

SpRL-CWW: Spatial Relation Classification with Independent Multi-class Models

Eric Nichols
Honda Research Institute Japan Co., Ltd.
e.nichols@jp.honda-ri.com

Fadi Botros
University of Calgary
fnmbotro@ucalgary.ca

Abstract

In this paper we describe the SpRL-CWW entry into SemEval 2015: Task 8 SpaceEval. It detects spatial and motion relations as defined by the ISO-Space specifications in two phases: (1) it detects spatial elements and spatial/motion signals with a Conditional Random Field model that uses a combination of distributed word representations and lexico-syntactic features; (2) given relation candidate tuples, it simultaneously detects relation types and labels the spatial roles of participating elements by using a combination of syntactic and semantic features in independent multi-class classification models for each relation type. In evaluation on the shared task data, our system performed particularly well on detection of elements and relations in unannotated data.

1 Introduction

Understanding human language about location and motion is important for many applications including robotics, navigation systems, and wearable computing. Shared tasks dedicated to the problem of representing and detecting spatial and motion relations have been organized for SemEval 2012 (Kordjamshidi et al., 2012), 2013 (Kolomiyets et al., 2013), and 2015. In this paper we present SpRL-CWW, our entry to SemEval 2015 Task 8: SpaceEval, and present extended evaluation of our system to investigate the impact of the task annotations and system configurations on task performance.

2 SpaceEval Task Definition

Kordjamshidi et al. (2011) proposed the task of Spatial Role Labeling (SpRL) to detect spatial and motion relations in text. SpRL was modeled after

semantic role labeling (see (Fillmore et al., 2003; Màrquez et al., 2008)), with spatial indicators instead of predicates signaling the presence of relations, and spatial roles instead of semantic roles.

A canonical example of a spatial relation from (Kordjamshidi et al., 2011) is:

- (1) *Give me the [grey book]_{TR} [on]_{SP} the [large table]_{LM}.*

The spatial indicator (*SP*) *on* indicates that there is a *spatial* relation between the trajector (*TR*; primary object of spatial focus) and the landmark (*LM*; secondary object of spatial focus). SpRL was formalized as a task of classifying tuples of $\langle w_{SP}, w_{TR}, w_{LM} \rangle$ as spatial relations or not.

The SpRL task was reformulated and reintroduced in SpaceEval¹ using the ISO-Space annotation specifications (Pustejovsky et al., 2012). The biggest change was the decoupling of the semantic *type* and *role* of spatial relation arguments. A taxonomy of Spatial Element (SE) types was introduced to describe the meaning of arguments independent of their participation in relations, and spatial roles were treated as instance-specific annotations on spatial and motion relations.

The SE types introduced are: SPATIAL_ENTITY, PATH, PLACE, MOTION, NON_MOTION_EVENT, and MEASURE. Two types were also introduced to represent expressions that indicated the presence of relations: SPATIAL_SIGNAL and MOTION.

Spatial and motion relations were redefined as:

- MOVELINK: motion relation
- QSLINK: qualitative spatial relation
- OLINK: spatial orientation relation

