmaya Sachan\*    Shashank Srivastava\*    Eduard Hovy

Language Technologies Institute  
School of Computer Science  
Carnegie Mellon University

{kartikgo, sjauhar, huiyingl, mrinmays, shashans, hovy}@cs.cmu.edu

## Abstract

In this paper we present a novel approach to modelling distributional semantics that represents meaning as distributions over relations in syntactic neighborhoods. We argue that our model approximates meaning in compositional configurations more effectively than standard distributional vectors or bag-of-words models. We test our hypothesis on the problem of judging event coreferentiality, which involves compositional interactions in the predicate-argument structure of sentences, and demonstrate that our model outperforms both state-of-the-art window-based word embeddings as well as simple approaches to compositional semantics previously employed in the literature.

## 1 Introduction

Distributional Semantic Models (DSM) are popular in computational semantics. DSMs are based on the hypothesis that the meaning of a word or phrase can be effectively captured by the distribution of words in its neighborhood. They have been successfully used in a variety of NLP tasks including information retrieval (Manning et al., 2008), question answering (Tellex et al., 2003), word-sense discrimination (Schütze, 1998) and disambiguation (McCarthy et al., 2004), semantic similarity computation (Wong and Raghavan, 1984; McCarthy and Carroll, 2003) and selectional preference modeling (Erk, 2007).

A shortcoming of DSMs is that they ignore the syntax within the context, thereby reducing the distribution to a bag of words. Composing the

distributions for “Lincoln”, “Booth”, and “killed” gives the same result regardless of whether the input is “Booth killed Lincoln” or “Lincoln killed Booth”. But as suggested by Pantel and Lin (2000) and others, modeling the distribution over preferential attachments for each syntactic relation separately yields greater expressive power. Thus, to remedy the bag-of-words failing, we extend the generic DSM model to several relation-specific distributions over syntactic neighborhoods. In other words, one can think of the Structured DSM (SDSM) representation of a word/phrase as several vectors defined over the same vocabulary, each vector representing the word’s selectional preferences for its various syntactic arguments.

We argue that this representation not only captures individual word semantics more effectively than the standard DSM, but is also better able to express the semantics of compositional units. We prove this on the task of judging event coreference.

Experimental results indicate that our model achieves greater predictive accuracy on the task than models that employ weaker forms of composition, as well as a baseline that relies on state-of-the-art window based word embeddings. This suggests that our formalism holds the potential of greater expressive power in problems that involve underlying semantic compositionality.

## 2 Related Work

Next, we relate and contrast our work to prior research in the fields of Distributional Vector Space Models, Semantic Compositionality and Event Co-reference Resolution.

### 2.1 DSMs and Compositionality

The underlying idea that “a word is characterized by the company it keeps” was expressed by Firth

---

\*Equally contributing authors

(1957). Several works have defined approaches to modelling context-word distributions anchored on a target word, topic, or sentence position. Collectively these approaches are called Distributional Semantic Models (DSMs).

While DSMs have been very successful on a variety of tasks, they are not an effective model of semantics as they lack properties such as compositionality or the ability to handle operators such as negation. In order to model a stronger form of semantics, there has been a recent surge in studies that phrase the problem of DSM compositionality as one of vector composition. These techniques derive the meaning of the combination of two words  $a$  and  $b$  by a single vector  $c = f(a, b)$ . Mitchell and Lapata (2008) propose a framework to define the composition  $c = f(a, b, r, K)$  where  $r$  is the relation between  $a$  and  $b$ , and  $K$  is the additional knowledge used to define composition. While this framework is quite general, the actual models considered in the literature tend to disregard  $K$  and  $r$  and mostly perform component-wise addition and multiplication, with slight variations, of the two vectors. To the best of our knowledge the formulation of composition we propose is the first to account for both  $K$  and  $r$  within this compositional framework.

Dinu and Lapata (2010) and Séaghdha and Korhonen (2011) introduced a probabilistic model to represent word meanings by a latent variable model. Subsequently, other high-dimensional extensions by Rudolph and Giesbrecht (2010), Baroni and Zamparelli (2010) and Grefenstette et al. (2011), regression models by Guevara (2010), and recursive neural network based solutions by Socher et al. (2012) and Collobert et al. (2011) have been proposed. However, these models do not efficiently account for structure.

