UNITOR-HMM-TK: Structured Kernel-based Learning
for Spatial Role Labeling

Emanuele Bastianelli[(†)(⋆)], **Danilo Croce**[(‡)], **Daniele Nardi**[(⋆)], **Roberto Basili**[(‡)]

(†) DICII	(‡) DII	(⋆) DIAG
University of Roma Tor Vergata	University of Roma Tor Vergata	University of Roma La Sapienza
Rome, Italy	Rome, Italy	Rome, Italy
{bastianelli}@ing.uniroma2.it	{croce,basili}@info.uniroma2.it	{nardi}@dis.uniroma1.it

Abstract

In this paper the `UNITOR-HMM-TK` system participating in the Spatial Role Labeling task at SemEval 2013 is presented. The spatial roles classification is addressed as a sequence-based word classification problem: the SVM^{hmm} learning algorithm is applied, based on a simple feature modeling and a robust lexical generalization achieved through a Distributional Model of Lexical Semantics. In the identification of spatial relations, roles are combined to generate candidate relations, later verified by a SVM classifier. The Smoothed Partial Tree Kernel is applied, i.e. a convolution kernel that enhances both syntactic and lexical properties of the examples, avoiding the need of a manual feature engineering phase. Finally, results on three of the five tasks of the challenge are reported.

1 Introduction

Referring to objects or entities in the space, as well as to relations holding among them, is one of the most important functionalities in natural language understanding. The detection of spatial utterances thus finds many applications, such as in GPS navigation systems, or Human-Robot Interaction (HRI).

In Computational Linguistics, the task of recognizing spatial information is known as Spatial Role Labeling (SpRL), as discussed in (KordJamshidi et al., 2010). Let us consider the sentence:

[*A man*]$_{\text{TRAJECTOR}}$ *is sitting* [*on*]$_{\text{SPATIAL\_INDICATOR}}$

[*a chair*]$_{\text{LANDMARK}}$ *and talking on the phone.*   (1)

where three roles are labeled: the phrase "*A man*" refers to a TRAJECTOR, "*a chair*" to a LAND-

MARK and they are related by the spatial expression "*on*" denoted as SPATIAL INDICATOR. The last role establishes the type of the spatial relation, e.g. `Regional`. The ambiguity of natural language makes this task very challenging. For example, in the same Example 1, another preposition "*on*" can be considered, but the phrase "*the phone*" is not a spatial role, as it refers to a communication mean. This mainly depends on the semantics of the grammatical head words, i.e. *chair* and *phone*. Such phenomena are crucial in many learning frameworks, as in kernel-based learning (Shawe-Taylor and Cristianini, 2004), where the decision is based on the similarity between training and testing data.

This paper describes the `UNITOR-HMM-TK` system participating in the Semeval 2013 Spatial Role Labeling Task (Kolomiyets et al., 2013), addressing three of the five defined sub-tasks:

- **Task A: Spatial Role Classification**. It consists in labeling short sentences with spatial roles among SPATIAL INDICATOR, TRAJECTOR and LANDMARK.

- **Task B: Relation Identification**. It consists in the identification of relations among roles identified in Task A. This task does not involve the semantic relation classification.

- **Task C: Spatial Role Classification**. It consists in labeling short documents with spatial roles among the extended role set: TRAJECTOR, LANDMARK, SPATIAL INDICATOR, MOTION_INDICATOR, PATH, DIRECTION and DISTANCE.

The `UNITOR-HMM-TK` system addresses both the problems of identifying spatial roles and relations as a sequence of two main classification steps.

