# Cross-lingual Transfer Learning for Semantic Role Labeling in Russian

**Ilseyar Alimova**
Kazan Federal University
Kazan, Russia
alimovaIlseyar@gmail.com

**Elena Tutubalina**
Kazan Federal University
Kazan, Russia
elvtutubalina@kpfu.ru

**Alexander Kirillovich**
Kazan Federal University
Kazan, Russia
alik.kirillovich@gmail.com

## Abstract

This work is devoted to semantic role labeling (SRL) task in Russian. We investigate the role of transfer learning strategies between English FrameNet and Russian FrameBank corpora. We perform experiments with embeddings obtained from various types of multilingual language models, including BERT, XLM-R, MUSE, and LASER. For evaluation, we use a Russian FrameBank dataset. As source data for transfer learning, we experimented with the full version of FrameNet and the reduced dataset with a smaller number of semantic roles identical to FrameBank. Evaluation results demonstrate that BERT embeddings show the best transfer capabilities. The model with pretraining on the reduced English SRL data and fine-tuning on the Russian SRL data show macro-averaged F1-measure of 79.8%, which is above our baseline of 78.4%.

**Keywords**: Semantic Role Labeling, Transfer learning, Word embeddings, Deep Learning, FrameNet, FrameBank

## 1. Introduction

Semantic Role Labeling (SRL) is one of the most critical tasks in natural language processing (Palmer et al., 2010; Solovyev and Ivanov, 2016). The SRL aims to identify the situation a given sentence describes, find sentence constituents expressing the participants of this situation, and identify the roles the participants play.

Recent advances in multilingual neural network models offer new opportunities to improve SRL (Arkhipov et al., 2019; Okamura et al., 2018; Subburathinam et al., 2019). In this work, we take the task a step further from existing monolingual research (Shelmanov and Devyatkin, 2017; Larionov et al., 2019) by exploring multilingual transfer between semantic roles labeled datasets in different languages. Our goal is not to outperform state of the art models, but to ask whether we can transfer knowledge from a high-resource language, such as English, to a low-resource one, e.g., Russian, for SRL. In this work, we seek to answer the following research questions: **RQ1:** Will transfer learning (TL) help improve the results? **RQ2:** How the quality change if the roles in the training and target corpora will be the same? **RQ3:** Which multilingual pre-trained models are most effective for an SRL task?

We conducted experiments on two datasets: a database of Russian lexical constructions FrameBank and an English large-scale semantic database FrameNet. We consider four modern multilingual language models: BERT (Devlin et al., 2019), XLM-R (Conneau and Lample, 2019), MUSE (Lample et al., 2017), LASER (Artetxe and Schwenk, 2019). To our knowledge, this is the first work exploring the interlingual transfer ability for SRL in Russian.

## 2. Related Work

Various approaches have been proposed for the SRL task in English. Gildea and Jurafsky proposed the statistical classifiers with various lexical and syntactic features combined with knowledge of the predicate verb, noun, or adjective and the prior probabilities of multiple combinations of semantic roles

(Gildea and Jurafsky, 2002). The classifier was tested on the FrameNet corpus. The developed system performed 82% accuracy in identifying the semantic role of pre-segmented constituents and 63% of F-measure on simultaneously segmenting constituents and identifying their semantic role task. Pradhan et al. proposed the SRL system based on the Support Vector Machine classifier (Pradhan et al., 2005). The authors applied a new set of features, including dependency parses features extracted with a combination of Minipar syntactic parse, a chunked syntactic representation, and Charniak parses. The model with a single Charniak parser performed 83.7% of F-measure. A combination of syntactic parsers improved the results on 1,5% of F-measure. Collobert et al. proposed a simple multi-layer neural network that takes as an input the words decoded into a feature vector, by a lookup table operation (Collobert et al., 2011). However, their best system fell short of previous feature-based systems. The modern works apply complicated neural network architectures. He et al. introduced deep highway BiLSTM architecture with constrained decoding (He et al., 2017). The network achieved 83.2% of F-measure on the CoNLL 2005 test set and 83.4 of F-measure on CoNLL 2012 datasets.

