

Frame-Based Continuous Lexical Semantics through Exponential Family Tensor Factorization and Semantic Proto-Roles

Francis Ferraro and Adam Poliak and Ryan Cotterell and Benjamin Van Durme

Center for Language and Speech Processing

Johns Hopkins University

{ferraro, azpoliak, ryan.cotterell, vandurme}@cs.jhu.edu

Abstract

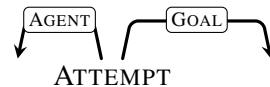
We study how different frame annotations complement one another when learning continuous lexical semantics. We learn the representations from a tensorized skip-gram model that consistently encodes syntactic-semantic content better, with multiple 10% gains over baselines.

1 Introduction

Consider “Bill” in Fig. 1: what is his involvement with the words “would try,” and what does this involvement *mean*? Word embeddings represent such meaning as points in a real-valued vector space (Deerwester et al., 1990; Mikolov et al., 2013). These representations are often learned by exploiting the frequency that the word cooccurs with contexts, often within a user-defined window (Harris, 1954; Turney and Pantel, 2010). When built from large-scale sources, like Wikipedia or web crawls, embeddings capture general characteristics of words and allow for robust downstream applications (Kim, 2014; Das et al., 2015).

Frame semantics generalize word meanings to that of analyzing structured and interconnected labeled “concepts” and abstractions (Minsky, 1974; Fillmore, 1976, 1982). These concepts, or roles, *implicitly* encode expected properties of that word. In a frame semantic analysis of Fig. 1, the segment “would try” *triggers* the ATTEMPT frame, filling the expected roles AGENT and GOAL with “Bill” and “the same tactic,” respectively. While frame semantics provide a structured form for analyzing words with crisp, categorically-labeled concepts, the encoded properties and expectations are implicit. What does it *mean* to fill a frame’s role?

Semantic proto-role (SPR) theory, motivated by Dowty (1991)’s thematic proto-role theory, offers an answer to this. SPR replaces categorical roles



She said Bill would try the same tactic again.

Figure 1: A simple frame analysis.

with judgements about multiple underlying properties about what is likely true of the entity filling the role. For example, SPR talks about how likely it is for Bill to be a willing participant in the ATTEMPT. The answer to this and other simple judgments characterize Bill and his involvement. Since SPR both captures the likelihood of certain properties and characterizes roles as groupings of properties, we can view SPR as representing a type of continuous frame semantics.

We are interested in capturing these SPR-based properties and expectations within word embeddings. We present a method that learns frame-enriched embeddings from millions of documents that have been semantically parsed with multiple different frame analyzers (Ferraro et al., 2014). Our method leverages Cotterell et al. (2017)’s formulation of Mikolov et al. (2013)’s popular skip-gram model as exponential family principal component analysis (EPCA) and tensor factorization. This paper’s primary contributions are: (i) enriching learned word embeddings with multiple, automatically obtained frames from large, disparate corpora; and (ii) demonstrating these enriched embeddings better capture SPR-based properties. In so doing, we also generalize Cotterell et al.’s method to arbitrary tensor dimensions. This allows us to include an arbitrary amount of semantic information when learning embeddings. Our variable-size tensor factorization code is available at <https://github.com/fmof/tensor-factorization>.

2 Frame Semantics and Proto-Roles

Frame semantics currently used in NLP have a rich history in linguistic literature. Fillmore (1976)’s frames are based on a word’s context and prototypical concepts that an individual word evokes; they intend to represent the meaning of lexical items by mapping words to real world concepts and shared experiences. Frame-based semantics have inspired many semantic annotation schemata and datasets, such as FrameNet (Baker et al., 1998), PropBank (Palmer et al., 2005), and Verbnet (Schuler, 2005), as well as composite resources (Hovy et al., 2006; Palmer, 2009; Banarescu et al., 2012).¹