573

In the first step, each word in the sentence is classified by a sequence-based classifier with respect to the possible spatial roles. It is in line with other methods based on sequence-based classifier for SpRL (Kordjamshidi et al., 2011; Kordjamshidi et al., 2012b). Our labeling has been inspired by the work in (Croce et al., 2012), where the SVM^{hmm} learning algorithm, formulated in (Altun et al., 2003), has been applied to the classical FrameNet-based Semantic Role Labeling. The main contribution in (Croce et al., 2012) is the adoption of shallow grammatical features (e.g. POS-tag sequences) instead of the full syntax of the sentence, in order to avoid over-fitting over training data. Moreover, lexical information has been generalized through the use of a Word Space, in line with (Schutze, 1998; Sahlgren, 2006): it consists in a Distributional Model of Lexical Semantics derived from the unsupervised analysis of an unlabeled large-scale corpus. The result is a geometrical space where words with similar meaning, e.g. involved in a paradigmatic or almost-synonymic relations, will be projected in similar vectors. As an example, we expect that a word like "*table*", maybe a LANDMARK in a training example, is more similar to "*chair*" as compared with "*phone*".

In the second step, all roles found in a sentence are combined to generate candidate relations, which are then verified by a Support Vector Machine (SVM) classifier. As the entire sentence is informative to determine the proper conjunction of all roles, we apply a kernel function within the classifier, that enhances both syntactic and lexical information of the examples. We adopted the *Smoothed Partial Tree Kernel* (SPTK), defined in (Croce et al., 2011): it is convolution kernel that allows to measure the similarity between syntactic structures, which are partially similar and whose nodes can differ, but are semantically related. Each example is represented as a tree structure directly derived from the sentence dependency parse, thus avoiding the manual definition of features. Similarity between lexical nodes is measured in the same Word Space mentioned above.

In the rest of the paper, Section 2 discusses the SVM^{hmm} based approach. The SPTK-based learning algorithm will be presented in Section 3. Finally, results obtained in the competition are discussed in Section 4.

2 Sequential Tagging for Spatial Role Classification

The system proposed for the *Spatial Role Classification* task is based on the SVM^{hmm} formulation discussed in (Altun et al., 2003). It extends classical SVMs by learning a discriminative model isomorphic to a k-order Hidden Markov Model through the Structural SVM formulation (Tsochantaridis et al., 2005). In the discriminative view of SVM^{hmm}, given an observed input word sequence $\mathbf{x} = (x_1 \ldots x_l) \in \mathcal{X}$ of feature vectors $x_1 \ldots x_l$, the model predicts a sequence of labels $\mathbf{y} = (y_1 \ldots y_l) \in \mathcal{Y}$ after learning a linear discriminant function $F : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ over input/output pairs. Each word is then modeled as a set of linear features that express lexical information as well as syntactic information surrogated by POS n-grams. With respect to other works using SVM^{hmm} for SpRL, such as (Kordjamshidi et al., 2012b), we investigate another set of possible features, as the ones proposed in (Croce et al., 2012): the aim is to provide an agile system that takes advantages in adopting only shallow grammatical features, thus ignoring the full syntactic information of a sentence. The syntactic features derived from a dependency parse process are surrogated by POS n-grams. According to this, our feature modeling adopts the `IOB` notation discussed in (Croce et al., 2012). It provides a class label for each token, mapping them into artificial classes representing the beginning (`B`), the inside (`I`) or ending (`O`) of a spatial role, plus the label of the classified role (i.e. B$_{\text{SPIND}}$ for the starting token of a SPATIAL INDICATOR); words external to every role are labeled with the special class (`_`). According to this notation, the labeling of Example 1 can be expressed as follows: "*A*/B$_{\text{TRAJ}}$ *man*/O$_{\text{TRAJ}}$ *is*/_ *sitting*/_ *on*/B$_{\text{SPIND}}$ *a*/B$_{\text{LAND}}$ *chair*/O$_{\text{LAND}}$ *and*/_ . . .*".

In order to reduce the complexity of the entire classification task, two phases are applied. In Task A, as in (Kordjamshidi et al., 2011), the first phase aims at labeling only SPATIAL INDICATOR, as they should relate remaining spatial expressions. For the same reason, in Task C we first label only SPATIAL INDICATOR and MOTION INDICATOR. Roles classified in this step are considered *pivot* and they can be used as features for the classification of the other roles: TRAJECTORS and LANDMARKS for

Task A while TRAJECTORS, LANDMARKS, PATHS, DISTANCES and DIRECTIONS for Task C.