The development of the Fremebank corpus led to the growth of studies devoted to SRL for the Russian language. Kuznetsov (2013) proposed a baseline system for SRL in Russian . The system consists of the following parts: text preprocessing module (morphological analysis, lemmatization, syntactic analysis), data enrichment module (mapping text segments annotation to syntax tree nodes), training module (feature extraction, classification, optimization). The system with verbs form, predicate lemmas, and syntactic features obtained 76.1% of F-measure. Adding a combination of semantic and syntactic features increased the results to 76.4% of F-measure. Shelmanov and Devyatkin (2017) applied two neural networks for SRL on the FrameBank corpus. The first neural network model has the simple architecture that acquires all features of an argument: sparse and dense, as a single vector and propagates them through three dense layers. The second complex neural network has the same types of layers. However, the first layer is split into several chunks: a chunk for categorical features, a chunk for an argument embedding, and a chunk for a predicate embedding. The categorical features include various morphological, the relative position of an argument in a sentence, predicate lemma, the preposition of an argument, and the name of a syntax link from argument to its parent features. The authors investigated the ability to learn a model for labeling arguments of “known” and “unknown” predicates that are present and not present in a training set, respectively. The complex neural network achieved 82.3% of micro F-score on “known” predicates and 66.7% on “unknown” predicates and outperformed the simple network on 6.2% and 34.8%, respectively. Larionov et al. (2019) evaluated various pretrained language models, including, word2vec, fasttext, ELMo, BERT, RuBERT. For “known” predicates, the ELMo-based model performed the highest micro F-measure (83.42%), and the RuBERT model outperformed other models in terms of macro F-measure (80.12%). For ‘unknown” predicates, ELMo performed the highest metrics both for macro and micro F-measures (37.64% and 55.50%, respectively). A recent study applied a frame-based approach for predicting sentiment attibutes towards named entities in political news (Rusnachenko et al., 2019; Loukachevitch and Rusnachenko, 2020).

To sum up, machine learning approaches with contextual embeddings have a high potential for the SRL task. More recently, multilingual embeddings have been used to achieve state-of-the-art performance on many NLP tasks such as named entity recognition and classification (Devlin et al., 2019; Conneau and Lample, 2019; Artetxe and Schwenk, 2019; Miftahutdinov et al., 2020). The goal of this study is to investigate cross-lingual transfer methods for SRL that exploit resources from existing high-resource language, i.e. English, and fine-tuning on Russian data.

## 3. Datasets

In this section, we describe two datasets for SRL in Russian and in English. Transfer learning aims to solve the problem on a “target” dataset using knowledge learned from a “source” dataset. We use the English FrameNet dataset (Baker et al., 1998) as source data and the Russian FrameBank dataset (Lyashevskaya and Kashkin, 2015; Lyashevskaya, 2012) as target data. We study two setups for the source side: (i) FULL data and (ii) REDUCED data setup that we describe in Section 3.3.

### 3.1. FrameBank

FrameBank[1] (Lyashevskaya and Kashkin, 2015; Lyashevskaya, 2012) is a database that consists of a dictionary of Russian lexical constructions and an annotated corpus of their realizations in contemporary written texts. In the dictionary, each verb or other predicate word is followed by a list of constructions in which it serves as a target word. Construction is a morphosyntactic template, where some elements are fixed lexical units, and some are variable slots. A typical construction describes the argument structure of a verb. It consists of the one fixed element, representing the verb and one or more variable slots, representing arguments of this verb. Less frequent constructions are ones describing the argument structure of other parts of speech (POS) and complex idiomatic phrases.

Description of construction elements includes:

- the syntactic rank (Subject, Object, Predicate, Peripheral, Clause);
- the morphosyntactic features (POS, case, and preposition marking);
- the semantic role (Agent, Patient, Instrument, Theme, etc.);
- the lexical-semantic class (person, animal, building, abstract entity, etc.).

The annotated corpus consists of construction realization examples in the Russian National Corpus. Each example is linked to the construction it instantiates, and the parts of the example sentence are linked to the construction slots. These parts are annotated by their actual syntactic and morphological features, which can be different from the features, prescribed by the corresponding construction.

The publicly available version of FrameBank contains 16123 constructions for 1589 target words, realized by 52737 annotation sets.

### 3.2. FrameNet

FrameNet[2] (Baker et al., 1998) is a large-scale semantic resource, organized as a network of frames. A frame is a description of an abstract situation and its participants, called frame elements. For example, the frame elements of the *Commerce_buy* frame are *Buyer*, *Goods*, *Seller*, etc. Frames are interlinked by several relation types, including inheritance, perspective on, subframes, etc. For example, the *Commerce_buy* frame and its frame elements inherit from the *Getting* frame, and its *Recipient*, *Theme*, *Source* frame elements, respectively. A frame is associated with lexical units, i.e., disambiguated words, evoking this frame. For example, the *Commerce_buy* frame is evoked by “buy” and “purchase” lexical units.