Thematic Roles and Proto Roles These resources map words to their meanings through discrete/categorically labeled frames and roles; sometimes, as in FrameNet, the roles can be very descriptive (e.g., the DEGREE role for the AFFIRM_OR_DENY frame), while in other cases, as in PropBank, the roles can be quite general (e.g., ARG0). Regardless of the actual schema, the roles are based on thematic roles, which map a predicate’s arguments to a semantic representation that makes various semantic distinctions among the arguments (Dowty, 1989).² Dowty (1991) claims that thematic role distinctions are not atomic, i.e., they can be deconstructed and analyzed at a lower level. Instead of many discrete thematic roles, Dowty (1991) argues for *proto-thematic roles*, e.g. PROTO-AGENT rather than AGENT, where distinctions in proto-roles are based on clusterings of logical entailments. That is, PROTO-AGENTS often have certain properties in common, e.g., manipulating other objects or willingly participating in an action; PROTO-PATIENTS are often changed or affected by some action. By decomposing the meaning of roles into properties or expectations that can be reasoned about, proto-roles can be seen as including a form of vector representation within structured frame semantics.

3 Continuous Lexical Semantics

Word embeddings represent word meanings as elements of a (real-valued) vector space (Deerwester et al., 1990). Mikolov et al. (2013)’s word2vec methods—skip-gram (SG) and continuous bag of

¹See Petrucci and de Melo (2014) for detailed descriptions on frame semantics’ contributions to applied NLP tasks.

²Thematic role theory is rich, and beyond this paper’s scope (Whitehead, 1920; Davidson, 1967; Cresswell, 1973; Kamp, 1979; Carlson, 1984).

words (CBOW)—repopularized these methods. We focus on SG, which predicts the context i around a word j , with learned representations \mathbf{c}_i and \mathbf{w}_j , respectively, as $p(\text{context } i \mid \text{word } j) \propto \exp(\mathbf{c}_i^\top \mathbf{w}_j) = \exp(\mathbf{1}^\top(\mathbf{c}_i \odot \mathbf{w}_j))$, where \odot is the Hadamard (pointwise) product. Traditionally, the context words i are those words within a small window of j and are trained with negative sampling (Goldberg and Levy, 2014).

3.1 Skip-Gram as Matrix Factorization

Levy and Goldberg (2014b), and subsequently Keerthi et al. (2015), showed how vectors learned under SG with the negative sampling are, under certain conditions, the factorization of (shifted) positive pointwise mutual information. Cotterell et al. (2017) showed that SG is a form of exponential family PCA that factorizes the matrix of word/context cooccurrence counts (rather than shifted positive PMI values). With this interpretation, they generalize SG from matrix to tensor factorization, and provide a theoretical basis for modeling higher-order SG (or additional context, such as morphological features of words) within a word embeddings framework.

Specifically, Cotterell et al. recast higher-order SG as maximizing the log-likelihood

$$\sum_{ijk} \mathcal{X}_{ijk} \log p(\text{context } i \mid \text{word } j, \text{feature } k) \quad (1)$$

$$= \sum_{ijk} \mathcal{X}_{ijk} \log \frac{\exp(\mathbf{1}^\top(\mathbf{c}_i \odot \mathbf{w}_j \odot \mathbf{a}_k))}{\sum_{i'} \exp(\mathbf{1}^\top(\mathbf{c}_{i'} \odot \mathbf{w}_j \odot \mathbf{a}_k))}, \quad (2)$$

where \mathcal{X}_{ijk} is a cooccurrence count 3-tensor of words j , surrounding contexts i , and features k .

3.2 Skip-Gram as n-Tensor Factorization

When factorizing an n -dimensional tensor to include an arbitrary number of L annotations, we replace *feature* k in Equation (1) and \mathbf{a}_k in Equation (2) with each annotation type l and vector α_l included. $\mathcal{X}_{i,j,k}$ becomes $\mathcal{X}_{i,j,l_1,\dots,l_L}$, representing the number of times word j appeared in context i with features l_1 through l_L . We maximize

$$\sum_{i,j,l_1,\dots,l_L} \mathcal{X}_{i,j,l_1,\dots,l_L} \log \beta_{i,j,l_1,\dots,l_L}$$

$$\beta_{i,j,l_1,\dots,l_L} \propto \exp(\mathbf{1}^\top(\mathbf{c}_i \odot \mathbf{w}_j \odot \alpha_{l_1} \odot \dots \odot \alpha_{l_L})).$$

4 Experiments

Our end goal is to use multiple kinds of automatically obtained, “in-the-wild” frame se-

semantic parses in order to improve the semantic content—specifically SPR-type information—within learned lexical embeddings. We utilize majority portions of the Concretely Annotated New York Times and Wikipedia corpora from Ferraro et al. (2014). These have been annotated with three frame semantic parses: FrameNet from Das et al. (2010), and both FrameNet and PropBank from Wolfe et al. (2016). In total, we use nearly five million frame-annotated documents.