¹<http://alt.qcri.org/semEval2015/task8/>

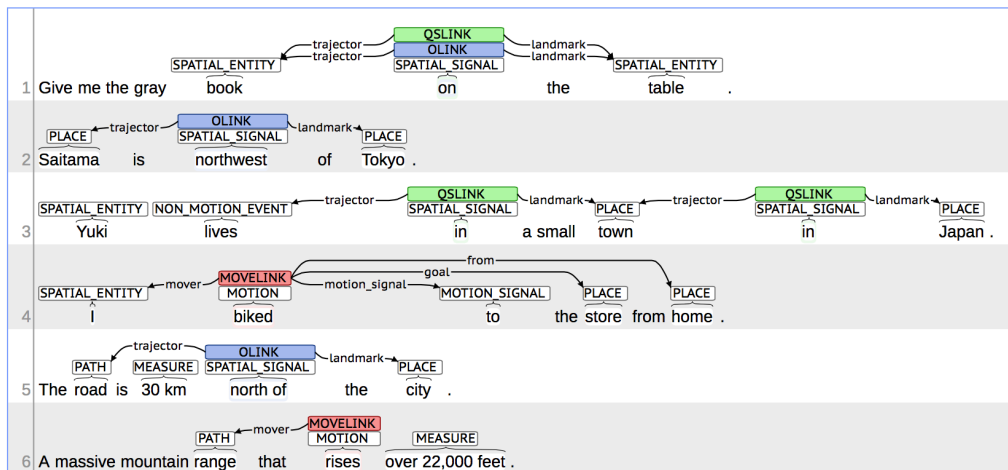


Figure 1: Example relations from the SpaceEval shared task. Only annotations that are targets are shown.

1. Only Unannotated Text is Provided
 - a. SE: precision, recall, and F1
 - b. SE: precision, recall, and F1 for each type, and an overall precision, recall, and F1
 - d. MOVELINK, QSLINK, OLINK: precision, recall, and F1
 - e. MOVELINK, QSLINK, OLINK: precision, recall, and F1 for each attribute, and an overall precision, recall, and F1
3. Spatial Elements, their Types, and their Attributes are Provided
 - a. MOVELINK, QSLINK, OLINK: precision, recall, and F1
 - b. MOVELINK, QSLINK, OLINK: precision, recall, and F1 for each attribute, and an overall precision, recall, and F1

Figure 2: SpaceEval task configurations participated in by SpRL-CWW.

Examples of SpaceEval annotations are given in Figure 1. The training data for SpaceEval consists of portions of the corpora from past SemEval SpRL tasks as well as a new dataset consisting of passages from guidebooks. Following the schema described in this section, a total of 6,782 spatial elements and signals comprising 2,186 relations were annotated.

We participated in the task configurations given in Figure 2, as defined by the official SpaceEval task description.

3 Related Research

KUL-SKIP-CHAIN-CRF (Kordjamshidi et al., 2011) was a skip-chain CRF-based sequential labeling model. It used a combination of lexico-syntactic information and semantic role information and used *preposition templates* to represent long distance dependencies. It was used as a baseline system in the

SemEval 2012 and 2013 SpRL tasks.

UTD-SpRL (Roberts and Harabagiu, 2012) was an entry into the SemEval 2012 SpRL task that adopted a joint relation detection and role labeling approach with the motivation that roles in spatial relations were dependent on each other. The approach used heuristics to gather spatial relation candidate tuples. A hand-crafted dictionary was used to detect SPATIAL_INDICATOR candidates, and noun phrase heads were treated as TRAJECTOR and LANDMARK candidates. A model for relation classification and role labeling was then trained with libLINEAR using POS, lemma, and dependency-path-based features, with feature selection used to prune away ineffective features.

UNITOR-HMM-TK (Bastianelli et al., 2013) was an entry into the SemEval 2013 SpRL task. It used a pipeline approach with three sub-tasks: (1) spatial indicator detection, (2) spatial role² classification and (3) spatial relation identification.

Spatial indicators and roles were detected with sequential labeling using SVM^{hmm} with detected indicators used as features for spatial role labeling. In addition, shallow grammatical features in the form of POS n-grams were used in place of richer syntactic information in order to avoid overfitting. The model also used PMI-score based word space representations as described in (Sahlgren, 2006).

UNITOR-HMM-TK’s approach to spatial relation identification avoided feature engineering by employing an SVM model with a smoothed partial tree

²Referred to as *spatial annotations* in the paper.