For the classification of SPATIAL and MOTION INDICATOR, each word, such as the first "*on*" occurrence in the Example 1, is modeled through the following features: its lemma (*on*) and POS tag (`IN`); the left and right lexical contexts, represented by the n words before (*man*::`NN` *is*::`VBZ` *sitting*::`VBG`) and after (*a*::`DT` *chair*::`NN` *and*::`CC`); the left and right syntactic contexts as the POS n-grams occurring *before* (i.e. `NN_VBZ VBZ_VBG NN_VBZ_VBG`) and *after* (i.e. `DT_NN NN_CC DT_NN_CC`) the word.

For the TRAJECTOR and LANDMARK classification in Task A, each word is represented by the same features described above, plus the following ones (with respect to Example 1, the token relative to the word *man*): lemma of the SPATIAL INDICATOR (*on*); *Positional Feature*: distance from the SPATIAL INDICATOR in terms of number of tokens (-3); relative position with respect to the SPATIAL INDICATOR, that is *before* or *after* (*before*); a boolean feature that indicates whether or not the current token is a SPATIAL INDICATOR; the number of words composing the SPATIAL INDICATOR (here 1).

In Task C, for the classification with respect to the complete set of roles, each word is modeled by the previous features together with the following: distance from the MOTION INDICATOR in terms of number of tokens; relative position with respect to the MOTION INDICATOR (before and after); a boolean feature that indicates whether or not the current token is a MOTION INDICATOR; the number of words that composes the MOTION INDICATOR. In both Tasks A and C the symbols `SI` and `MI` to represent a SPATIAL INDICATOR or a MOTION INDICATOR are used respectively to represent the target pivot role within any n-gram.

In order to increase the robustness of our modeling, we extended the lexical information with features derived from a distributional analysis over large texts. In essence, we represent the lexical semantic similarity between different words with similar meaning. We extend a supervised approach through the adoption of vector based models of lexical meaning: a large-scale corpus is statistically analyzed and a Word Space, (Sahlgren, 2006), is acquired as follows. A word-by-context matrix M is obtained through a large scale corpus analysis.

Then the *Latent Semantic Analysis* (Landauer and Dumais, 1997) technique is applied to reduce the space dimensionality. Moreover it provides a way to project a generic word w_i into a k-dimensional space where each row corresponds to the representation vector $\vec{w_i}$. In such a space, the distance between vectors reflects the similarity between corresponding words. The resulting feature vector representing w_i is then augmented with $\vec{w_i}$, as in (Croce et al., 2010), where the benefits of such information have been reported in the FrameNet-based Semantic Role Labeling task.

3 Relation identification

The `UNITOR-HMM-TK` system tackles *Relation Identification* task by determining which spatial roles, discovered in the previous classification phase, can be combined to determine valid spatial relations. Our method is inspired by the work of (Roberts and Harabagiu, 2012), where all possible spatial roles are first generated through heuristics and then combinatorially combined to acquire candidate relations; valid spatial relations are finally determined using a SVM classifier. We aim at reducing the potentially huge search space, by considering only spatial roles proposed by our sequential tagging approach, described in Section 2. Most importantly, we avoid the manual feature engineering phase of (Roberts and Harabagiu, 2012). Candidate relations are not represented as vectors, whose dimensions are manually defined features useful for the target classification. We directly apply the Smoothed Partial Tree-Kernel (SPTK), proposed in (Croce et al., 2011), to estimate the similarity among a specific tree representation.

