

Neural Models of Selectional Preferences for Implicit Semantic Role Labeling

Minh Le and Antske Fokkens

Computational Lexicology and Terminology Lab
Vrije Universiteit Amsterdam, the Netherlands
{m.n.le, antske.fokkens}@vu.nl

Abstract

Implicit Semantic Role Labeling is a challenging task: it requires high-level understanding of the text while annotated data is very limited. Due to the lack of training data, most researches either resort to simplistic machine learning methods or focus on automatically acquiring training data. In this paper, we explore the possibilities of using more complex and expressive machine learning models trained on a large amount of explicit roles. In addition, we compare the impact of one-way and multi-way selectional preference with the hypothesis that the added information in multi-way models are beneficial. Although our models surpass a baseline that uses prototypical vectors for SemEval-2010, we otherwise face mostly negative results. Selectional preference models perform lower than the baseline on ON5V, a dataset of five ambiguous and frequent verbs. They are also outperformed by the Naïve Bayes model of Feizabadi and Pado (2015) on both datasets. We conclude that, even though multi-way selectional preference improves results for predicting explicit semantic roles compared to one-way selectional preference, it harms performance for implicit roles. We release our source code, including the reimplementations of two previously unavailable systems to enable further experimentation.

Keywords: neural network, implicit semantic role labeling, selectional preferences

1. Introduction

Defined as the recovery of semantic roles beyond immediate syntactic structure, implicit Semantic Roles Labeling (iSRL) can contribute valuable information for obtaining complete semantic interpretations of text. Yet, it has been elusive since its first shared task eight years ago (Ruppenhofer et al., 2010).

The main difficulty faced by researchers is the small size of training data. Compared to traditional SRL datasets, SemEval-2010 is hundreds-fold smaller, containing only slightly more than a hundred of training examples (Table 1). Early work applying traditional semantic role labeling (SRL) techniques to iSRL was met with deflating results. Therefore, researchers limited themselves to simplistic machine learning models such as Naïve Bayes (Feizabadi and Pado, 2015, among others) or abandoned machine learning altogether (Laparra and Rigau, 2013). Several studies were devoted to the automatic expansion of training data (see Section 2 for an overview).

This paper presents an attempt to recover implicit semantic roles using neural networks. We take advantage of the fact that OntoNotes contains a vast amount of manually annotated explicit semantic roles from which we can learn the selectional preference of frames (e.g. *look.O1* prefers animate fillers for role *A0* (looker)). A neural network is used to capture complex interactions between a predicate, a target role and its co-occurring roles. In addition, we compare the impact of one-way selectional preference, taking only the selectional preference of the predicate for the target role into account, to multi-way selectional preference, which uses information from all semantic roles related to the predicate.

The contribution of this paper is twofold: First, we experimented with a class of simple neural models for iSRL and two types of selection preference. While the results are mostly negative, they highlight the importance of discourse

	SemEval		OntoNotes
	Train	Test	
Words	8K	9K	1,700K
Frames	344	371	7,007
Predicates	811	1,008	324,996
Predicates with DNI	102	118	0

Table 1: Statistics of an iSRL dataset (SemEval-2010, PropBank version) and a traditional SRL dataset (OntoNotes).

information (see Section 4.4 and 4.5) and suggest future directions that should (not) be taken.

The second contribution lies in addressing the challenges we met in carrying out this research and interpreting our results. The nature of these challenges lies in the fact that (1) all resources for implicit Semantic Role Labeling are small, (2) previous approaches differ in the dataset and the metrics they use for evaluation, and (3) to our knowledge, none of the existing systems is available as open source code. This has led to a situation that is typical for challenging tasks using small datasets: it is almost impossible to determine what the state-of-the-art approach is and how new work relates to this. Even results from papers that are evaluated on the same dataset are difficult to compare, because differences in results can be due to the difference in features, machine learning algorithm, method of extending data, heuristics or (as pointed out in Fokkens et al. (2013)) choices in preprocessing and data preparation. As part of this research, we built an experimental platform for iSRL. This platform provides open source implementations for the experiments reported in this paper, for the system described in Schenk and Chiarcos (2016) which inspired our own approach and for Feizabadi and Pado (2015)'s approach which provided state-of-the-art performance on SemEval-2010.