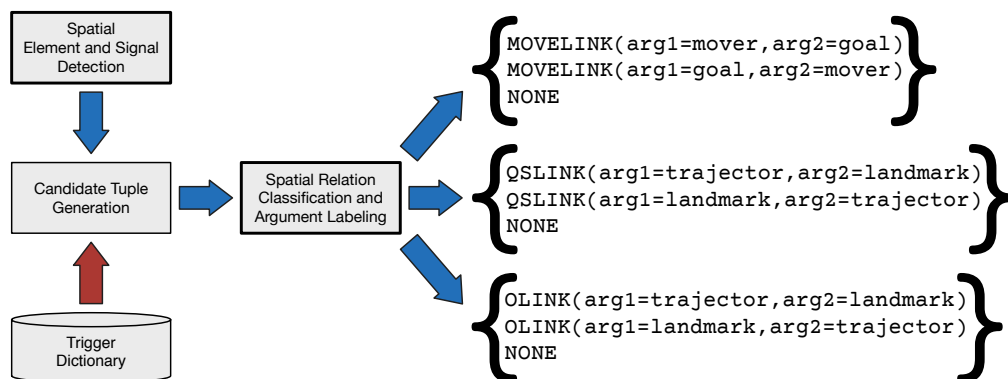


Figure 3: The SpRL-CWW system architecture. Spatial elements and signals are detected, from which relation candidate tuples are generated, and then relations with their arguments labeled are identified by a separate classifier for each relation type. The red arrow indicates special trigger dictionary processing that is only carried out for SpaceEval tasks 1d and 1e, and for Setting F of the relation classification task extended evaluation in Table 3.

EF.1	Raw string in a 5-word window (i.e. <i>Saitama is <u>northwest</u> of Tokyo</i>)
EF.2	Lemma in a 5-word window (i.e. <i>Saitama be <u>northwest</u> of Tokyo</i>)
EF.3	POS in a 5-word window (i.e. <i>NNP VBZ <u>RB</u> IN NNP</i>)
EF.4	Named Entity in a 5-word window (i.e. <i>LOC NONE <u>NONE</u> NONE LOC</i>)
EF.5	Lemma concatenated with the POS in a 3-word window (i.e. <i>be::VBZ <u>northwest::RE</u> of::IN</i>)
EF.6	Named Entity concatenated with the POS in a 3-word window (i.e. <i>NONE::VBZ <u>NONE::RB</u> NONE ::IN</i>)
EF.7	Direct dependency on the head of the sentence if present (i.e. <i>advmod</i>)
EF.8	Direct dependency on the head of the sentence concatenated with the lemma of the head (i.e. <i>advmod::be</i>)
EF.9	300-dimension GloVe word vector
EF.10	POS bigrams for a 5-word window (i.e. <i>NNP_VBZ VBZ_RB RB_IN IN_NNP</i>)
EF.11	Raw string n-grams for 3-word window (i.e. <i>is_northwest northwest_of</i>)

Figure 4: Features for spatial element/signal detection for the sentence “*Saitama is northwest of Tokyo.*”

kernel over modified dependency trees to capture syntactic information.

More recent work on spatial relation identification includes (Kordjamshidi and Moens, 2014).

4 Spatial Element and Signal Detection

4.1 Approach

SpRL-CWW uses a feature-rich CRF model to jointly label spatial elements and spatial/motion signals. Previous approaches (Kordjamshidi et al., 2011;

Bastianelli et al., 2013) proposed a two-step sequential labeling method for this task. In the first step, they label spatial signals³ since they indicate the presence of a relation, which spatial roles depend on. In the second step, they label all the other spatial roles in the sentence using the extracted signals as features. However, any errors made in the first step will deteriorate the performance of the second. Furthermore, for SpaceEval 2015 the spatial element annotations are less likely to depend on the presence of a relation and can be detected independently. Thus, our system avoids the performance degradation associated with pipeline approaches by combining the two steps.

SpRL-CWW’s CRF model labels each word in a sentence with one of the labels described in Section 2, or with NONE. In line with UNITOR-HMM-TK (Bastianelli et al., 2013), shallow lexico-syntactic features are applied instead of the full syntax of the sentence to avoid over-fitting the training data. We use word vectors trained on Web-scale corpora for a fine-grained lexical representation.

An example of our feature representation for the sentence “*Saitama is northwest of Tokyo.*” is given in Figure 4.