Tree kernels exploit syntactic similarity through the idea of convolutions among substructures. Any tree kernel computes the number of common substructures between two trees T_1 and T_2 without explicitly considering the whole fragment space. Its general equation is reported hereafter:

$$TK(T_1, T_2) = \sum_{n_1 \in N_{T_1}} \sum_{n_2 \in N_{T_2}} \Delta(n_1, n_2)$$

where N_{T_1} and N_{T_2} are the sets of the T_1's and T_2's nodes respectively, and $\Delta(n_1, n_2)$ is equal to the number of common fragments rooted in the n_1 and n_2 nodes[1]. The SVM classifier is thus trained in

---

[1]To have a similarity score between 0 and 1, a normalization

a implicit very high-dimensional space, where each dimension reflects a possible tree sub-structure, thus avoiding the need of an explicit feature definition. The function Δ determines the nature of such space. For example, Syntactic Tree Kernel (STK) are used to model complete context free rules as in (Collins and Duffy, 2001).

The algorithm for SPTK (Croce et al., 2011) pushes for more emphasis on lexical nodes. The Δ function allows to recursively matches tree structures and lexical nodes: this allows to match fragments having same structure but different lexical nodes, by assigning a score proportional to the product of the lexical similarities, thus generalizing grammatical and lexical information in training data. While similarity can be modeled directly over lexical resources, e.g. WordNet as discussed in (Pedersen et al., 2004), their development can be very expensive, thus limiting the coverage of the resulting convolution kernel, especially in specific application domains. Again, a Word Space model is adopted: given two words, the term similarity function σ is estimated as the cosine similarity between the corresponding projections.

As proposed in (Croce et al., 2011), the SPTK is applied to examples modeled according the *Grammatical Relation Centered Tree* (GRCT) representation, which is derived from the original dependency parse structure. Figure 1 shows the GRCT for Example 1: non-terminal nodes reflect syntactic relations, such as subject (NSUBJ); pre-terminals are the POS, such as nouns (NN), and leaves are lexemes, such as *man::n*[2]. Non-terminal nodes associated with a role are enriched with the role name, e.g. NSUBJ$_{\text{TRAJ}}$. All nodes not covering any role are pruned out, so that all information not concerning spatial aspects that would introduce noise is ignored.

In this setting, positive examples are provided by considering sentences labeled by roles involved in a valid relation. The definition of negative examples is more difficult. We considered all roles labelled by the SVM^{hmm} based system, discussed in Section 2. For each incorrect labeling over the an-

---

in the kernel space, i.e. $\frac{TK(T_1, T_2)}{\sqrt{TK(T_1, T_1) \times TK(T_2, T_2)}}$ is applied.

[2]Each word is lemmatized to reduce data sparseness, but they are enriched with POS tags to avoid confusing words from different grammatical categories.
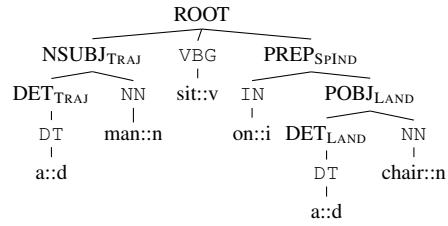

Figure 1: GRCT representation of a *positive* example derived from a correct labeling from Example 1
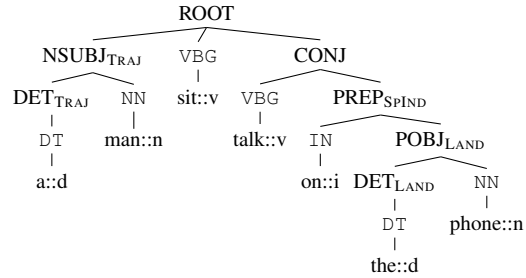

Figure 2: GRCT representation of a *negative* example derived from a wrong labeling from Example 1

notated material, a set of negative examples is acquired by combining all proposed roles. In order to avoid over-fitting, a n-fold schema has been applied: it is needed to avoid the SVM^{hmm} labeling the same sentences used for training. Moreover, constraints over the relation are imposed to avoid violations of the Spatial Role theory: in Task B each relation must be composed at least by a SPATIAL_INDICATOR, LANDMARK and a TRAJECTOR or by a SPATIAL_INDICATOR, implicit LANDMARK and a TRAJECTOR. Let us consider a possible labeling of Example 1: "[A man]$_{\text{TRAJ}}$ *is sitting* [on]$_{\text{SPIND}}$ [a chair]$_{\text{LAND}}$ *and talking* [on]$_{\text{SPIND}}$[the phone]$_{\text{LAND}}$"; here, the second SPATIAL_INDICATOR "*on*" and the LANDMARK "*the phone*" are incorrectly labeled. A negative example is thus obtained by considering these roles togheter with the TRAJECTOR "*the phone*", as shown in Figure 2. Other two negative examples can be generated by combining the remaining two roles.