The rest of the paper is organized as follows: Section 2 summarizes the foundation of iSRL and related work. In Section 3, we outline our models of selectional preference. Section 4 quantifies the effectiveness of selectional preference with regard to iSRL. Section 5 concludes the work and outlines future directions of research.

2. Background and Related Work

In this section, we explain what implicit Semantic Role Labeling entails. This is followed by an overview of previous work on this task. Next, we address related work that uses selectional preferences. Consider the following sentence from SemEval-2010 training set:

- (1) Apparently [the tenants]_{A0} had [brought]_{bring.01} [little or nothing]_{A1} with them, and all the furniture down to the smallest details had been taken over with [the house]_{A2}.

The roles *A0* and *A1* of the predicate *bring.01* can be filled with phrases in the immediate syntactic structure while the filler of *A2* falls into a separate clause. Typically, a SRL system would annotate the fillers for *A0* and *A1* and ignore *A2*. It is therefore called a Null Instantiation (NI).

Null-instantiations can be indefinite (INI) and definite (DNI). To reuse examples from Ruppenhofer et al. (2010), in the blog headline *More babbling about what it means to know*, the subject of knowing is not expected to be instantiated within the discourse. In contrast, in the sentence *Don't tell me you didn't know!*, the hearer expects a concrete filler for the role of what (s)he should know and it can be expected to be present in previous context. The first example is a case of INI while the second is a DNI.

2.1. Previous work on iSRL

Traditional SRL techniques led to very low results for iSRL due to data sparseness (Chen et al., 2010; Tonelli and Delmonte, 2010). Researchers therefore explored simpler alternatives such as BayesNet (Silberer and Frank, 2012; Roth and Frank, 2013; Roth and Frank, 2015), Naïve Bayes (Feizabadi and Pado, 2015), and memory-based learning (Schenk et al., 2015). Others proposed non-parametric approaches such as observed frequency (Laparra and Rigau, 2012), prototypical vectors (Schenk and Chiarcos, 2016) and other heuristics (Laparra and Rigau, 2013; Gorinski et al., 2013).

In addition to methods of machine learning and heuristics, previous work investigated the possibilities of increasing training data. Feizabadi and Pado (2015) combine multiple corpora and apply domain adaptation methods to deal with the difference in genre. They demonstrated that combining two iSRL corpora led to improved performance. Silberer and Frank (2012) and Roth and Frank (2015) used heuristics to generate iSRL training examples from manually and automatically annotated SRL corpora. This work differs from these approaches, because their research focused on creating iSRL training examples of reasonable quality rather than using a SRL resource directly.

2.2. Selectional preferences

Selectional preference has a long research tradition (Katz and Fodor, 1963) and has been applied in various tasks such as syntactic parsing (Zhou et al., 2011), textual inference (Ritter et al., 2010), and semantic role labeling (Zapirain et al., 2013). The idea is simple: a role is filled with some words more frequently than others. For example, *the man* is much more likely a filler for the role *A0* (leader) of the predicate *lead.01* than e.g. *the bottle* (an inanimate object) although one can construct a grammatically and semantically correct example for each filler.

Next to the role's semantics, co-occurring roles also have an influence. For example, if *lead.01*'s role *A4* (goal) is filled by *the guest house*, *the nation* is an implausible filler for *A1* (thing led), while it is perfectly plausible had we not known what fills *A4*. This is known as *multi-way selectional preference* (van de Cruys, 2014).