4.2 Evaluation

4.2.1 Setup

Sentences were processed with Stanford CoreNLP (Manning et al., 2014) for POS tagging, lemmatiza-

³Also known as *spatial indicators*.

Task	Overall Precision	Overall Recall	Overall F1	Mean F1	Overall Accuracy
1a	0.84	0.83	0.83	0.83	0.89
1b	0.77	0.76	0.76	0.76	0.91
1d	0.56	0.51	0.53	0.40	0.57
1e	0.03	0.04	0.03	0.03	0.25
3a	0.78	0.57	0.66	0.57	0.86
3b	0.05	0.06	0.05	0.05	0.48

Table 1: Official SpaceEval submission results.

Label	Training			Test		
	5-fold cross validation			P	R	F1
MEASURE	0.889	0.707	0.788	0.869	0.726	0.791
MOTION	0.823	0.700	0.756	0.808	0.733	0.769
MOTION_SIGNAL	0.766	0.600	0.673	0.801	0.772	0.786
NON_MOTION_EVENT	0.663	0.371	0.476	0.688	0.478	0.564
PATH	0.815	0.614	0.701	0.759	0.519	0.617
PLACE	0.802	0.777	0.789	0.742	0.752	0.747
SPATIAL_ENTITY	0.793	0.653	0.716	0.858	0.763	0.808
SPATIAL_SIGNAL	0.750	0.603	0.668	0.740	0.681	0.709
OVERALL	0.795	0.674	0.730	0.785	0.712	0.746

Table 2: Spatial Element/Signal detection results on training data and test data. Results are reproduced independently of official evaluation.

tion, NER, and dependency parsing. The word representations are publicly-available 300-dimension GloVe⁴ word vectors trained on 42 billion tokens of Web data (Pennington et al., 2014). The model was trained using CRFsuite (Okazaki, 2007) with L-BFGS using L2 regularization with $\lambda_2 = 1 * 10^{-5}$.

4.2.2 Datasets

We evaluated our system on the SpaceEval training data as described in Section 2, and additionally on the SpaceEval Task 3 test data, which was distributed with gold labeled Spatial Elements, Indicators, and Motions. The test data consisted of 16 files with 317 sentences and 1,609 spatial roles.

4.2.3 Results

Official task results for spatial element/signal identification (Task 1a) and classification (Task 1b) are shown in Table 1.

We performed more detailed evaluation using 5-fold cross validation on the training data and on the released gold test data. Our results are presented in Table 2. These results and have an f1-score that is slightly lower than the official reported result.⁵Evaluation over the test data produced a

⁴<http://www-nlp.stanford.edu/projects/glove/>

⁵As the official evaluation data and scripts have not been fully released at the time of writing, it is not possible to determine the cause of the discrepancy in f1-scores. Comparison between strict and “relaxed” matching as used in prior SemEval SpRL tasks did not account for the difference.

Features representing the extracted trigger:	
RF.1	Raw string
RF.2	Lemma
RF.3	POS
RF.4	<u>RF.2 concatenated with RF.3</u>
Features representing each of the two arguments:	
RF.5	Raw string
RF.6	Lemma
RF.7	POS
RF.8	<u>RF.6 concatenated with RF.7</u>
RF.9	Spatial element type (i.e Place, Path, etc.)
RF.10	<u>RF.9 of each argument concatenated together</u>
RF.11	<u>RF.10 concatenated with RF.2</u>
RF.12	Direction of the argument with the respect to the extracted trigger (i.e left/right)
RF.13	<u>RF.12 of each argument concatenated together</u>
RF.14	<u>RF.13 concatenated with RF.2</u>
RF.15	Boolean value representing whether there are other spatial elements in between the argument and the extracted trigger
RF.16	RF.15 of each argument concatenated together
RF.17	Dependency path between the argument and the extracted trigger (i.e. $\uparrow conj \downarrow dep \downarrow nsubj$)
RF.18	<u>RF.17 of each argument concatenated together</u>
RF.19	Dependency path between the two arguments
RF.20	Length of the dependency path between the argument and the extracted trigger
RF.21	Bag-of-words of tokens in between the argument and the extracted trigger
RF.22	Number of tokens in between the argument and the extracted trigger
RF.23	RF.22 of each argument added together
RF.24	Boolean value representing whether either of the arguments are null values
Features representing the spatial elements that are directly to the left and to the right of the trigger:	
RF.25	Raw string
RF.26	Lemma
RF.27	POS
RF.28	<u>RF.26 concatenated with RF.27</u>
RF.29	Number of tokens in between the spatial element and the extracted trigger

