

# Composition of Word Representations Improves Semantic Role Labelling

Michael Roth and Kristian Woodsend

Institute for Language, Cognition and Computation

School of Informatics, University of Edinburgh

{mroth, kwoodsen}@inf.ed.ac.uk

## Abstract

State-of-the-art semantic role labelling systems require large annotated corpora to achieve full performance. Unfortunately, such corpora are expensive to produce and often do not generalize well across domains. Even in domain, errors are often made where syntactic information does not provide sufficient cues. In this paper, we mitigate both of these problems by employing distributional word representations gathered from unlabelled data. While straight-forward word representations of predicates and arguments improve performance, we show that further gains are achieved by composing representations that model the interaction between predicate and argument, and capture full argument spans.

## 1 Introduction

The goal of *semantic role labelling* (SRL) is to discover the relations that hold between a predicate and its arguments in a given input sentence (e.g., “who” did “what” to “whom”, “when”, “where”, and “how”). This semantic knowledge at the predicate-argument level is required by inference-based NLP tasks in order to identify meaning-preserving transformations, such as active/passive, verb alternations and nominalizations. Several manually-build semantic resources, including FrameNet (Ruppenhofer et al., 2010) and PropBank (Palmer et al., 2005), have been developed with the goal of documenting and providing examples of such transformations and how they preserve semantic role information. Given that labelled corpora are inevitably restricted in size and coverage, and that syntactic cues are not by themselves unambiguous or sufficient, the success of systems that automatically provide corresponding analyses has been limited in practice.

Recent work on SRL has explored approaches that can leverage unlabelled data, following a semi-supervised (Fürstenauf and Lapata, 2012; Titov and Klementiev, 2012) or unsupervised learning paradigm (Abend et al., 2009; Titov and Klementiev, 2011). Unlabelled data provides additional statistical strength and can lead to more consistent models. For instance, latent representations of words can be computed, based on distributional similarity or language modelling, which can be used as additional features during traditional supervised learning. Although we would expect that extra features would improve classifier performance, this seems in part counter-intuitive. Just because one word has a specific representation does not mean that it should be assigned a specific argument label. Instead, one would expect a more complex interplay between predicate, argument and the context they appear in.

In this paper, we investigate the impact of distributional word representations for SRL. Initially, we augment the feature space with word representations for a predicate and its argument head. Furthermore, we use a compositional approach to model a representation of the full argument, by composing a joint representation of all words in the argument span, and we also investigate the interaction between predicate and argument, using a compositional representation of the dependency path. We demonstrate the benefits of these compositional features using a state-of-the-art semantic role labeller, which we evaluate on the English part of the CoNLL-2009 data set.

## 2 Related Work

Research into using distributional information in SRL dates back to Gildea and Jurafsky (2002), who used distributions over verb-object co-occurrence clusters to improve coverage in argument classification. The distribution of a word over these soft clusters assignments was added as

features to their classifier. The SRL system by Croce et al. (2010) combines argument clustering based on co-occurrence frequencies with a language model. Collobert et al. (2011) used distributional word representations in a neural network model that can update representations during training. Zapirain et al. (2013) suggested distributional information as a basis for a selectional preference model that can be used as a single additional feature for classifying potential arguments. Most recently, Hermann et al. (2014) used distributional word representations within pre-defined syntactic contexts as input to a classifier which learns to distinguish different predicate senses.

A complementary line of research explores the representation of sequence information. Prominent examples are the works by Deschacht and Moens (2009) and Huang and Yates (2010) who learned and applied Hidden Markov Models to assign state variables to words and word spans, which serve as supplementary features for classification. One drawback of this approach is that state variables are discrete and the number of states (i.e., their granularity) has to be chosen in advance.

The popularity of distributional methods for word representation has been a motivation for developing representations of larger constructions such as phrases and sentences, and there have been several proposals for computing the meaning of word combinations in vector spaces. Mitchell and Lapata (2010) introduced a general framework where composition is formulated as a function  $f$  of two vectors  $u$  and  $v$ . Depending on how  $f$  is chosen, different composition models arise, the simplest being an additive model where  $f(u, v) = u + v$ . To capture relational functions, Baroni and Zamparelli (2010) expanded on this approach by representing verbs, adjectives and adverbs by matrices which can modify the properties of nouns (represented by vectors). Socher et al. (2012) combined word representations with syntactic structure information, through a recursive neural network that learns vector space representations for multi-word phrases and sentences. An empirical comparison of these composition methods was provided in (Blacoe and Lapata, 2012).