Extracting Counts The baseline extraction we consider is a standard sliding window: for each word w_j seen $\geq T$ times, extract all words w_i two to the left and right of w_j . These counts, forming a matrix, are then used within standard `word2vec`. We also follow Cotterell et al. (2017) and augment the above with the signed number of tokens separating w_i and w_j , e.g., recording that w_i appeared two to the left of w_j ; these counts form a 3-tensor.

To turn semantic parses into tensor counts, we first identify relevant information from the parses. We consider all parses that are triggered by the target word w_j (seen $\geq T$ times) and that have at least one role filled by some word in the sentence. We organize the extraction around roles and what fills them. We extract every word w_r that fills all possible triggered frames; each of those frame and role labels; and the distance between filler w_r and trigger w_j . This process yields a 9-tensor \mathcal{X} .³ Although we **always** treat the trigger as the “original” word (e.g., word j , with vector \mathbf{w}_j), later we consider (1) what to include from \mathcal{X} , (2) what to predict (what to treat as the “context” word i), and (3) what to treat as auxiliary features.

Data Discussion The baseline extraction methods result in roughly symmetric target and surrounding word counts. This is not the case for the frame extraction. Our target words must trigger some semantic parse, so our target words are actually target triggers. However, the surrounding context words are those words that fill semantic roles. As shown in Table 1, there are an order-of-magnitude fewer triggers than target words, but up to an order-of-magnitude *more* surrounding words.

Implementation We generalize Levy and Goldberg (2014a)’s and Cotterell et al. (2017)’s code

³ Each record consists of the trigger, a role filler, the number of words between the trigger and filler, and the relevant frame and roles from the three semantic parsers. Being automatically obtained, the parses are overlapping and incomplete; to properly form \mathcal{X} , one can implicitly include special `<NO.FRAME>` and `<NO.ROLE>` labels as needed.

	windowed	frame
# target words	232	35.9 (triggers)
	<i>404</i>	<i>45.7</i> (triggers)
# surrounding words	232	531 (role fillers)
	<i>404</i>	<i>2,305</i> (role fillers)

Table 1: Vocabulary sizes, in thousands, extracted from Ferraro et al. (2014)’s data with both the standard sliding context window approach (§3) and the frame-based approach (§4). Upper numbers (Roman) are for newswire; lower numbers (italics) are Wikipedia. For both corpora, 800 total FrameNet frame types and 5100 PropBank frame types are extracted.

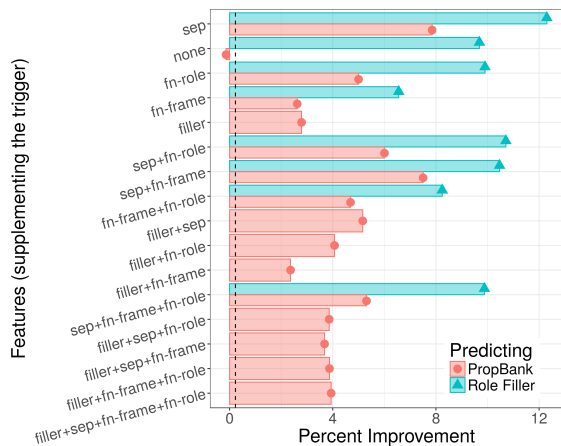
to enable any arbitrary dimensional tensor factorization, as described in §3.2. We learn 100-dimensional embeddings for words that appear at least 100 times from 15 negative samples.⁴ The implementation is available at <https://github.com/fmof/tensor-factorization>.

Metric We evaluate our learned (trigger) embeddings \mathbf{w} via QVEC (Tsvetkov et al., 2015). QVEC uses canonical correlation analysis to measure the Pearson correlation between \mathbf{w} and a collection of *oracle* lexical vectors \mathbf{o} . These oracle vectors are derived from a human-annotated resource. For QVEC, higher is better: a higher score indicates \mathbf{w} more closely correlates (positively) with \mathbf{o} .