Figure 5: Features for joint spatial relation classification and role labeling. Underlined features are withheld from quadratic feature Settings D and E of Table 3.

slightly higher f1-score than on the training data. We theorize that this is due to cross-fold validation using a smaller dataset for its model.

5 Spatial Relation Classification and Argument Labeling

5.1 Approach

To identify spatial relations, the SpRL-CWW system determines which spatial elements and signals, can be combined to form valid spatial relations. Since the type of a relation (MOVELINK, QSLINK, or OLINK) is dependent upon its arguments, our method, inspired by UTD-SpRL (Roberts and Harabagiu, 2012), jointly classifies spatial relations and labels participating arguments in one classification step. We aim to simplify our model and improve learning by only

Setting	Regularization	Features	SEs	Triggers	P	R	F1
A	no	all	gold	gold	0.560	0.500	0.527
B	yes	all	gold	gold	0.599	0.496	0.544
C	yes	-SE types	gold	gold	0.597	0.430	0.500
D	yes	all + semantic types	gold	gold	0.636	0.501	0.561
E	yes	pre-quadratic	gold	gold	0.575	0.411	0.480
F	yes	quadratic	gold	gold	0.762	0.345	0.463
G	yes	all	predicted	dictionary	0.423	0.427	0.425
H	yes	all	predicted	predicted	0.382	0.364	0.372

Table 3: Settings for extended relation detection evaluation over the SpaceEval 2015 training data. All evaluation is conducted with 5-fold cross validation, the full RE feature set from Figure 5, gold standard SEs, and gold standard triggers. The overall precision, recall, and f1-scores are reported for each setting with the highest performing in bold. Setting A was used for our official submission. Where indicated, L2 regularization was performed with $\lambda_2 = 1 * 10^{-14}$.

considering relations that contain a trigger and by labeling only the following attributes which correspond to primary spatial and motion roles:

- MOVELINK: trigger, mover, goal
- QSLINK and OLINK: trigger, trajectory, landmark

5.1.1 Candidate Trigger Extraction

First, candidate triggers are extracted from each sentence. The model we presented for detecting signals in Section 4.1 has a high f1-score but low precision. Because we want to prioritize recall for generating candidate tuples, when classifying relations on unannotated text, dictionaries of triggers automatically compiled from the training data are used to extract potential triggers from sentences. These dictionaries are used in Task 1d and 1e in Figure 2. In Task 3, where gold spatial roles are provided, MOTIONS are used as potential MOVELINK triggers, SPATIAL_SIGNALS are used as potential QSLINK and OLINK triggers. Evaluation of the trigger dictionaries shows that they have much higher recall than CRF models⁶. Additional relation classification evaluation in Table 3 show that the dictionaries (Setting F) achieve an f1-score improvement of 0.055 over the CRF models (Setting G).

5.1.2 Candidate Tuple Generation

All possible candidate relations in a sentence are then generated using the extracted triggers and the spatial elements in the sentence. A candidate tuple consists of an extracted trigger and two other spatial elements: arg1 and arg2. Since some relations, such as the one represented in Figure 1 Example 6, can have undefined arguments, tuples with undefined arguments are also generated. For Example 4

⁶In particular, recall for SPATIAL_SIGNALS increases from 0.603 to 0.936 and MOTION recall increases from 0.700 to 0.812 on the SpaceEval test data.

in Figure 1, the following candidate tuples will be generated for MOVELINK classification:

- < trigger:biked, arg1:I, arg2:store >
- < trigger:biked, arg1:I, arg2:home >
- < trigger:biked, arg1:I, arg2: \emptyset >
- < trigger:biked, arg1:home, arg2:store >
- < trigger:biked, arg1:home, arg2: \emptyset >
- < trigger:biked, arg1:store, arg2: \emptyset >

Each tuple is represented by three main groups of features outlined in Figure 5. A one-against-all multi-class classifier is then applied to classify each candidate relation tuple into one of three possible classes. Three independent classifiers are trained, one for each spatial relation type, using Vowpal Wabbit (Agarwal et al., 2011). The classes used by the MOVELINK classifier are:

Class 1 - REL(arg1=mover, arg2=goal)

Class 2 - REL(arg1=goal, arg2=mover)

Class 3 - NONE

The classes used by the QSLINK and OLINK classifiers are:

Class 1 - REL(arg1=trajectory, arg2=landmark)

Class 2 - REL(arg1=landmark, arg2=trajectory)

Class 3 - NONE

5.2 Evaluation

5.2.1 Setup

Once again, Stanford CoreNLP was used for POS tagging, lemmatization and dependency parsing. The classification models were trained with Vowpal Wabbit’s one-against-all multi-class classifier using its online stochastic gradient descent implementation with all the default settings.

Relation Type	P	R	F1
QSLINK	0.661	0.538	0.594
MOVELINK	0.571	0.451	0.504
OLINK	0.691	0.517	0.591
OVERALL	0.636	0.501	0.561

Table 4: SpRL-CWW’s relation classification results for the highest-performing Setting D.

5.2.2 Datasets

We evaluated our system on the trial and training data that was released for SpaceEval, with the exception of 9 files that didn’t have spatial relations annotated. Since our system focuses on relations with a trigger, we filtered out the relations that contained no trigger. The resulting dataset of 1,801 relations was used for training and evaluation.

5.2.3 Results

Official task results for relation classification are shown in Table 1. Task 1d results use the SEs that were detected in the previous step (Task 1b). Task 3a results are for relation classification using gold spatial elements and signals.

6 Discussion

Participation in SpaceEval raised several questions which we attempt to answer by conducting extended evaluation of our system on the SpaceEval training data using 5-fold cross validation⁷. The settings and results are summarized in Table 3.

Which features were effective?

The feature ablation results in Table 5 show the three features with the largest contribution to SE and SI classification. They verify the contribution of word vectors trained on Web-scale data and support Bastianelli’s et al. (2013)’s claim that shallow grammatical information is essential.

Does the fine-grained SpaceEval annotation scheme help or hinder?

In order to explore this, we compare the top performing setting with SE type-related features (Setting B) to a setting with them removed (Setting C). Absence of these features decrease the f1-score by 0.044, providing evidence that fine-grained SE types help relation classification, though the relation and spatial role taxonomy requires consideration.

⁷Partitions were made by taking a stratified split of the document set when ordered by decreasing size.

Features	P	R	F1	$\Delta F1$
all	0.795	0.674	0.730	-
-EF.1	0.807	0.604	0.691	-0.039
-EF.9	0.808	0.602	0.690	-0.040
-EF.10	0.761	0.600	0.671	-0.059

Table 5: The three spatial element classification features with the largest delta in feature ablation.

Furthermore, each gold Spatial Signal that was provided for Task 3 had one of three possible semantic types; *DIRECTIONAL*, *TOPOLOGICAL* or *DIR_TOP* (both). Instead of using all Spatial Signals as candidate triggers for QSLINKs and OLINKs, we only considered *TOPOLOGICAL* Spatial Signals as candidate triggers for QSLINK and *DIRECTIONAL* Spatial Signals as candidate triggers for OLINK. This setting (Setting D) achieved the highest f1-score and recall, demonstrating the importance of Spatial Signal semantic types in relation classification. Full relation classification results for Setting D are summarized in Table 4.⁸

Is less (or no) feature engineering feasible?