One-way selectional preferences have been applied to implicit semantic role labeling before. Silberer and Frank (2012)'s system include a feature calculated using weighted similarity to head words that are observed to fill a role. The selectional preference model itself is described in (Erk, 2007) and (Resnik, 1996). A simpler model that uses unweighted similarity is used by Schenk and Chiarcos (2016). Our results show that adding multi-way selectional preference improves results on explicit semantic roles, but not for iSRL. Recently work by Do et al. (2017) is closest to our work but they apply their methods on nominal data and did not compare one-way and multi-way selectional preference.

2.3. Neural networks

Recent years have witnessed a surge of research interest in neural networks for natural language processing (Goldberg, 2016). Plenty of models have been proposed for various tasks (Godbole et al., 2015; Zhou and Xu, 2015; Andor et al., 2016, among many others). Apart from Do et al. (2017) who uses a different architecture for a different version of the task, we are not aware of work that applies neural networks to iSRL.

3. Models

In this study, we focus exclusively on *DNI resolution*, the last and hardest step in iSRL. For each test case, we assume that the predicate *p* is already identified and disambiguated, the target role *r** is given, and overt roles are coupled with their fillers $\{(r_j, g_j) | j = 1..m\}$. The goal is to rank the correct filler highest among the candidates $\{c_i | i = 1..n\}$. To test a simple multi-way selectional preference model, we use the following formula to assign a score for each candidate:

$$s(c_i) = \bigoplus_{j=1}^m f(p, r^*, r_j, g_j, c_i) \quad (1)$$

where *m* is the number of explicit roles known to the system, \bigoplus is a aggregation function (e.g. *sum* or *max*). In the case of *one-way* selectional preference, the formula degenerates into:

$$s'(c_i) = f'(p, r^*, c_i) \quad (2)$$

f and f' are neural networks that have the same architecture, except the number of inputs. The precise form of the neural networks and the aggregation function is determined via a hyperparameter search (see Section 4.3).

Role fillers $\{g_j\}$ and candidates $\{c_i\}$ can be transformed into features by extracting the head word, but other features can also be used. Together with predicate and role names, they are embedded into a vector space. The embedding matrix is trainable and can be initialized with pretrained word vectors for better performance.

Compared to a model that computes prototypical vectors as the average of observed vectors in the fashion of Schenk and Chiarcos (2016), our neural models have (at least) two advantages:

- Distributed representation is used to represent predicates and roles, not only fillers, allowing the model to work in cases of unseen predicates or predicate-role combinations.
- The representation enables the sharing of statistical strength between predicates, i.e. rare predicates can get more accurate predictions by means of resemblance to frequent predicates.

An additional motivation is that multi-way preference can offer a solution to the context-dependent nature of semantic role labeling. Our current results, however, do not provide sufficient evidence to support such a claim.

4. Experiments

We evaluate our models for DNI resolution by comparing our model to a baseline and to Feizabadi and Pado (2015). In addition, we perform an ablation analysis to find out which components of the model are useful.

4.1. Data

We train our selectional preference model on OntoNotes (Weischedel et al., 2013), a balanced 1.7M words corpus with over 320K manually annotated predicates and their explicit arguments.

SemEval-2010 (Ruppenhofer et al., 2010) is a standard dataset to evaluate iSRL systems. It contains chapters of Sherlock Holmes, one for training and two for testing, annotated with both implicit and explicit semantic roles. The organizers provide two versions of the same dataset: one annotated with FrameNet roles and the other PropBank. Because OntoNotes was compiled using PropBank, we also use the PropBank version of SemEval-2010.

Note that OntoNotes differs from SemEval-2010 in task (explicit versus implicit SRL), genres (news, weblogs and conversations versus novel) and time period (20th century versus 19th century). Training on OntoNotes SRL and testing on SemEval-2010 iSRL can be seen as a form of domain adaptation and requires powerful generalization.

We also test our models on ON5V (Moor et al., 2013) which poses a different challenge. Implicit semantic roles were manually annotated on top of explicit semantic roles and other linguistic information on a selection of

Predicate	Role	Filler head	Filler full
pay.01	A2	refugees	the refugees
pay.01	A2	U.S.	U.S.
pay.01	A2	families	the victims' families
pay.01	A2	trust	this trust
pay.01	A2	Warner	AOL Time Warner
pay.01	A2	they	they
pay.01	A2	lenders	lenders
pay.01	A2	lawyers	lawyers
pay.01	A2	one	one

Table 2: Examples from ON5V showing the diversity of fillers in terms of semantic types, part-of-speech, and topics.