In this work, we use type-based continuous representations of words to compose representations of multiple word sequences and spans, which can then be incorporated directly as features into SRL systems.

Distributional Feature	Computation
<b>Argument</b> $a$	$\vec{a}$
<b>Predicate</b> $p$	$\vec{p}$
<b>Predicate-argument Interaction</b>	$\vec{a} + \vec{p}$
<b>Argument Span</b> $w_1 \dots w_n$	$\sum_i \vec{w}_i$
<b>Dependency Path</b> from $a$ to $p$	$\sum_{w \in \text{path}(a,p)} \vec{w}$

Table 1: Features based on distributional word representations and additive composition. Vector  $\vec{w}$  denotes the representation of word  $w$ .

### 3 Method

Following the set-up of the CoNLL shared task in 2009, we consider predicate-argument structures that consist of a verbal or nominal predicate  $p$  and PropBank-labelled arguments  $a_i \in \{a_1 \dots a_n\}$ , where each  $a_i$  corresponds to the head word of the phrase that constitutes the respective argument. Traditional semantic role labelling approaches compute a set of applicable features on each pair  $\langle p, a_i \rangle$ , such as the observed lemma type of a word and the grammatical relation to its head, that serve as indicators for a particular role label.

The disadvantage of this approach lies in the fact that indicator features such as word and lemma type are often sparse in training data and hence do not generalize well across domains. In contrast, features based on distributional representations (e.g., raw co-occurrence frequencies) can be computed for every word, given that it occurs in some unlabelled corpus. In addition to this obvious advantage for out-of-domain settings, distributional representations can provide a more robust input signal to the classifier, for instance by projecting a matrix of co-occurrence frequencies to a lower-dimensional space. We hence hypothesize that such features enable the model to become more robust out-of-domain, while providing higher precision in-domain.

Although simply including the components of a word representation as features to a classifier can lead to immediate improvements in SRL performance, this observation seems in part counter-intuitive. Just because one word has a specific representation does not mean that it should be assigned a specific argument label. In fact, one would expect a more complex interplay between the representation of an argument  $a_i$  and the context it appears in. To model aspects of this interplay, we define an extended set of features that

further includes representations for the combination of  $p$  and  $a_i$ , the set of words in the dependency path between  $p$  and  $a_i$ , and the set of words in the full span of  $a_i$ . We compute additive compositional representations of multiple words, using the simplest method of Mitchell and Lapata (2010) where the composed representation is the uniformly weighted sum of each single representation. Our full set of feature types based on distributional word representations is listed in Table 1.

## 4 Experimental Setup

We evaluate the impact of different types of features by performing experiments on a benchmark dataset for semantic role labelling. To assess the gains of distributional representations realistically, we incorporate the features described in Section 3 into a state-of-the-art SRL system. The following paragraphs summarize the details of our experimental setup.

**Semantic Role Labeller.** In all our experiments, we use the publicly available system by Björkelund et al. (2010).<sup>1</sup> This system combines the first-ranked SRL system and the first-ranked syntactic parser in the CoNLL 2009 shared task for English (Björkelund et al., 2009; Bohnet, 2010). To the best of our knowledge, this combination represents the current state-of-the-art for semantic role labelling following the Prop-Bank/NomBank paradigm (Palmer et al., 2005; Meyers et al., 2004). To re-train and evaluate models with different feature sets, we use the same training, development and test sets as provided in the CoNLL shared task (Hajič et al., 2009). Although the employed system features a full syntactic-semantic parsing pipeline, we only modify the feature sets of the two components directly related to the actual role labelling task, namely argument identification and argument classification.