In FrameNet, the network of frames is complemented by the corpus of annotated sentences. In each sentence, one word (typically, a verb) is a lexical unit, evoking a frame, and other sentence constituents express elements of this frame. For example, in the “John bought a car from Mary” sentence, “bought” is a lexical unit, evoking the *Commerce_buy* frame, while “John”, “a car” and “Mary” express *Buyer*, *Goods*, *Seller* frame elements, respectively.

The currents version of FrameNet contains 1224 frames, evoked by 13676 lexical units and 202970 annotation sets.

### 3.3. Linking of FrameNet roles to FrameBank

Concerning the SRL task, there are several significant differences between FrameNet and FrameBank. First, in FrameNet, the frames are defined as generalized language-independent situations, while any FrameBank construction is defined for a particular target word. Second, FrameNet frame elements are defined locally for each particular frame (for example, the *Commercial_transaction* frame defines the roles of the buyer, seller, good, etc., and the *Theft* frame defines the roles of perpetrator, victim, good, etc.). In contrast, FrameBank roles are defined globally for all constructions (for example, the construction for the word *kupit* ‘to buy’ and for the word *ukrast* ‘to steal’ use the roles from the same globally defined pool: agent, patient, theme, etc.).

[1] https://github.com/olesar/FrameBank
[2] https://framenet.icsi.berkeley.edu/

To mitigate these differences in the FrameNet, we study two setups for the source side. First, for the FULL setup, we use the entire FrameNet data as a starting point to train a neural model for SRL. We used lexical units as predicates, words annotated in corpus with frame element as arguments, and frame element's name as roles. Second, for the REDUCED setup, we left examples with roles present in the FrameBank corpus. To identify matching roles, we took the translation of roles provided in the FrameBank corpus. The final REDUCED FrameNet corpus includes a total of 86951 examples and 19 roles. For each annotated corpus, we created all possible pairs of predicate and argument and obtained 505 940 samples.

## 4. Experiments and Evaluation

In this section, we describe the model architecture, pretrained language models, and results of experiments.

### 4.1. Model

We implemented the neural network proposed in (Larionov et al., 2019). The network contains three input layers for the embedding of an argument, the embedding of a predicate and feature embeddings, and sparse categorical features. The input data fed separately to dense and batch normalization layers. Concatenated outputs of the first layer are fed to dense, batch normalization and dropout layers. The last output dense layer with a softmax activation function makes a classification.

The model takes as an input following features:

- Various morphological characteristics of both an argument and a predicate (case, valency, verb form, etc.);
- Relative position of an argument in a sentence concerning a predicate.
- Preposition of an argument extracted from a syntax tree (including a complex preposition as a single string);
- Name of a syntactic link that connects an argument token to its parent in the syntax tree;
- Argument and predicate lemmas.

We used a maximum of 50 and 15 epochs to train the model on FrameNet and FrameBank, respectively. For both corpora, we utilized the batch size of 32 and Adam optimizer. We applied the implementation of the model from this repository[3].

### 4.2. Pretrained Language Models

We consider four modern multilingual language models: BERT (Devlin et al., 2019), XLM-R (Conneau and Lample, 2019), MUSE (Lample et al., 2017), LASER (Artetxe and Schwenk, 2019). For arguments and predicates consisting of several words, we used averaged vectors. Further, we provide a detailed description of each model.

**BERT** (Bidirectional Encoder Representations from Transformers) is a recent neural network model for NLP presented by Google (Devlin et al., 2019). BERT is based on bidirectional attention-based Transformer architecture (Vaswani et al., 2017). In particular, we applied BERT$_{base}$, Multilingual Cased (Multi-BERT), which is pretrained on 104 languages and has 12 heads, 12 layers, 768 hidden units per layer, and a total of 110M parameters. For each predicate and argument, we use the BERT output layer to obtain embeddings without using context in sentences. Besides, we obtained contextualized vectors, when the whole sentence was fed to the input of the network (**BERT-context**).