OntoNotes documents. The authors chose five highly frequent verbs to annotate in order to create “high-volume of annotations for specific verb predicates”.¹ As a result, the words and phrases that fill each role are very diverse, as illustrated by the examples in Table 2. To achieve high performance on this dataset, a model needs to be selective yet general enough to encompass different types of fillers.

4.2. Baseline

Our baseline is inspired by Schenk and Chiarcos (2016). For every <predicate, role> pair found in our training set, it computes a prototypical vector and, at test time, returns the candidate that is closest to the prototypical vector. Following their best model, we use the pretrained embeddings from Collobert et al. (2011).

Due to some differences between research questions and experimental setup, the results cannot be compared to Schenk and Chiarcos’s algorithm directly. Firstly, for a fair comparison with selectional preference-based models, we use only the head word of each candidate (whereas they average all words in the phrase). Secondly, we train and evaluate on PropBank-style datasets while they use FrameNet-style data.

4.3. Experimental Setup

We use the baseline described in the previous section and a Naïve Bayes model trained on SemEval-2010 data (Feizabadi and Pado, 2015) to compare to our model’s performance. To quantify the effect of different aspects of the model, we investigate the following variants:

ONEWAY captures one-way selectional preference and represents fillers by their syntactic head.

MULTIWAY captures multi-way selectional preference and represents fillers by their syntactic head.

SYNSEM uses richer features for fillers rather than selectional preferences. We use five syntactic and semantic features from Feizabadi and Pado (2015), namely, *Expected roles*, *Semantic Type*, *Word Frequency*, *POS*, and *Constituent type*.²

¹They are: *pay*, *give*, *bring*, *leave*, *put*.

²See Table 2 in their paper. We did not use their discourse features because they require iSRL annotations which is not available in OntoNotes.

	Train (%)	Validation (%)
ONEWAY	59.52	46.61±0.23
MULTIWAY	58.52	47.64±0.19

Table 3: Accuracy of selectional preference models on OntoNotes (for validation set we report mean and standard deviation over 5 runs).

SYNSEM+ONEWAY combines richer features with one-way selectional preference.

SYNSEM+MULTIWAY takes into account co-occurring roles to capture multi-way selectional preference.

We construct one training example for each argument found in OntoNotes and split the data into 90% for training and 10% for development. Models are trained to choose the right filler for each target role with as input: the predicate, the role and, if applicable, other explicit arguments. We evaluate on the NI-only test set from SemEval-2010 using the standard evaluation script (Ruppenhofer et al., 2010).

We initiated the embedding matrix with 27K vectors from the pretrained embeddings of Collobert et al. (2011). We use AdaGrad (Duchi et al., 2011) for optimization; the initial learning rate was customized for each model to avoid gradient explosion. All models were trained until no improvement was observed on the development set (but not more than 1,000 epochs, for practical reasons). To account for random initialization in neural networks, we run each model 15 times and average the results. An arbitrary but fixed random seed was used for each run to ensure reproducibility.

All hyperparameters were tuned on OntoNotes development set. We tested sum and max for aggregation function; sigmoid, tanh, and cube for activation function (Chen and Manning, 2014); different strength of dropout (Hinton et al., 2012), regularization, and learning rate. Because of limited computational resource, we performed a random hyperparameter search to find the best setting. As discussed in Section 3, fillers can be represented in different ways. We observed that using both the head word and the closest coreferent non-pronoun head word is better than using the head word only on our development set. Notice that gold coreference chains are assumed to be available at test time and were used in previous work (Silberer and Frank, 2012) as well as the system we compare to (Feizabadi and Pado, 2015).