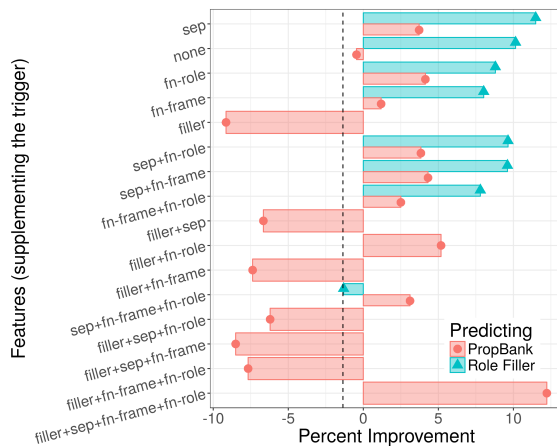
Evaluating Semantic Content with SPR Motivated by Dowty (1991)’s proto-role theory, Reisinger et al. (2015), with a subsequent expansion by White et al. (2016), annotated thousands of predicate-argument pairs (v, a) with (boolean) applicability and (ordinal) likelihoods of well-motivated semantic properties applying to/being true of a .⁵ These likelihood judgments, under the SPR framework, are converted from a five-point Likert scale to a 1–5 interval scale. Because the predicate-argument pairs were extracted from previously annotated dependency trees, we link each property with the dependency relation joining v and a when forming the oracle vectors; each component of an oracle vector \mathbf{o}_v is the unity-normalized sum of likelihood judgments for joint property and grammatical relation, using the interval responses when the property is applicable and discarding non-applicable properties, i.e. treating the response as 0. Thus, the combined 20 properties of Reisinger et al. (2015) and White et al. (2016)—together with the four basic grammatical

⁴In preliminary experiments, this occurrence threshold did not change the overall conclusions.

⁵ We use the training portion of <http://decomp.net/wp-content/uploads/2015/08/UniversalDecompositionalSemantics.tar.gz>.



(a) Changes in SPR-QVEC for *Annotated NYT*.



(b) Changes in SPR-QVEC for Wikipedia.

Figure 2: Effect of frame-extracted tensor counts on our SPR-QVEC evaluation. Deltas are shown as relative percent changes vs. the `word2vec` baseline. The dashed line represents the 3-tensor `word2vec` method of Cotterell et al. (2017). Each row represents an ablation model: `sep` means the prediction relies on the token separation distance between the frame and role filler, `fn-frame` means the prediction uses FrameNet frames, `fn-role` means the prediction uses FrameNet roles, and `filler` means the prediction uses the tokens filling the frame role. Read from top to bottom, additional contextual features are denoted with a +. Note when `filler` is used, we only predict PropBank roles.

relations *nsubj*, *dobj*, *iobj* and *nsubjpass*—result in 80-dimensional oracle vectors.⁶

Predict Fillers or Roles? Since SPR judgments are between predicates and arguments, we predict the words filling the roles, and treat all other frame information as auxiliary features. SPR annotations were originally based off of (gold-standard) PropBank annotations, so we also train a model to predict PropBank frames and roles, thereby treating role-filling text and all other frame information as auxiliary features. In early experiments, we found it beneficial to treat the FrameNet annotations additively and not distinguish one system’s output from another. Treating the annotations additively serves as a type of collapsing operation. Although \mathcal{X} started as a 9-tensor, we only consider up to 6-tensors: trigger, role filler, token separation between the trigger and filler, PropBank frame and role, FrameNet frame, and FrameNet role.

Results Fig. 2 shows the overall percent change for SPR-QVEC from the filler and role prediction models, on newswire (Fig. 2a) and Wikipedia (Fig. 2b), across different ablation models. We indicate additional contextual features being used with a +: `sep` uses the token separation distance between the frame and role filler, `fn-frame` uses FrameNet frames, `fn-role` uses FrameNet roles, `filler` uses the tokens filling the frame

role, and `none` indicates no additional information is used when predicting. The 0 line represents a plain `word2vec` baseline and the dashed line represents the 3-tensor baseline of Cotterell et al. (2017). Both of these baselines are windowed: they are restricted to a local context and cannot take advantage of frames or any lexical signal that can be derived from frames.