**Word Representations.** As a baseline, we simply added as features the word representations of the predicate and argument head involved in a classification decision (first two lines in Table 1). We experimented with a range of publicly available sets of word representations, including embeddings from various neural language models

Development	dims	P	R	F <sub>1</sub>
None	–	86.1	81.0	83.5
Brown clusters <sup>2</sup>	320	86.2	81.3	83.7
Neural LM <sup>2</sup>	50	86.2	<b>81.4</b>	<b>83.7</b>
Neural LM+Global <sup>3</sup>	50	86.2	81.4	83.7
HLBL <sup>2</sup>	50	<b>86.3</b>	81.3	83.7
H-PCA <sup>4</sup>	50	86.2	81.3	83.7
Eigenwords <sup>5</sup>	50	86.2	81.3	83.6

Table 2: Results on the CoNLL-2009 development set, using off-the-shelf word representations for predicates and argument as additional features. Performance numbers in percent.

(Mnih and Hinton, 2009; Collobert et al., 2011; Huang et al., 2012), eigenvectors (Dhillon et al., 2011), Brown clusters (Brown et al., 1992), and post-processed co-occurrence counts (Lebret and Collobert, 2014). Results on the development set for various off-the-shelf representations are shown in Table 2. The numbers reveal that any kind of word representation can be employed to improve results. We choose to perform all follow-up experiments using the 50-dimensional embeddings induced by Turian et al. (2010), using the method by Collobert et al., as they led to slightly better results in F<sub>1</sub>-score than other representations. No significant differences were observed, however, using other types of representations or vector sizes.

## 5 Results

We evaluate our proposed set of additional features on the CoNLL-2009 in-domain and out-of-domain test sets, using the aforementioned SRL system and word representations. All results are computed using the system’s built-in preprocessing pipeline and re-trained models for argument identification and classification. We report labelled precision, recall and semantic F1-score as computed by the official scorer.

The upper part of Table 3 shows SRL performance on the in-domain CoNLL-2009 test set, with and without (**Original**) additional features based on distributional representations. The results reveal that any type of additional feature helps to improve precision and recall in this setting (from 85.2% F<sub>1</sub>-score up to 85.5%), with significant gains for 4 of the 5 additional features (computed using a randomization test; cf. Yeh, 2000). Interestingly, we find that the features do not seem

<sup>1</sup><http://code.google.com/p/mate-tools/>

<sup>2</sup><http://metaoptimize.com/projects/wordreprs/>

<sup>3</sup><http://ai.stanford.edu/%7eeehuang/>

<sup>4</sup><http://lebret.ch/words/>

<sup>5</sup><http://www.cis.upenn.edu/%7eungar/eigenwords/>

In-domain	P	R	F <sub>1</sub>
Original	87.4	83.1	85.2
Original + Argument	87.6	83.3	85.4**
Original + Predicate	87.4	83.2	85.2
Original + Interaction	87.5	83.3	85.3**
Original + Span	<b>87.6</b>	<b>83.5</b>	<b>85.5**</b>
Original + Path	87.5	83.4	85.4**
Original + All	87.6	83.4	85.5**

  

Out-of-domain	P	R	F <sub>1</sub>
Original	76.9	71.7	74.2
Original + Argument	77.4	71.9	74.5
Original + Predicate	77.3	72.2	74.7*
Original + Interaction	77.2	72.0	74.5
Original + Span	77.3	72.3	74.7*
Original + Path	77.2	72.3	74.7*
Original + All	<b>77.5</b>	<b>73.0</b>	<b>75.2**</b>

Table 3: Results on both CoNLL-2009 test sets. All numbers in percent. Significant differences from Original in terms of F<sub>1</sub>-score are marked by asterisks (\* p<0.05, \*\* p<0.01).

to have a cumulative effect here, as indicated by the results with all features (+**All**, 85.5% F<sub>1</sub>). We conjecture that this is due to the high volume of existing in-domain training data, which renders our full feature set redundant. To test this conjecture, we further assess performance on the out-of-domain test set of the CoNLL-2009 shared task.

The results for the out-of-domain experiment are summarized in the lower part of Table 3. We again observe that each single feature type improves classification, with absolute gains being slightly higher than in the in-domain setting. More interestingly though, we find that the complete feature set boosts performance even further, achieving an overall gain in precision and recall of 0.6 and 1.3 percentage points, respectively. The resulting F<sub>1</sub>-score of 75.2 lies even higher than the top score for this particular data set reported in the CoNLL shared task (Zhao et al., 2009; 74.6 F<sub>1</sub>).