Pantel and Lin (2000) and Erk and Padó (2008) attempt to include syntactic context in distributional models. A quasi-compositional approach was attempted in Thater et al. (2010) by a combination of first and second order context vectors. But they do not explicitly construct phrase-level meaning from words which limits their applicability to real world problems. Furthermore, we also include structure into our method of composition. Prior work in structure aware methods to the best of our knowledge are (Weisman et al., 2012) and (Baroni and Lenci, 2010). However, these methods do not explicitly model composition.

## 2.2 Event Co-reference Resolution

While automated resolution of entity coreference has been an actively researched area (Haghighi and Klein, 2009; Stoyanov et al., 2009; Raghunathan et al., 2010), there has been relatively little work on event coreference resolution. Lee et al. (2012) perform joint cross-document entity and event coreference resolution using the two-way feedback between events and their arguments. We, on the other hand, attempt a slightly different problem of making co-referentiality judgements on event-coreference candidate pairs.

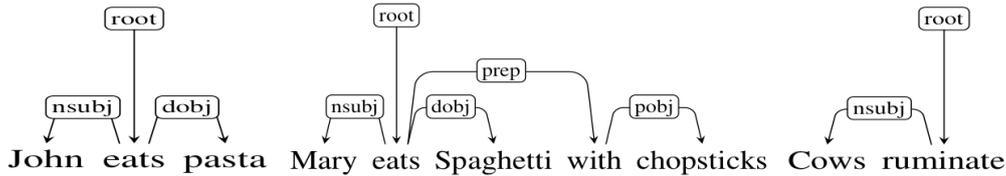
## 3 Structured Distributional Semantics

In this paper, we propose an approach to incorporate structure into distributional semantics (more details in Goyal et al. (2013)). The word distributions drawn from the context defined by a set of relations anchored on the target word (or phrase) form a set of vectors, namely a matrix for the target word. One axis of the matrix runs over all the relations and the other axis is over the distributional word vocabulary. The cells store word counts (or PMI scores, or other measures of word association). Note that collapsing the rows of the matrix provides the standard dependency based distributional representation.

### 3.1 Building Representation: The PropStore

To build a lexicon of SDSM matrices for a given vocabulary we first construct a proposition knowledge base (the PropStore) created by parsing the Simple English Wikipedia. Dependency arcs are stored as 3-tuples of the form  $\langle w_1, r, w_2 \rangle$ , denoting an occurrence of words  $w_1$ , word  $w_2$  related by  $r$ . We also store sentence indices for triples as this allows us to achieve an intuitive technique to achieve compositionality. In addition to the words' surface-forms, the PropStore also stores their POS tags, lemmas, and Wordnet supersenses. This helps to generalize our representation when surface-form distributions are sparse.

The PropStore can be used to query for the expectations of words, supersenses, relations, etc., around a given word. In the example in Figure 1, the query  $(SST(W_1) = \text{verb.consumption}, ?, \text{dobj})$  i.e. "what is consumed" might return expectations [pasta:1, spaghetti:1, mice:1 ...]. Relations and POS tags are obtained using a dependency parser Tratz and Hovy (2011), supersense tags using sst-light Ciaramita and Altun (2006), and lemmas us-



- 1) { (John/NNP/john/Noun.person , nsubj, eats/VBG/eat/verb.consumption ),  
 (eats/VBG/eat/verb.consumption, dobj, pasta/NN/pasta/noun.food) }
- 2) { (Mary/NNP/mary/Noun.person), nsubj, (eats/VBG/eat/verb.consumption) ... }
- 3) { (Cows/NNP/cow/Noun.animal),nsubj,(ruminates/VBG/ruminates/verb.consumption) }

Figure 1: Sample sentences & triples

ing Wordnet Fellbaum (1998).