The source code of all experiments, including random seeds and replication instructions, is publicly available at: <https://bitbucket.org/ctrl/isrl-sp>.

4.4. Results on SemEval-2010

Table 3 shows the performance of selectional preference models with regard to resolving the explicit roles of OntoNotes. Selectional preference alone (without the help of syntactic structures) can find the correct filler in more than 47% of the cases. We observe a small but statistically significant ($p < 0.05$) improvement on the validation set by adding multiway selectional preference.

The results in Table 4 show that our models significantly increase the F_1 score above the baseline on SemEval-2010

	P	R	F_1
Baseline	26.85	21.80	24.07
F&P	35.04	30.83	32.80
ONEWAY	28.80	28.67	28.74 ± 1.63
MULTIWAY	27.02	26.82	26.92 ± 1.42
SYNSEM	16.63	16.54	16.58 ± 1.51
ONEWAY+SYNSEM	24.05	23.91	23.98 ± 5.05
MULTIWAY+SYNSEM	17.29	17.29	17.29 ± 0.00

Table 4: Results on SemEval-2010. Results of neural models are averaged over 15 runs. F_1 scores are reported with mean and standard deviation when possible.

	P	R	F_1
Baseline	13.00	13.00	13.00
F&P	16.72	15.19	15.90
ONEWAY	10.64	10.64	10.64 ± 1.44
MULTIWAY	9.14	9.14	9.14 ± 1.46
SYNSEM	5.92	5.92	5.92 ± 2.10
ONEWAY+SYNSEM	10.37	10.37	10.37 ± 5.50
MULTIWAY+SYNSEM	1.24	1.24	1.24 ± 0.00

Table 5: Results on ON5V

dataset. Both neural models show significant improvement in precision and an even bigger improvement in recall. This can be attributed to their ability to generalize to unseen predicate-role combinations and abstract away from observed ones in their hidden layer. Contrary to our expectation, MULTIWAY is inferior to ONEWAY ($p < 0.05$).

ONEWAY and MULTIWAY do not outperform F&P which is simpler in terms of machine learning architecture, but is trained on in-domain, iSRL data, and uses more features. To bridge the gap between the models, we also integrate Feizabadi and Pado’s syntactic and semantic features into our neural models but they do not lead to improved performance.

Table 3 and Table 4 reveal an increase in random fluctuation when moving from OntoNotes to SemEval-2010, probably because of a difference of some orders of magnitude in size. Moreover, SYNSEM+ONEWAY gets an F_1 -score of 16.54% for one of its runs (lower than the mean of all other neural models) and 29.43% for another (higher than all means). These observations stress the importance of running experiments multiple times when random factors (such as parameter initialization and the order of training examples) are involved. Based on a single run, a model might be heavily over- or underrated.

4.5. Results on ON5V

In Table 5, we report the results of models on ON5V (Moor et al., 2013). Again, the Naïve Bayes model using both local and discourse information proposed by Feizabadi and Pado (2015) clearly provides the best performance, whereas neural models do not show improvement over the baseline ($p < 0.05$).

The disappointing performance points to its inherent limitation: it expects one prototypical filler per (predicate, role) pair. As shown in Table 2, this assumption breaks in ON5V, resulting in a lower mean and higher variance.

We expected that MULTIWAY would alleviate this problem by varying the predicted vector based on surrounding roles. While it achieves that for explicit SRL on OntoNotes (Table 3), the result does not carry over to ON5V. Local syntactic and semantic information do not improve results for SemEval. This applies even more strongly to ON5V. SYNSEM leads to very low results when standing alone and does not improve performance when combined with ONEWAY or MULTIWAY (the difference between ONEWAY and its combination with SYNSEM is not statistically significant). In comparison with F&P, this result emphasizes the importance of discourse information in the task.