We next investigate the benefits of compositional representations over features for single words by assessing their impact on the overall result in an ablation study. Table 4 shows results of ablation tests performed for the three compositional feature types **Interaction**, **Span** and **Path** on the out-of-domain test set. The results reveal

Out-of-domain	P	R	F <sub>1</sub>
Original	76.9	71.7	74.2
Full (Original+All)	<b>77.5</b>	<b>73.0</b>	<b>75.2</b>
Full –Interaction	77.2	72.5	74.8
Full –Span	77.2	72.3	74.7
Full –Path	77.6	72.3	74.8

Table 4: Results of an ablation study over features based on compositional representations. All numbers in percent.

a considerable loss in recall, indicating the importance of including compositional word representations and confirming our intuition that they can provide additional gains over simple type-level representations. In the next section, we discuss this result in more detail and provide examples of improved classification decisions.

## 6 Discussion

As a more detailed qualitative analysis, we examined the impact of word representations on SRL performance with respect to different argument labels and predicate types. Results on the in-domain data set, shown in the upper part of Table 5, suggest that most improvements in terms of precision are gained for verbal predicates, while nominal predicates primarily benefit from higher recall. One reason for the latter observation might be that arguments of nominal predicates are generally much harder to identify for the **Original** model, as the cues provided by indicator features on words and syntax are often inconclusive. For verbal predicates, the word representations mainly provide reinforcing signals to the classifier, improving its precision at a slight cost of recall.

The results on the out-of-domain data set provide more insights regarding the suitability of word representations for generalization. As shown in the lower half of Table 5, the additional features on average have a positive impact on precision and recall. For verbal predicates, we observe only one case, namely A0, in which improvements in recall came with a decline in precision. Regarding nominal predicates, the trend is similar to what we have seen in the in-domain setting, with most gains being achieved in terms of recall.

Apart from assessing quantitative effects, we further examined cases that directly show the qualitative gains of the compositional features defined

Sentence with <i>predicate</i> and [gold argument <sub>label</sub> ]	Original	Features required for correction
(1) He did not resent [their <sub>A0</sub> ] <i>supervision</i>	A1	<b>Interaction</b>
(2) [He <sub>A1</sub> ] is getting plenty of <i>rest</i>	no label	<b>Interaction, Path</b>
(3) [He <sub>A0</sub> ] rose late and went down to <i>have</i> breakfast.	no label	<b>Path</b>
(4) He was able to <i>sit</i> [for hours <sub>AM-TMP</sub> ].	A2	<b>Span</b>
(5) Because he had <i>spoken</i> [too softly <sub>AM-MNR</sub> ].	AM-TMP	<b>Span</b>

Table 6: Example sentences in which distributional features compensated for errors made by **Original**.

In-domain	verbal		nominal	
Label	P	R	P	R
A0	+0.4	+0.4	-0.1	+2.4
A1	+0.2	-0.4	+0.6	+1.5
A2	+1.7	-1.5	-	+2.5
AM-ADV	+0.8	+0.2	-9.9	-3.1
AM-DIS	+0.3	-3.2	-	-
AM-LOC	+0.8	+1.1	+0.6	+3.0
AM-MNR	-0.5	-1.2	+2.7	+0.3
AM-TMP	-1.2	-0.7	-1.9	+3.3

  

Out-of-domain	verbal		nominal	
Label	P	R	P	R
A0	-0.9	+2.5	-2.5	-0.4
A1	+1.7	+0.8	+1.0	+3.7
A2	+1.4	+0.7	-2.5	+3.2
AM-ADV	+5.6	+0.7	-	-
AM-DIS	+7.3	-	-	-
AM-LOC	+0.7	+2.4	-	+15.0
AM-MNR	+6.4	+10.5	+9.7	+10.7
AM-TMP	+1.6	+1.8	-6.7	+1.1

Table 5: Differences in precision and recall per argument label and predicate word category. All numbers represent absolute percentage points.

in Section 3. Table 6 lists examples from the out-of-domain data set that were misclassified by the **Original** model but could be correctly predicted using our enhanced feature set. As illustrated by Examples (1) and (2), the **Interaction** feature seems to help recall by guiding classification decisions towards more meaningful and complete structures.