**XLM-R** improves the multilingual BERT model by incorporating a cross-lingual task of translation language modeling, which performs masked language modeling on a concatenation of parallel bilingual sentence pairs (Ruder et al., 2019). The model is also based on Transformer architecture (Vaswani et al.,

[3]https://github.com/IINemo/isanlp_srl_framebank

2017). We applied the XLM-R Masked Language Model, which is pretrained on 2.5 TB of CommonCrawl data, in 100 languages, with 8 heads, 6 layers, 1024 hidden units per layer.

**MUSE** (Multilingual Unsupervised and Supervised Embeddings) is a sentence encoding model simultaneously trained on multiple tasks and multiple languages able to create a single embedding space to 30 languages (Lample et al., 2017). The vectors obtained with fastText library (Bojanowski et al., 2017) pretrained on texts from Wikipedia. The length of the obtained vectors is 300.

**LASER** (Language-Agnostic SEntence Representations) is a library to calculate and use multilingual sentence embeddings (Artetxe and Schwenk, 2019). LASER is based on encoder-decoder architecture proposed in (Schwenk, 2018). The model was trained on Wikipedia texts and Billions of High-Quality Parallel Sentences on the WEB in 93 languages. The length of the obtained vectors is 1024.

### 4.3. Corpora preprocessing

For FrameBank corpus, we made the same text processing as in (Larionov et al., 2019). We filtered the dataset keeping only predicates with at least 10 examples and dropped infrequent roles, for which the dataset contains less than 180 samples. The final corpus version contains 52,751 examples for 44 unique semantic roles.

To obtain features the following linguistic processing steps were performed:

- tokenization and sentence splitting with NLTK library (Schneider and Wooters, 2017);
- lemmatization, POS-tagging, and morphological analysis with MyStem library (Segalovich, 2003);
- syntax parsing via UDPipe parser (Straka and Straková, 2017) with model trained on SynTagRus (Nivre et al., 2008).

These steps are implemented using a publicly available IsaNLP library[4].

### 4.4. Results

We compare all models in terms of macro-averaged precision (P), recall (R), and F1-measure (F). Training and testing sets of FrameBank are adopted from (Larionov et al., 2019) for a fair comparison. The results of multilingual models as well as state-of-the-art **RuBERT** model from (Larionov et al., 2019) are presented in Table 1. **RuBERT** is the Russian Cased BERT pretrained on the Russian part of Wikipedia and news data (Kuratov and Arkhipov, 2019); it has 12 heads, 12 layers, 768 hidden units per layer, and a total of 180M parameters; Multi-BERT was used for initialization, while the vocabulary of Russian subtokens was built on the training dataset.

There are several conclusions to be drawn based on the results in Table 1. First, the models with BERT-context and XLM-R embeddings show the best results among non-pretrained models in terms of F-measure (78.4% and 78.3%, respectively). The model with BERT-context embeddings achieves the highest precision (82.8%), while the model with XLM-R demonstrates the highest recall (76.5%). The model with BERT-based embeddings for individual words shows lower scores than the BERT-context model, where sentences were used to obtain embeddings.

Second, the pretraining on FULL FrameNet improves results for all models except model with XLM-R embeddings. The model achieves the best improvement with BERT embeddings (+2.1%). The model with BERT-context embeddings obtains the best results in terms of recall (83.2%) and F-measure (79.0%) among models pre-trained on full FrameNet corpus. Pretraining on full FrameNet led to an increase in recall metrics for all models on 2.8-8.6%, while the precision metric reduced on 4.6-7.4%.

Third, for the REDUCED setup, the lower number of training examples from FrameNet improves the results for the model with BERT-context embeddings only (+0.8% of F-measure). The precision of the model with BERT-context embedding improves on 6%, while recall reduces on 4.8% compared to a model trained on full FrameNet corpus.

[4] https://github.com/IINemo/isanlp

| Model | P | R | F |
|---|---|---|---|
| BERT, Multilingual (Larionov et al., 2019) | - | - | .757 |
| RuBERT, Russian (Larionov et al., 2019) | - | - | .801 |
| **Without pretraining on FrameNet** | | | |
| BERT | .820 | .739 | .766 |
| BERT-context | **.828** | .746 | .784 |
| XLM-R | .815 | .765 | .783 |
| MUSE | .818 | .733 | .772 |
| LASER | .811 | .720 | .762 |
| **Pretrained on the Full FrameNet** | | | |
| BERT | .768 | .808 | .787 |
| BERT-context | .754 | **.832** | .790 |
| XLM-R | .759 | .793 | .773 |
| MUSE | .772 | .782 | .777 |
| LASER | .756 | .774 | .764 |
| **Pretrained on Reduced FrameNet** | | | |
| BERT | .766 | .805 | .784 |
| BERT-context | .814 | .784 | **.798** |
| XLM-R | .739 | .793 | .762 |
| MUSE | .736 | .784 | .758 |
| LASER | .747 | .786 | .764 |

Table 1: The model performance results.