Overall, we notice that we obtain large improvements from models trained on lexical signals that have been *derived* from frame output (`sep` and `none`), even if the model *itself* does not incorporate any frame labels. The embeddings that predict the role filling lexical items (the green triangles) correlate higher with SPR oracles than the embeddings that predict PropBank frames and roles (red circles). Examining Fig. 2a, we see that both model types outperform both the `word2vec` and Cotterell et al. (2017) baselines in nearly all model configurations and ablations. We see the highest improvement when predicting role fillers given the frame trigger and the number of tokens separating the two (the green triangles in the `sep` rows).

Comparing Fig. 2a to Fig. 2b, we see newswire is more amenable to predicting PropBank frames and roles. We posit this is a type of out-of-domain error, as the PropBank parser was trained on newswire. We also find that newswire is overall more amenable to incorporating limited frame-based features, particularly when predicting PropBank using lexical role fillers as part of the con-

⁶ The full cooccurrence among the properties and relations is relatively sparse. Nearly two thirds of all non-zero oracle components are comprised of just fourteen properties, and only the *nsubj* and *dobj* relations.

anticipated		anticipated	
Filler sep		PropBank sep	
1 foresaw	6 pondered	1 anticipate	6 intimidated
2 figuring	7 kidded	2 anticipating	7 separating
3 alleviated	8 constituted	3 anticipates	8 separates
4 craved	9 uttering	4 stabbing	9 drag
5 jeopardized	10 forgiven	5 separate	10 guarantee

invented		invented	
Filler sep		PropBank sep	
1 pioneered	6 tolerated	1 invent	6 aspire
2 scratch	7 resurrected	2 document	7 documenting
3 complemented	8 sweated	3 documented	8 aspires
4 competed	9 fancies	4 invents	9 inventing
5 consoled	10 concocted	5 documents	10 swinging

producing		producing	
Filler sep		PropBank sep	
1 containing	6 storing	1 produces	6 ridden
2 contains	7 reproduce	2 produce	7 improves
3 manufactures	8 store	3 produced	8 surround
4 contain	9 exhibiting	4 prized	9 surrounds
5 consume	10 furnish	5 originates	10 originating

Figure 3: K -Nearest Neighbors for three randomly sampled trigger words, from two newswire models.

textual features. We hypothesize this is due to the significantly increased vocabulary size of the Wikipedia role fillers (c.f., Tab. 1). Note, however, that by using all available schema information when predicting PropBank, we are able to compensate for the increased vocabulary.

In Fig. 3 we display the ten nearest neighbors for three randomly sampled trigger words according to two of the highest performing newswire models. They each condition on the trigger and the role filler/trigger separation; these correspond to the `sep` rows of Fig. 2a. The left column of Fig. 3 predicts the role filler, while the right column predicts PropBank annotations. We see that while both models learn inflectional relations, this quality is prominent in the model that predicts PropBank information while the model predicting role fillers learns more non-inflectional paraphrases.

5 Related Work

The recent popularity of word embeddings have inspired others to consider leveraging linguistic annotations and resources to learn embeddings. Both Cotterell et al. (2017) and Levy and Goldberg (2014a) incorporate additional syntactic and morphological information in their word embeddings. Rothe and Schütze (2015)’s use lexical resource entries, such as WordNet synsets, to improve pre-computed word embeddings. Through generalized CCA, Rastogi et al. (2015) incorporate paraphrased FrameNet training data. On the applied side, Wang and Yang (2015) used frame embeddings—produced by training `word2vec` on tweet-derived semantic frame (names)—as additional features in downstream prediction.

Teichert et al. (2017) similarly explored the relationship between semantic frames and thematic proto-roles. They proposed using a Conditional Random Field (Lafferty et al., 2001) to jointly and conditionally model SPR and SRL. Teichert et al. (2017) demonstrated slight improvements in jointly and conditionally predicting PropBank (Bonial et al., 2013)’s semantic role labels and Reisinger et al. (2015)’s proto-role labels.

6 Conclusion

We presented a way to learn embeddings enriched with multiple, automatically obtained frames from large, disparate corpora. We also presented a QVEC evaluation for semantic proto-roles. As demonstrated by our experiments, our extension of Cotterell et al. (2017)’s tensor factorization enriches word embeddings by including syntactic-semantic information not often captured, resulting in consistently higher SPR-based correlations. The implementation is available at <https://github.com/fmof/tensor-factorization>.