4 Results

In this section experimental results of the UNITOR-HMM-TK system in the Spatial Role Labeling task at SemEval 2013 are reported. In Tasks A and B, the dataset is a corrected version

of the same training dataset employed in (Kord-jamshidi et al., 2012a)[3]. The dataset for Task C was part of the Confluence corpus[4]. More details about the dataset are provided in (Kolomiyets et al., 2013). In all experiments, sentences are processed with the Stanford CoreNLP[5], for Part-of-Speech tagging, lemmatization (Task A and C) and dependency parsing (Task B).

The sequential labeling system described in Section 2 has been made available by the SVM^{hmm} software[6]. The estimation of the semantically Smoothed Partial Tree Kernel (SPTK), described in Section 3 is made available by an extended version of SVM-LightTK software[7] (Moschitti, 2006), implementing the smooth matching between tree nodes. Similarity between lexical nodes is estimated as the cosine similarity in the co-occurrence Word Space described above, as in (Croce et al., 2011).

The co-occurrence Word Space is acquired through the distributional analysis of the UkWaC corpus (Baroni et al., 2009). First, all words occurring more than 100 times (i.e. the *targets*) are represented through vectors. The original space dimensions are generated from the set of the 20,000 most frequent words (i.e. *features*) in the UkWaC corpus. One dimension describes the Pointwise Mutual Information score between one feature, as it occurs on a left or right window of 3 tokens around a target. Left contexts of targets are treated differently from the right ones, in order to capture asymmetric syntactic behaviors (e.g., useful for verbs): 40,000 dimensional vectors are thus derived for each target. The Singular Value Decomposition is applied and the space dimensionality is reduced to $k = 100$.

4.1 Results in Task A

Two different runs were submitted for Task A. The first takes into account all roles labeled accordingly to the approach described in Section 2. Results, in term of precision, recall and F-measure for each spatial role are shown in Table 1. The second run considers only those roles composing the relations that

---
[3]The initial number of sentences was of 600, but it decreased after the elimination of 21 duplicated sentences.

[4]Three of the original 95 files were ignored because of some issues with their format. See http://confluence.org

[5]http://nlp.stanford.edu/software/corenlp.shtml

[6]http://www.cs.cornell.edu/People/tj/svm_light/svm_hmm.html

[7]http://disi.unitn.it/moschitti/Tree-Kernel.htm

are positively classified in Task B and it will be discussed in Section 4.2.

A tuning phase has been carried out through a 10-fold cross validation: it allowed to find the best classifier parameters. The evaluation of the system performances is measured using a character based measure, i.e. considering the number of characters in the span that overlap a role in the gold-standard test.

Spatial Role	Precision	Recall	F-Measure
SPATIAL INDICATOR	0.967	0.889	0.926
TRAJECTOR	0.684	0.681	0.682
LANDMARK	0.741	0.835	0.785

Table 1: Task A results (first run)

The overall performances of the first run are very promising in terms of both precision and recall. In particular, the SPATIAL INDICATOR labeling achieves a significant F-Measure of 0.926 with a precision of 0.967. The sequence labeling approach provides good results for the LANDMARK and the TRAJECTOR roles too. Unfortunately, these results are not comparable with the performances obtained the last year edition of the SpRL task, where a grammatical head word-based measure has been applied.

The main difficulty in the SPATIAL INDICATOR classification concerns the tagging of a larger or smaller span for the roles, as for *"at the back"* that is tagged as *"at the back of"*. On the contrary, for roles like *"to the left and the right"* the system produces a tag covering just the first three words, *"to the left"*, because this shortest sequence was far more represented within the training set. Some roles corresponding to unknown word sequences, such as *"on the very right"*, were not labeled, leading to the little drop in terms of recall for the SPATIAL INDICATOR.