We attempt this by automatically generating features using Vowpal Wabbit’s quadratic feature generation. We disable all features underlined in Figure 5) and instruct VW to automatically construct features by generating all possible feature combinations. Settings E and F compare the base feature set before and after quadratic features are added. While quadratic features achieve a lower f1-score, they have the highest precision of all settings, suggesting feature generation may be useful for increasing precision of relation classification, but the low f1-score of Setting F indicates care is needed in selecting the base feature set. We are exploring feature engineering reduction further with a phrase vector-based model inspired by (Hermann et al., 2014).

7 Conclusion

In this paper we presented the SpRL-CWW entry to SpaceEval 2015: Task 8. Official evaluation showed that it performed especially well on unannotated data. Extended evaluation verified the contribution of Web-scale word vectors, trigger dictionaries, and SE type information; and automatic feature generation showed promise. For future work, we plan to explore phrase vector-based approaches to SpRL.

⁸We thank an anonymous reviewer for the suggestion to use Spatial Signal semantic types.

Acknowledgments

This research was supported by Honda Research Institute Japan, Co., Ltd. We also thank the anonymous reviewers for their many fruitful suggestions.

References

- Alekh Agarwal, Olivier Chapelle, Miroslav Dudík, and John Langford. 2011. A reliable effective terascale linear learning system. *CoRR*, abs/1110.4198.
- Emanuele Bastianelli, Danilo Croce, Roberto Basili, and Daniele Nardi. 2013. UNITOR-HMM-TK: Structured kernel-based learning for spatial role labeling. *Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*.
- Charles J. Fillmore, Christopher R. Johnson, and Miriam R.L. Petruck. 2003. Background to FrameNet. *International Journal of Lexicography*, 16.3:235–250.
- Karl Moritz Hermann, Dipanjan Das, Jason Weston, and Kuzman Ganchev. 2014. Semantic frame identification with distributed word representations. In *Proceedings of ACL*, June.
- Oleksandr Kolomiyets, Parisa Kordjamshidi, Steven Bethard, and Marie-Francine Moens. 2013. SemEval-2013 task 3: Spatial role labeling. In *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013), Second joint conference on lexical and computational semantics, Atlanta, USA, 14-15 June 2013*, pages 255–266.
- Parisa Kordjamshidi and Marie-Francine Moens. 2014. Global machine learning for spatial ontology population. *Web Semantics: Science, Services and Agents on the World Wide Web*.
- Parisa Kordjamshidi, Martijn Van Otterlo, and Marie-Francine Moens. 2011. Spatial role labeling: Towards extraction of spatial relations from natural language. *ACM Transactions on Speech and Language Processing (TSLP)*, 8(3):4.
- Parisa Kordjamshidi, Steven Bethard, and Marie-Francine Moens. 2012. SemEval-2012 task 3: Spatial role labeling. In **SEM 2012: The First Joint Conference on Lexical and Computational Semantics – Volume 1: Proceedings of the Main Conference and the Shared Task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012), SemEval-2012*, pages 365–373, June.
- Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 55–60.
- Lluís Màrquez, Xavier Carreras, Kenneth C Litkowski, and Suzanne Stevenson. 2008. Semantic role labeling: an introduction to the special issue. *Computational Linguistics*, 34(2):145–159.
- Naoaki Okazaki. 2007. CRFsuite: a fast implementation of conditional random fields (CRFs). <http://www.chokkan.org/software/crfsuite/>.
- Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Doha, Qatar, October.
- James Pustejovsky, Jessica Moszkowicz, and Marc Verhagen. 2012. A linguistically grounded annotation language for spatial information. *TAL*, 53(2).
- Kirk Roberts and Sanda Harabagiu. 2012. UTD-SpRL: A joint approach to spatial role labeling. In **SEM 2012: The First Joint Conference on Lexical and Computational Semantics – Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)*, pages 419–424.
- Magnus Sahlgren. 2006. *The Word-Space Model*. Ph.D. thesis, University of Stockholm (Sweden).