### 3.2 Mimicking Compositionality

For representing intermediate multi-word phrases, we extend the above word-relation matrix symbolism in a bottom-up fashion using the PropStore. The combination hinges on the intuition that when lexical units combine to form a larger syntactically connected phrase, the representation of the phrase is given by its own distributional neighborhood within the embedded parse tree. The distributional neighborhood of the net phrase can be computed using the PropStore given syntactic relations anchored on its parts. For the example in Figure 1, we can compose  $SST(w_1) = \text{Noun.person}$  and  $\text{Lemma}(W_1) = \text{eat}$  appearing together with a nsubj relation to obtain expectations around “people eat” yielding [pasta:1, spaghetti:1 ...] for the *object* relation, [room:2, restaurant:1 ...] for the *location* relation, etc. Larger phrasal queries can be built to answer queries like “What do people in China eat with?”, “What do cows do?”, etc. All of this helps us to account for both relation  $r$  and knowledge  $K$  obtained from the PropStore within the compositional framework  $c = f(a, b, r, K)$ .

The general outline to obtain a composition of two words is given in Algorithm 1, which returns the distributional expectation around the composed unit. Note that the entire algorithm can conveniently be written in the form of database queries to our PropStore.

---

#### Algorithm 1 ComposePair( $w_1, r, w_2$ )

---

- $M_1 \leftarrow \text{queryMatrix}(w_1)$  (1)
  - $M_2 \leftarrow \text{queryMatrix}(w_2)$  (2)
  - SentIDs  $\leftarrow M_1(r) \cap M_2(r)$  (3)
  - return  $((M_1 \cap \text{SentIDs}) \cup (M_2 \cap \text{SentIDs}))$  (4)
- 

For the example “noun.person nsubj eat”, steps

(1) and (2) involve querying the PropStore for the individual tokens, noun.person and eat. Let the resulting matrices be  $M_1$  and  $M_2$ , respectively. In step (3), SentIDs (sentences where the two words appear with the specified relation) are obtained by taking the intersection between the nsubj component vectors of the two matrices  $M_1$  and  $M_2$ . In step (4), the entries of the original matrices  $M_1$  and  $M_2$  are intersected with this list of common SentIDs. Finally, the resulting matrix for the composition of the two words is simply the union of all the relationwise intersected sentence IDs. Intuitively, through this procedure, we have computed the expectation around the words  $w_1$  and  $w_2$  when they are connected by the relation “r”.

Similar to the two-word composition process, given a parse subtree  $T$  of a phrase, we obtain its matrix representation of empirical counts over word-relation contexts (described in Algorithm 2). Let the  $E = \{e_1 \dots e_n\}$  be the set of edges in  $T$ ,  $e_i = (w_{i1}, r_i, w_{i2}) \forall i = 1 \dots n$ .

---

#### Algorithm 2 ComposePhrase( $T$ )

---

- SentIDs  $\leftarrow$  All Sentences in corpus
  - for**  $i = 1 \rightarrow n$  **do**
  - $M_{i1} \leftarrow \text{queryMatrix}(w_{i1})$
  - $M_{i2} \leftarrow \text{queryMatrix}(w_{i2})$
  - SentIDs  $\leftarrow$  SentIDs  $\cap (M_{i1}(r_i) \cap M_{i2}(r_i))$
  - end for**
  - return  $((M_{11} \cap \text{SentIDs}) \cup (M_{12} \cap \text{SentIDs}) \dots \cup (M_{n1} \cap \text{SentIDs}) \cup (M_{n2} \cap \text{SentIDs}))$
- 

The phrase representations becomes sparser as phrase length increases. For this study, we restrict phrasal query length to a maximum of three words.

### 3.3 Event Coreferentiality

Given the SDSM formulation and assuming no sparsity constraints, it is possible to calculate

SDSM matrices for composed concepts. However, are these correct? Intuitively, if they truly capture semantics, the two SDSM matrix representations for “Booth assassinated Lincoln” and “Booth shot Lincoln with a gun” should be (almost) the same. To test this hypothesis we turn to the task of predicting whether two event mentions are coreferent or not, even if their surface forms differ. It may be noted that this task is different from the task of full event coreference and hence is not directly comparable to previous experimental results in the literature. Two mentions generally refer to the same event when their respective actions, agents, patients, locations, and times are (almost) the same. Given the non-compositional nature of determining equality of locations and times, we represent each event mention by a triple  $\mathbf{E} = (e, a, p)$  for the event, agent, and patient.