Another issue in the TRAJECTOR and LANDMARK labeling is due to the absence of specific role sequences in the training set, such as LANDMARK-TRAJECTOR-SPATIAL INDICATOR labeled in the test sentence *"there is a [coffee table]*LANDMARK *with a [sofa]*TRAJECTOR *[around]*SP.IND*"*: the SVM^{hmm} classifier in fact tends to discard any sequence unseen during training. Another issue concerns the difficulty in assigning the TRAJECTOR role to the proper SPATIAL INDICATOR: in the sentence *"a bench with a person lying on it"* where both *"a bench"* and *"a person"* are tagged as TRAJECTOR.

4.2 Results in Task B

Task B has been tackled using the SPTK-based Relation Identification approach, described in Section 3. In particular, the SVM classifier is fed with 741 positive examples, corresponding to the number of gold relations, while the negative examples generation process, described in Section 3, yielded 2,256 examples. The same Word Space described in the previous section has been used to compute the semantic similarity within the SPTK. For the tuning phase, a 80-20 fixed split has been applied.

For this task, two different measures are presented. The *Relaxed* measure considers a relation correct if each role composing it has at least one character overlapping the corresponding gold role. The *Strict* measure considers a relation correct only if each role in it has all the characters overlapping with the gold role. The first measure is more comparable with the one used in (Kordjamshidi et al., 2012a), where a relation is considered correct only if each grammatical head word of the involved roles were correctly labeled. The results achieved in this task by our system are reported in Table 2.

Spatial Role	Precision	Recall	F-Measure
RELAXED	0.551	0.391	0.458
STRICT	0.431	0.306	0.358

Table 2: Task B results

The problem for this task is more challenging. In fact, the overall task is strictly biased by the quality of the SVM^{hmm} based classifier and inherits all the limitations underlined in Section 4.2. This mostly affects the recall, because every error generated during the role classification is cumulative and losing only one role in Task A implies a misclassification of the whole relation. However, it is important to notice that these results have been achieved without any manual feature engineering nor any heuristics or hand coded lexical resource.

Spatial Role	Precision	Recall	F-Measure
SPATIAL INDICATOR	0.968	0.585	0.729
TRAJECTOR	0.682	0.493	0.572
LANDMARK	0.801	0.560	0.659

Table 3: Task A results (second run)

In the second run of Task A, we evaluate the contribution of this syntactic information to filter out roles. In Table 3 results of the second run for Task A are reported (see previous Section). As expected, the recall measure shows a performance drop with respect to results shown in Table 1: the results proposed in the first run represents an upperbound to the recall as any novel role is added here. However, the precision measure for the LANDMARK role classification is improved of about 10%.

Spatial Role	Precision	Recall	F-Measure
SPATIAL INDICATOR	0.609	0.479	0.536
MOTION INDICATOR	0.892	0.294	0.443
TRAJECTOR	0.565	0.317	0.406
LANDMARK	0.662	0.476	0.554
PATH	0.775	0.295	0.427
DIRECTION	0.312	0.229	0.264
DISTANCE	0.946	0.331	0.490

Table 4: Task C results

4.3 Results in Task C

In Task C the extended set of roles is considered. According to this, the number of possible labels to be learnt by the system increases, thus making the problem more challenging. As for Task A, here the SVM^{hmm} has been trained over the whole training set, using a 10-fold cross validation in the tuning phase. Moreover, the sentences of the Confluence corpus are far more complex than the ones from the CLEF corpus. Confluence sentences have a more narrative nature with respect to the CLEF sentences, that are simple description of images. The combination of these two factors resulted in a large drop in the performance, especially for the recall.