5. Conclusions

In this paper, we investigated the use of more expressive machine learning models for implicit Semantic Role Labeling. We proposed novel neural models that use selectional preference and applied them to iSRL. Our empirical results show that neural models are only better than a lookup table of prototypical vectors in a natural setting such as SemEval-2010 while underperforming for highly frequent and ambiguous words in ON5V. Furthermore, the added expressive power does not help neural models to overcome a simpler model trained on in-domain data and equipped with discourse features (though it should be noted that we tested only a small family of simple architectures, cf. Do et al. (2017)). Multi-way preference is found to be helpful in the case of (explicit) semantic role labeling but not for iSRL. Although the results are mostly negative, our research provides hard-earned insights into this challenging task which we believe will be useful for researchers.

We release all of our models and the implementation of Schenk and Chiarcos (2016) and Feizabadi and Pado (2015)'s models as open-source software. We also report the fluctuation of results which stresses the importance of measuring a model multiple times when stochastic factors are involved.

Overall, this paper provides a solid basis for further research. Our observations on fluctuation and significance suggest more evaluation data may be needed to identify the true impact of specific models and features.

6. Acknowledgements

The work presented in this paper was funded by the Netherlands Organization for Scientific Research (NWO) via the Spinoza grant, awarded to Piek Vossen and via VENI grant 275-89-029 awarded to Antske Fokkens. Experiments were carried out on the Dutch national e-infrastructure with the support of SURF Cooperative.

7. Bibliographical References

Andor, D., Alberti, C., Weiss, D., Severyn, A., Presta, A., Ganchev, K., Petrov, S., and Collins, M. (2016). Globally Normalized Transition-Based Neural Networks. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, volume 1, pages 2442–2452, Berlin, Germany. Association for Computational Linguistics.

Chen, D. and Manning, C. (2014). A Fast and Accurate Dependency Parser using Neural Networks. In *EMNLP 2014*, pages 740–750. ACL.

Chen, D., Schneider, N., Das, D., and Smith, N. A. (2010). SEMAFOR: Frame argument resolution with log-linear models. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pages 264–267, Uppsala, Sweden. ACL.

Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., and Kuksa, P. (2011). Natural language processing (almost) from scratch. *Journal of Machine Learning Research*, 12:2493–2537.

Do, Q. N. T., Bethard, S., and Moens, M.-F. (2017). Improving Implicit Semantic Role Labeling by Predicting Semantic Frame Arguments. *arXiv:1704.02709*.

Duchi, J., Hazan, E., and Singer, Y. (2011). Adaptive Sub-gradient Methods for Online Learning and Stochastic Optimization. *Journal of Machine Learning Research*, 12:2121–2159.

Erk, K. (2007). A Simple, Similarity-based Model for Selectional Preferences. In *ACL 2007*, pages 216–223, Prague, Czech Republic.

Feizabadi, P. S. and Pado, S. (2015). Combining Seemingly Incompatible Corpora for Implicit Semantic Role Labeling. In **SEM 2015*, pages 40–50.

Fokkens, A., Erp, M., Postma, M., Pedersen, T., Vossen, P., and Freire, N. (2013). Offspring from Reproduction Problems: What Replication Failure Teaches Us. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1691–1701. Association for Computational Linguistics.

Godbole, V., Liu, W., and Togneri, R. (2015). An Investigation of Neural Embeddings for Coreference Resolution. *Computational Linguistics and Intelligent Text Processing*, 9041:241–251.

Goldberg, Y. (2016). A Primer on Neural Network Models for Natural Language Processing. *Journal of Artificial Intelligence Research*, 57:345–420.

Gorinski, P., Ruppenhofer, J., and Sporleder, C. (2013). Towards Weakly Supervised Resolution of Null Instantiations. *Proceedings of IWCS*, pages 119–130.

Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*, pages 1–18.

Katz, J. J. and Fodor, J. A. (1963). The structure of a semantic theory. *Language*, 39(2):170–210.

Laparra, E. and Rigau, G. (2012). Exploiting Explicit Annotations and Semantic Types for Implicit Argument Resolution. In *Semantic Computing (ICSC), 2012 IEEE Sixth International Conference on*, pages 75–78, 9.