Improvements using the **Path** feature can be observed in cases where nested syntactic structures need to be processed, as required in Example (2). In another instance, Example (3), the following path is predicted between argument and predicate: *He*  $\xrightarrow{\text{SBJ}}$  *rose*  $\xleftarrow{\text{COORD}}$  *and*  $\xleftarrow{\text{CONJ}}$  *went*  $\xleftarrow{\text{OPRD}}$  *to*  $\xleftarrow{\text{IM}}$  *have*.

Such cases are particularly problematic for the **Original** model because long and potentially erroneous paths are sparse in the training data.

Further gains in performance are achieved using the **Span** feature, which enables the model to better handle infrequent and out-of-vocabulary words occurring in an argument span, including “hours” and “softly” in Example (4) and (5), respectively.

## 7 Conclusions

In this paper, we proposed to enhance the feature space of a state-of-the-art semantic role labeller by applying and composing distributional word representations. Our results indicate that combining such features with standard syntactic cues leads to more precise and more robust models, with significant improvements both in-domain and out-of-domain. Ablation tests on an out-of-domain data set have shown that gains in recall are mostly due to features based on composed representations. Given the novelty of these features for SRL, we believe that this insight is remarkable and deserves further investigation. In future work, we plan to apply more sophisticated models of compositionality to better represent predicate-argument structures and to guide classification decisions towards outcomes that are semantically more plausible. We anticipate that this line of research will also be of interest for a range of related tasks beyond traditional SRL, including predicate-argument structure alignment (Roth and Frank, 2012) and implicit argument linking (Gerber and Chai, 2012).

## Acknowledgements

This work has been supported by the FP7 Collaborative Project S-CASE (Grant Agreement No 610717) funded by the European Commission (Michael Roth), and by EPSRC (EP/K017845/1) in the framework of the CHIST-ERA READERS project (Kristian Woodsend).

## References

- Omri Abend, Roi Reichart, and Ari Rappoport. 2009. Unsupervised argument identification for semantic role labeling. 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*, pages 28–36, Suntec, Singapore, August.
- 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*, pages 1183–1193, Cambridge, MA, October. Association for Computational Linguistics.
- Anders Björkelund, Love Hafdell, and Pierre Nugues. 2009. Multilingual semantic role labeling. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning: Shared Task*, pages 43–48. Association for Computational Linguistics.
- Anders Björkelund, Bernd Bohnet, Love Hafdell, and Pierre Nugues. 2010. A high-performance syntactic and semantic dependency parser. In *Coling 2010: Demonstration Volume*, pages 33–36, Beijing, China, August. Coling 2010 Organizing Committee.
- William Blacoe and Mirella Lapata. 2012. A comparison of vector-based representations for semantic composition. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 546–556, Jeju Island, Korea, July. Association for Computational Linguistics.
- Bernd Bohnet. 2010. Top accuracy and fast dependency parsing is not a contradiction. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pages 89–97, Beijing, China, August.
- Peter F Brown, Peter V Desouza, Robert L Mercer, Vincent J Della Pietra, and Jenifer C Lai. 1992. Class-based n-gram models of natural language. *Computational linguistics*, 18(4):467–479.
- Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. 2011. Natural language processing (almost) from scratch. *The Journal of Machine Learning Research*, 12:2493–2537.
- Danilo Croce, Cristina Giannone, Paolo Annesi, and Roberto Basili. 2010. Towards open-domain semantic role labeling. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 237–246, Uppsala, Sweden, July. Association for Computational Linguistics.
- Koen Deschacht and Marie-Francine Moens. 2009. Semi-supervised semantic role labeling using the Latent Words Language Model. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 21–29, Singapore, August.
- Paramveer Dhillon, Dean P Foster, and Lyle H Ungar. 2011. Multi-view learning of word embeddings via CCA. In *Advances in Neural Information Processing Systems*, pages 199–207.
- Hagen Fürstenaу and Mirella Lapata. 2012. Semi-supervised semantic role labeling via structural alignment. *Computational Linguistics*, 38(1):135–171.
- Matthew Gerber and Joyce Chai. 2012. Semantic Role Labeling of Implicit Arguments for Nominal Predicates. *Computational Linguistics*, 38(4):755–798.
- Daniel Gildea and Daniel Jurafsky. 2002. Automatic labeling of semantic roles. *Computational linguistics*, 28(3):245–288.
- Jan Hajič, Massimiliano Ciaramita, Richard Johansson, Daisuke Kawahara, Maria Antònia Martí, Lluís Màrquez, Adam Meyers, Joakim Nivre, Sebastian Padó, Jan Štěpánek, et al. 2009. The conll-2009 shared task: Syntactic and semantic dependencies in multiple languages. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning: Shared Task*, pages 1–18.
- Karl Moritz Hermann, Dipanjan Das, Jason Weston, and Kuzman Ganchev. 2014. Semantic frame identification with distributed word representations. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, pages 1448–1458, Baltimore, Maryland, June. Association for Computational Linguistics.
- Fei Huang and Alexander Yates. 2010. Open-domain semantic role labeling by modeling word spans. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 968–978, Uppsala, Sweden, July.
- Eric Huang, Richard Socher, Christopher Manning, and Andrew Ng. 2012. Improving word representations via global context and multiple word prototypes. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 873–882, Jeju Island, Korea, July.
- Rémi Lebret and Ronan Collobert. 2014. Word embeddings through hellinger pca. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pages 482–490, Gothenburg, Sweden, April. Association for Computational Linguistics.
- A. Meyers, R. Reeves, C. Macleod, R. Szekely, V. Zielinska, B. Young, and R. Grishman. 2004. The nombank project: An interim report. In A. Meyers, editor, *HLT-NAACL 2004 Workshop: Frontiers in Corpus Annotation*, pages 24–31, Boston, Massachusetts, USA, May.