In our corpus, most event mentions are verbs. However, when nominalized events are encountered, we replace them by their verbal forms. We use SRL Collobert et al. (2011) to determine the agent and patient arguments of an event mention. When SRL fails to determine either role, its empirical substitutes are obtained by querying the PropStore for the most likely word expectations for the role. It may be noted that the SDSM representation relies on syntactic dependency relations. Hence, to bridge the gap between these relations and the composition of semantic role participants of event mentions we empirically determine those syntactic relations which most strongly co-occur with the semantic relations connecting events, agents and patients. The triple  $(e, a, p)$  is thus the composition of the triples  $(a, relationset_{agent}, e)$  and  $(p, relationset_{patient}, e)$ , and hence a complex object. To determine equality of this complex composed representation we generate three levels of progressively simplified event constituents for comparison:

**Level 1: Full Composition:**

$$M_{full} = ComposePhrase(e, a, p).$$

**Level 2: Partial Composition:**

$$M_{part:EA} = ComposePair(e, r, a)$$

$$M_{part:EP} = ComposePair(e, r, p).$$

**Level 3: No Composition:**

$$M_E = queryMatrix(e)$$

$$M_A = queryMatrix(a)$$

$$M_P = queryMatrix(p)$$

To judge coreference between events  $\mathbf{E1}$  and  $\mathbf{E2}$ , we compute pair-

wise similarities  $\text{Sim}(M1_{full}, M2_{full})$ ,  $\text{Sim}(M1_{part:EA}, M2_{part:EA})$ , etc., for each level of the composed triple representation. Furthermore, we vary the computation of similarity by considering different levels of granularity (lemma, SST), various choices of distance metric (Euclidean, Cityblock, Cosine), and score normalization techniques (Row-wise, Full, Column-collapsed). This results in 159 similarity-based features for every pair of events, which are used to train a classifier to decide conference.

## 4 Experiments

We evaluate our method on two datasets and compare it against four baselines, two of which use window based distributional vectors and two that employ weaker forms of composition.

### 4.1 Datasets

**IC Event Coreference Corpus:** The dataset (Hovy et al., 2013), drawn from 100 news articles about violent events, contains manually created annotations for 2214 pairs of co-referent and non-coreferent events each. Where available, events’ semantic role-fillers for *agent* and *patient* are annotated as well. When missing, empirical substitutes were obtained by querying the PropStore for the preferred word attachments.

**EventCorefBank (ECB) corpus:** This corpus (Bejan and Harabagiu, 2010) of 482 documents from Google News is clustered into 45 topics, with event coreference chains annotated over each topic. The event mentions are enriched with semantic roles to obtain the canonical event structure described above. Positive instances are obtained by taking pairwise event mentions within each chain, and negative instances are generated from pairwise event mentions across chains, but within the same topic. This results in 11039 positive instances and 33459 negative instances.

### 4.2 Baselines

To establish the efficacy of our model, we compare SDSM against a purely window-based baseline (DSM) trained on the same corpus. In our experiments we set a window size of seven words. We also compare SDSM against the window-based embeddings trained using a recursive neural network (SENNA) (Collobert et al., 2011) on both datasets. SENNA embeddings are state-of-the-art for many NLP tasks. The second baseline uses

	IC Corpus				ECB Corpus			
	Prec	Rec	F-1	Acc	Prec	Rec	F-1	Acc
SDSM	<b>0.916</b>	0.929	<b>0.922</b>	<b>0.906</b>	0.901	0.401	<b>0.564</b>	<b>0.843</b>
Senna	0.850	0.881	0.865	0.835	0.616	<b>0.408</b>	0.505	0.791
DSM	0.743	0.843	0.790	0.740	0.854	0.378	0.524	0.830
MVC	0.756	<b>0.961</b>	0.846	0.787	<b>0.914</b>	0.353	0.510	0.831
AVC	0.753	0.941	0.837	0.777	0.901	0.373	0.528	0.834

Table 1: Cross-validation Performance on IC and ECB dataset

SENNA to generate level 3 similarity features for events’ individual words (agent, patient and action). As our final set of baselines, we extend two simple techniques proposed by (Mitchell and Lapata, 2008) that use element-wise addition and multiplication operators to perform composition. We extend it to our matrix representation and build two baselines AVC (element-wise addition) and MVC (element-wise multiplication).