Laparra, E. and Rigau, G. (2013). ImpAr: A Deterministic Algorithm for Implicit Semantic Role Labelling. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1180–1189.

Moor, T., Roth, M., and Frank, A. (2013). Predicate-specific Annotations for Implicit Role Binding: Corpus

- Annotation, Data Analysis and Evaluation Experiments. In *Proceedings of the 10th International Conference on Computational Semantics*, pages 369–375.
- Resnik, P. (1996). Selectional constraints: an information-theoretic model and its computational realization. *Cognition*, 61(508):127–159.
- Ritter, A., Etzioni, M., and Etzioni, O. (2010). A Latent Dirichlet Allocation method for Selectional Preferences. *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 423–434.
- Roth, M. and Frank, A. (2013). Automatically Identifying Implicit Arguments to Improve Argument Linking and Coherence Modeling. *Proceedings of the Second Joint Conference on Lexical and Computational Semantics (*SEM)*, 1:306–316.
- Roth, M. and Frank, A. (2015). Inducing Implicit Arguments from Comparable Texts: A Framework and its Applications. *Computational Linguistics*, 41(4):625–664.
- Ruppenhofer, J., Sporleder, C., Morante, R., Baker, C., and Palmer, M. (2010). SemEval-2010 Task 10: Linking Events and Their Participants in Discourse. In *Proceedings of the 5th International Workshop on Semantic Evaluation, ACL 2010*, pages 45–50, Uppsala, Sweden.
- Schenk, N. and Chiarcos, C. (2016). Unsupervised Learning of Prototypical Fillers for Implicit Semantic Role Labeling. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1473–1479, San Diego, California, 6. Association for Computational Linguistics.
- Schenk, N., Chiarcos, C., and Sukhareva, M. (2015). Towards the Unsupervised Acquisition of Implicit Semantic Roles. In *RANLP 2015*, pages 570–578.
- Silberer, C. and Frank, A. (2012). Casting Implicit Role Linking As an Anaphora Resolution Task. In *Proceedings of 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 '12*, pages 1–10, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Tonelli, S. and Delmonte, R. (2010). VENSES++: Adapting a deep semantic processing system to the identification of null instantiations. In *Proceedings of the 5th International Workshop on Semantic Evaluations, ACL 2010*, number July, pages 296–299, Uppsala, Sweden. Association for Computational Linguistics.
- van de Cruys, T. (2014). A Neural Network Approach to Selectional Preference Acquisition. *EMNLP 2014*, pages 26–35.
- Zapirain, B., Agirre, E., Màrquez, L., and Surdeanu, M. (2013). Selectional Preferences for Semantic Role Classification. *Computational Linguistics*, 39(3).
- Zhou, J. and Xu, W. (2015). End-to-end learning of semantic role labeling using recurrent neural networks. *ACL 2015*, pages 1127–1137.
- Zhou, G., Zhao, J., Liu, K., and Cai, L. (2011). Exploiting Web-Derived Selectional Preference to Improve Statistical Dependency Parsing. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 1556–1565.

8. Language Resource References

- Moor, Tatjana and Roth, Michael and Frank, Anette. (2013). *The OntoNotes 5 Verbs Non-local Role Linking Data Set (ON5V)*. Heidelberg Universität.
- Ruppenhofer, Joseph and Sporleder, Caroline and Morante, Roser and Baker, Colin and Palmer, Martha. (2010). *SemEval-2010 Task 10: Linking Events and Their Participants in Discourse*. Association of Computational Linguistics.
- Weischedel, Ralph and Palmer, Martha and Marcus, Mitchell and Hovy, Eduard and Pradhan, Sameer and Ramshaw, Lance and Xue, Nianwen and Taylor, Ann and Kaufman, Jeff and Franchini, Michelle and El-Bachouti, Mohammed and Belvin, Robert and Houston, Ann. (2013). *OntoNotes Release 5.0 LDC2013T19*. Linguistic Data Consortium, ISLRN 151-738-649-048-2.