- Jeff Mitchell and Mirella Lapata. 2010. Composition in Distributional Models of Semantics. *Cognitive Science*, 34(8):1388–1429.
- Andriy Mnih and Geoffrey Hinton. 2009. A scalable hierarchical distributed language model. In *Advances in Neural Information Processing Systems*, volume 21, pages 1081–1088.
- Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71–106.
- Michael Roth and Anette Frank. 2012. Aligning predicate argument structures in monolingual comparable texts: A new corpus for a new task. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics*, pages 218–227, Montreal, Canada, June.
- Josef Ruppenhofer, Michael Ellsworth, Miriam R. L. Petruck, Christopher R. Johnson, and Jan Schefczyk. 2010. *FrameNet II: Extended Theory and Practice*.
- 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*, pages 1201–1211, Jeju Island, Korea, July. Association for Computational Linguistics.
- Ivan Titov and Alexandre Klementiev. 2011. A bayesian model for unsupervised semantic parsing. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 1445–1455, Portland, Oregon, USA, June.
- Ivan Titov and Alexandre Klementiev. 2012. Semi-supervised semantic role labeling: Approaching from an unsupervised perspective. In *Proceedings of COLING 2012*, pages 2635–2652, Mumbai, India, December.
- Joseph Turian, Lev-Arie Ratinov, and Yoshua Bengio. 2010. Word representations: A simple and general method for semi-supervised learning. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 384–394, Uppsala, Sweden, July. Association for Computational Linguistics.
- Alexander Yeh. 2000. More accurate tests for the statistical significance of result differences. In *Proceedings of the 18th International Conference on Computational Linguistics*, pages 947–953, Saarbrücken, Germany, August.
- Beñat Zepirain, Eneko Agirre, Lluís Màrquez, and Mihai Surdeanu. 2013. Selectional preferences for semantic role classification. *Computational Linguistics*, 39(3):631–663.
- Hai Zhao, Wenliang Chen, Jun’ichi Kazama, Kiyotaka Uchimoto, and Kentaro Torisawa. 2009. Multilingual dependency learning: Exploiting rich features for tagging syntactic and semantic dependencies. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL 2009): Shared Task*, pages 61–66, Boulder, Colorado, USA, June.