As shown by the results in Table 4, DIRECTION is the most difficult role to be classified, probably because it is represented by many different word sequences. Other roles are found in few instances, but almost all correct, as for DISTANCE and MOTION INDICATOR. The high value of *Precision* for the DISTANCE role is justified by the fact that when this role is composed by a number, (i.e. *"530 meters"*), the system identified and classified it well, while for a representation with only words (i.e. *"very close"*) the system did not retrieved it at all.

References

Y. Altun, I. Tsochantaridis, and T. Hofmann. 2003. Hidden Markov support vector machines. In *Proceedings of the International Conference on Machine Learning*.

Marco Baroni, Silvia Bernardini, Adriano Ferraresi, and Eros Zanchetta. 2009. The wacky wide web: a collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation*, 43(3):209–226.

Michael Collins and Nigel Duffy. 2001. Convolution kernels for natural language. In *Proceedings of Neural Information Processing Systems (NIPS'2001)*, pages 625–632.

Danilo Croce, Cristina Giannone, Paolo Annesi, and Roberto Basili. 2010. Towards open-domain semantic role labeling. In *ACL*, pages 237–246.

Danilo Croce, Alessandro Moschitti, and Roberto Basili. 2011. Structured lexical similarity via convolution kernels on dependency trees. In *Proceedings of EMNLP*, Edinburgh, Scotland, UK.

Danilo Croce, Giuseppe Castellucci, and Emanuele Bastianelli. 2012. Structured learning for semantic role labeling. *In Intelligenza Artificiale*, 6(2):163–176, January.

Oleksandr Kolomiyets, Parisa Kordjamshidi, Steven Bethard, and Marie-Francine Moens. 2013. Semeval-2013 task 2: Spatial role labeling. In *Proceedings of the 7th International Workshop on Semantic Evaluation*. Association for Computational Linguistics.

Parisa KordJamshidi, Martijn van Otterlo, and Marie-Francine Moens. 2010. Spatial role labeling: Task definition and annotation scheme. In *LREC*.

Parisa Kordjamshidi, Martijn Van Otterlo, and Marie-Francine Moens. 2011. Spatial role labeling: Towards extraction of spatial relations from natural language. *ACM Trans. Speech Lang. Process.*, 8(3):4:1–4:36, December.

Parisa Kordjamshidi, Steven Bethard, and Marie-Francine Moens. 2012a. Semeval-2012 task 3: Spatial role labeling. In *SemEval 2012*, pages 365–373, Montréal, Canada, 7-8 June. Association for Computational Linguistics.

Parisa Kordjamshidi, Paolo Frasconi, Martijn Van Otterlo, Marie-Francine Moens, and Luc De Raedt. 2012b. Relational learning for spatial relation extraction from natural language. In *Proceedings of the 21st international conference on Inductive Logic Programming*, ILP'11, pages 204–220, Berlin, Heidelberg. Springer-Verlag.

T. Landauer and S. Dumais. 1997. A solution to plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, 104(2):211–240.

Alessandro Moschitti. 2006. Efficient convolution kernels for dependency and constituent syntactic trees. In *ECML*, pages 318–329, Berlin, Germany, September.

Ted Pedersen, Siddharth Patwardhan, and Jason Michelizzi. 2004. WordNet::Similarity - Measuring the Relatedness of Concept. In *Proc. of 5th NAACL*, Boston, MA.

Kirk Roberts and Sanda Harabagiu. 2012. Utd-sprl: A joint approach to spatial role labeling. In *SemEval 2012)*, pages 419–424, Montréal, Canada, 7-8 June. Association for Computational Linguistics.

Magnus Sahlgren. 2006. *The Word-Space Model*. Ph.D. thesis, Stockholm University.

Hinrich Schutze. 1998. Automatic word sense discrimination. *Journal of Computational Linguistics*, 24:97–123.

John Shawe-Taylor and Nello Cristianini. 2004. *Kernel Methods for Pattern Analysis*. Cambridge University Press.

Ioannis Tsochantaridis, Thorsten Joachims, Thomas Hofmann, and Yasemin Altun. 2005. Large margin methods for structured and interdependent output variables. *J. Machine Learning Reserach.*, 6, December.