Acknowledgments

This work was supported by Johns Hopkins University, the Human Language Technology Center of Excellence (HLTCOE), DARPA DEFT, and DARPA LORELEI. We would also like to thank three anonymous reviewers for their feedback. The views and conclusions contained in this publication are those of the authors and should not be interpreted as representing official policies or endorsements of DARPA or the U.S. Government.

References

- Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. [The berkeley framenet project](#). In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics - Volume 1*. Association for Computational Linguistics, Stroudsburg, PA, USA, ACL ’98, pages 86–90. <https://doi.org/10.3115/980845.980860>.
- Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2012. Abstract meaning representation (amr) 1.0 specification. In *Parsing on Freebase from Question-Answer Pairs*. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. Seattle: ACL, pages 1533–1544.
- Claire Bonial, Kevin Stowe, and Martha Palmer. 2013. [Renewing and revising semlink](#). In *Proceedings of*

- the 2nd Workshop on Linked Data in Linguistics (LDL-2013): Representing and linking lexicons, terminologies and other language data. Association for Computational Linguistics, Pisa, Italy, pages 9–17. <http://www.aclweb.org/anthology/W13-5503>.
- Greg N Carlson. 1984. Thematic roles and their role in semantic interpretation. *Linguistics* 22(3):259–280.
- Ryan Cotterell, Adam Poliak, Benjamin Van Durme, and Jason Eisner. 2017. Explaining and generalizing skip-gram through exponential family principal component analysis. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*. Valencia, Spain.
- Maxwell John Cresswell. 1973. *Logics and languages*. London: Methuen [Distributed in the U.S.A. By Harper & Row].
- Dipanjan Das, Nathan Schneider, Desai Chen, and Noah A Smith. 2010. Probabilistic frame-semantic parsing. In *Human language technologies: The 2010 annual conference of the North American chapter of the association for computational linguistics*. Association for Computational Linguistics, pages 948–956.
- Rajarshi Das, Manzil Zaheer, and Chris Dyer. 2015. Gaussian lda for topic models with word embeddings. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Association for Computational Linguistics, Beijing, China, pages 795–804. <http://www.aclweb.org/anthology/P15-1077>.
- Donald Davidson. 1967. The logical form of action sentences. In Nicholas Rescher, editor, *The Logic of Decision and Action*, University of Pittsburgh Press.
- Scott Deerwester, Susan T. Dumais, George W. Furnas, Thomas K. Landauer, and Richard Harshman. 1990. Indexing by latent semantic analysis. *JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE* 41(6):391–407.
- David Dowty. 1991. Thematic proto-roles and argument selection. *Language* 67(3):547–619.
- David R Dowty. 1989. On the semantic content of the notion of thematic role. In *Properties, types and meaning*, Springer, pages 69–129.
- Francis Ferraro, Max Thomas, Matthew R. Gormley, Travis Wolfe, Craig Harman, and Benjamin Van Durme. 2014. Concretely Annotated Corpora. In *4th Workshop on Automated Knowledge Base Construction (AKBC)*.
- Charles Fillmore. 1982. Frame semantics. *Linguistics in the morning calm* pages 111–137.
- Charles J Fillmore. 1976. Frame semantics and the nature of language*. *Annals of the New York Academy of Sciences* 280(1):20–32.
- Yoav Goldberg and Omer Levy. 2014. word2vec explained: Deriving Mikolov et al.’s negative-sampling word-embedding method. *arXiv preprint arXiv:1402.3722*.
- Zellig S Harris. 1954. Distributional structure. *Word* 10(2-3):146–162.
- Eduard Hovy, Mitchell Marcus, Martha Palmer, Lance Ramshaw, and Ralph Weischedel. 2006. Ontonotes: the 90% solution. In *Proceedings of the human language technology conference of the NAACL, Companion Volume: Short Papers*. Association for Computational Linguistics, pages 57–60.
- Hans Kamp. 1979. Events, instants and temporal reference. In *Semantics from different points of view*, Springer, pages 376–418.
- S. Sathiya Keerthi, Tobias Schnabel, and Rajiv Khanna. 2015. Towards a better understanding of predict and count models. *arXiv preprint arXiv:1511.0204*.
- Yoon Kim. 2014. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, Doha, Qatar, pages 1746–1751. <http://www.aclweb.org/anthology/D14-1181>.
- John D. Lafferty, Andrew McCallum, and Fernando C. N. Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the Eighteenth International Conference on Machine Learning*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, ICML ’01, pages 282–289. <http://dl.acm.org/citation.cfm?id=645530.655813>.
- Omer Levy and Yoav Goldberg. 2014a. Dependency-based word embeddings. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Association for Computational Linguistics, Baltimore, Maryland, pages 302–308. <http://www.aclweb.org/anthology/P14-2050>.
- Omer Levy and Yoav Goldberg. 2014b. Neural word embedding as implicit matrix factorization. In *Advances in neural information processing systems*. pages 2177–2185.
- Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.
- Marvin Minsky. 1974. A framework for representing knowledge. MIT-AI Laboratory Memo 306.
- Martha Palmer. 2009. Semlink: Linking propbank, verbnet and framenet. In *Proceedings of the Generative Lexicon Conference*. GenLex-09, 2009 Pisa, Italy, pages 9–15.

- Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The proposition bank: An annotated corpus of semantic roles. *Computational linguistics* 31(1):71–106.
- Miriam R. L. Petrucci and Gerard de Melo, editors. 2014. *Proceedings of Frame Semantics in NLP: A Workshop in Honor of Chuck Fillmore (1929-2014)*. Association for Computational Linguistics, Baltimore, MD, USA. <http://www.aclweb.org/anthology/W14-30>.
- Pushpendre Rastogi, Benjamin Van Durme, and Raman Arora. 2015. **Multiview LSA: Representation Learning via Generalized CCA**. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, Denver, Colorado, pages 556–566. <http://www.aclweb.org/anthology/N15-1058>.
- Drew Reisinger, Rachel Rudinger, Francis Ferraro, Craig Harman, Kyle Rawlins, and Benjamin Van Durme. 2015. Semantic proto-roles. *Transactions of the Association for Computational Linguistics (TACL)* 3:475–488.
- Sascha Rothe and Hinrich Schütze. 2015. **Autoextend: Extending word embeddings to embeddings for synsets and lexemes**. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Association for Computational Linguistics, Beijing, China, pages 1793–1803. <http://www.aclweb.org/anthology/P15-1173>.
- Karin Kipper Schuler. 2005. Verbnet: A broad-coverage, comprehensive verb lexicon .
- Adam Teichert, Adam Poliak, Benjamin Van Durme, and Matthew Gormley. 2017. Semantic proto-role labeling. In *AAAI Conference on Artificial Intelligence*.
- Yulia Tsvetkov, Manaal Faruqui, Wang Ling, Guillaume Lample, and Chris Dyer. 2015. **Evaluation of word vector representations by subspace alignment**. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Lisbon, Portugal, pages 2049–2054. <http://aclweb.org/anthology/D15-1243>.
- Peter D Turney and Patrick Pantel. 2010. From frequency to meaning: Vector space models of semantics. *Journal of artificial intelligence research* 37:141–188.
- William Yang Wang and Diyi Yang. 2015. **That’s so annoying!!!: A lexical and frame-semantic embedding based data augmentation approach to automatic categorization of annoying behaviors using #petpeeve tweets**. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Lisbon, Portugal, pages 2557–2563. <http://aclweb.org/anthology/D15-1306>.
- Aaron Steven White, Drew Reisinger, Keisuke Sakaguchi, Tim Vieira, Sheng Zhang, Rachel Rudinger, Kyle Rawlins, and Benjamin Van Durme. 2016. **Universal decompositional semantics on universal dependencies**. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Austin, Texas, pages 1713–1723. <https://aclweb.org/anthology/D16-1177>.
- Alfred North Whitehead. 1920. *The concept of nature: the Tarnier lectures delivered in Trinity College, November 1919*. Kessinger Publishing.
- Travis Wolfe, Mark Dredze, and Benjamin Van Durme. 2016. **A study of imitation learning methods for semantic role labeling**. In *Proceedings of the Workshop on Structured Prediction for NLP*. Association for Computational Linguistics, Austin, TX, pages 44–53. <http://aclweb.org/anthology/W16-5905>.