### 4.3 Discussion

Among common classifiers, decision-trees (J48) yielded best results in our experiments. Table 1 summarizes our results on both datasets.

The results reveal that the SDSM model consistently outperforms DSM, SENNA embeddings, and the MVC and AVC models, both in terms of F-1 score and accuracy. The IC corpus comprises of domain specific texts, resulting in high lexical overlap between event mentions. Hence, the scores on the IC corpus are consistently higher than those on the ECB corpus.

The improvements over DSM and SENNA embeddings, support our hypothesis that syntax lends greater expressive power to distributional semantics in compositional configurations. Furthermore, the increase in predictive accuracy over MVC and AVC shows that our formulation of composition of two words based on the relation binding them yields a stronger form of compositionality than simple additive and multiplicative models.

Next, we perform an ablation study to determine the most predictive features for the task of event coreferentiality. The forward selection procedure reveals that the most informative attributes are the level 2 compositional features involving the agent and the action, as well as their individual level 3 features. This corresponds to the intuition that the agent and the action are the principal determiners for identifying events. Features involving the patient and level 1 features are least

useful. This is probably because features involving full composition are sparse, and not as likely to provide statistically significant evidence. This may change as our PropStore grows in size.

## 5 Conclusion and Future Work

We outlined an approach that introduces structure into distributed semantic representations gives us an ability to compare the identity of two representations derived from supposedly semantically identical phrases with different surface realizations. We employed the task of event coreference to validate our representation and achieved significantly higher predictive accuracy than several baselines.

In the future, we would like to extend our model to other semantic tasks such as paraphrase detection, lexical substitution and recognizing textual entailment. We would also like to replace our syntactic relations to semantic relations and explore various ways of dimensionality reduction to solve this problem.

## Acknowledgments

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions to improve the quality of the paper. This work was supported in part by the following grants: NSF grant IIS-1143703, NSF award IIS-1147810, DARPA grant FA87501220342.

## References

- Marco Baroni and Alessandro Lenci. 2010. Distributional memory: A general framework for corpus-based semantics. *Comput. Linguist.*, 36(4):673–721, December.
- Marco Baroni and Roberto Zamparelli. 2010. Nouns are vectors, adjectives are matrices: representing adjective-noun constructions in semantic space. In *Proceedings of the 2010 Conference on Empirical*

- Methods in Natural Language Processing*, EMNLP '10, pages 1183–1193, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Cosmin Adrian Bejan and Sanda Harabagiu. 2010. Unsupervised event coreference resolution with rich linguistic features. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, ACL '10, pages 1412–1422, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Massimiliano Ciaramita and Yasemin Altun. 2006. Broad-coverage sense disambiguation and information extraction with a supersense sequence tagger. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, EMNLP '06, pages 594–602, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. 2011. Natural language processing (almost) from scratch. *J. Mach. Learn. Res.*, 999888:2493–2537, November.
- Georgiana Dinu and Mirella Lapata. 2010. Measuring distributional similarity in context. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, EMNLP '10, pages 1162–1172, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Katrin Erk and Sebastian Padó. 2008. A structured vector space model for word meaning in context. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '08, pages 897–906, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Katrin Erk. 2007. A simple, similarity-based model for selectional preferences.
- Christiane Fellbaum. 1998. *WordNet: An Electronic Lexical Database*. Bradford Books.
- John R. Firth. 1957. A Synopsis of Linguistic Theory, 1930-1955. *Studies in Linguistic Analysis*, pages 1–32.
- Kartik Goyal, Sujay Kumar Jauhar, Mrinmaya Sachan, Shashank Srivastava, Huiying Li, and Eduard Hovy. 2013. A structured distributional semantic model: Integrating structure with semantics. In *Proceedings of the 1st Continuous Vector Space Models and their Compositionality Workshop at the conference of ACL 2013*.
- Edward Grefenstette, Mehrnoosh Sadrzadeh, Stephen Clark, Bob Coecke, and Stephen Pulman. 2011. Concrete sentence spaces for compositional distributional models of meaning. In *Proceedings of the Ninth International Conference on Computational Semantics*, IWCS '11, pages 125–134, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Emiliano Guevara. 2010. A regression model of adjective-noun compositionality in distributional semantics. In *Proceedings of the 2010 Workshop on GEometrical Models of Natural Language Semantics*, GEMS '10, pages 33–37, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Aria Haghighi and Dan Klein. 2009. Simple coreference resolution with rich syntactic and semantic features. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3 - Volume 3*, EMNLP '09, pages 1152–1161, Stroudsburg, PA, USA. Association for Computational Linguistics.
- E.H. Hovy, T. Mitamura, M.F. Verdejo, J. Araki, and A. Philpot. 2013. Events are not simple: Identity, non-identity, and quasi-identity. In *Proceedings of the 1st Events Workshop at the conference of the HLT-NAACL 2013*.
- Heeyoung Lee, Marta Recasens, Angel Chang, Mihai Surdeanu, and Dan Jurafsky. 2012. Joint entity and event coreference resolution across documents. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, EMNLP-CoNLL '12, pages 489–500, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. 2008. *Introduction to Information Retrieval*. Cambridge University Press, New York, NY, USA.
- Diana McCarthy and John Carroll. 2003. Disambiguating nouns, verbs, and adjectives using automatically acquired selectional preferences. *Comput. Linguist.*, 29(4):639–654, December.
- Diana McCarthy, Rob Koeling, Julie Weeds, and John Carroll. 2004. Finding predominant word senses in untagged text. In *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, ACL '04, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Jeff Mitchell and Mirella Lapata. 2008. Vector-based models of semantic composition. In *Proceedings of ACL-08: HLT*, pages 236–244.
- Patrick Pantel and Dekang Lin. 2000. Word-for-word glossing with contextually similar words. In *Proceedings of the 1st North American chapter of the Association for Computational Linguistics conference*, NAACL 2000, pages 78–85, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Karthik Raghunathan, Heeyoung Lee, Sudarshan Rangarajan, Nathanael Chambers, Mihai Surdeanu, Dan Jurafsky, and Christopher Manning. 2010. A multi-pass sieve for coreference resolution. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, EMNLP '10,

- pages 492–501, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Sebastian Rudolph and Eugenie Giesbrecht. 2010. Compositional matrix-space models of language. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, ACL '10, pages 907–916, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Hinrich Schütze. 1998. Automatic word sense discrimination. *Comput. Linguist.*, 24(1):97–123.
- Diarmuid Ó Séaghdha and Anna Korhonen. 2011. Probabilistic models of similarity in syntactic context. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '11, pages 1047–1057, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Richard Socher, Brody Huval, Christopher D. Manning, and Andrew Y. Ng. 2012. Semantic compositionality through recursive matrix-vector spaces. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, EMNLP-CoNLL '12, pages 1201–1211, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Veselin Stoyanov, Nathan Gilbert, Claire Cardie, and Ellen Riloff. 2009. Conundrums in noun phrase coreference resolution: making sense of the state-of-the-art. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2 - Volume 2*, ACL '09, pages 656–664, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Stefanie Tellex, Boris Katz, Jimmy J. Lin, Aaron Fernandes, and Gregory Marton. 2003. Quantitative evaluation of passage retrieval algorithms for question answering. In *SIGIR*, pages 41–47.
- Stefan Thater, Hagen Fürstenu, and Manfred Pinkal. 2010. Contextualizing semantic representations using syntactically enriched vector models. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, ACL '10, pages 948–957, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Stephen Tratz and Eduard Hovy. 2011. A fast, accurate, non-projective, semantically-enriched parser. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '11, pages 1257–1268, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Hila Weisman, Jonathan Berant, Idan Szpektor, and Ido Dagan. 2012. Learning verb inference rules from linguistically-motivated evidence. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, EMNLP-CoNLL '12, pages 194–204, Stroudsburg, PA, USA. Association for Computational Linguistics.
- S. K. M. Wong and Vijay V. Raghavan. 1984. Vector space model of information retrieval: a reevaluation. In *Proceedings of the 7th annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '84, pages 167–185, Swinton, UK. British Computer Society.