To sum up, our results demonstrate that models, pretrained on the FULL version of FrameNet and fine-tuned on FrameBank, obtain higher recall and F-measure scores; from the other side, pretraining on English data for SRL decreases precision. The reducing number of FrameNet examples improves results for the model with BERT-context embeddings only.

## 5. Conclusion

We contribute to the transfer learning research by providing a first study on the effectiveness of exploiting English SRL data to boost Russian SRL performance. We study two setups for the source FrameBank dataset. Our experiments with several multilingual embeddings on the FrameBank dataset show that pretraining on the English FrameNet yield improvement for BERT-, LASER-, and MUSE-based models. Among four models, the model with BERT-based contextualized embeddings obtains the best macro-averaged F1-measure of 79.8%. We have demonstrated that it is beneficial to have the same set of roles in both corpora to order to boost the semantic role labeling performance.

We are currently working on the integration of FrameBank into the Linguistic Linked Open Data (LLOD) cloud (Cimiano et al., 2020; McCrae et al., 2016). According to our project, FrameBank will be interlinked with: **1)** the LLOD representation of FrameNet (Rospocher et al., 2019); **2)** other linguistic resources from the LLOD cloud, such as WordNet (McCrae et al., 2014), BabelNet (Ehrmann et al., 2014) and RuThes Cloud (Kirillovich et al., 2017; Galieva et al., 2017); and **3)** extralingual Linked Open Data resources, including DBpedia (Lehmann et al., 2015).

After that, we are going to retrain our model based on the newly obtained links. We hypothesize that these links can improve the accuracy of SRL against the baseline obtained in the presented paper.

## Acknowledgements

The work was funded by Russian Science Foundation according to the research project no. 19-71-10056.

## References

Arkhipov, M., Trofimova, M., Kuratov, Y., and Sorokin, A. (2019). Tuning multilingual transformers for named entity recognition on slavic languages. In Erjavec, T., Marcińczuk, M., Nakov, P., Piskorski, J., Pivovarova, L., Šnajder, J., Steinberger, J., and Yangarber, R., Eds., *Proceedings of the 7th Workshop on Balto-Slavic Natural Language Processing (BSNLP 2019)*, pages 89–93. Association for Computational Linguistics. `https://doi.org/10.18653/v1/W19-3712`.

Artetxe, M. and Schwenk, H. (2019). Margin-based parallel corpus mining with multilingual sentence embeddings. In Nakov, P. and Palmer, A., Eds., *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL 2019)*, pages 3197–3203. Association for Computational Linguistics. `https://doi.org/10.18653/v1/P19-1309`.

Baker, C. F., Fillmore, C. J., and Lowe, J. B. (1998). The berkeley framenet project. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics (COLING-ACL '98), Volume I*, pages 86–90. Université de Montréal. `https://doi.org/10.3115/980845.980860`.

Bojanowski, P., Grave, E., Joulin, A., and Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5:135–146. `https://doi.org/10.1162/tacl_a_00051`.

Cimiano, P., Chiarcos, C., McCrae, J. P., and Gracia, J. (2020). Linguistic linked open data cloud. In *Linguistic Linked Data: Representation, Generation and Applications*, pages 29–41. Springer. `https://doi.org/10.1007/978-3-030-30225-2_3`.

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(Aug):2493–2537. `https://dl.acm.org/doi/10.5555/1953048.2078186`.

Conneau, A. and Lample, G. (2019). Cross-lingual language model pretraining. In Wallach, H. M., othersLarochelle, H., Beygelzimer, A., d'Alché Buc, F., Fox, E. B., and Garnett, R., Eds., *Proceedings of the 33rd Conference on Advances in Neural Information Processing Systems (NIPS 2019)*, pages 7059–7069. Curran Associates, Inc. `https://papers.nips.cc/paper/8928-cross-lingual-language-model-pretraining`.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. In Burstein, J., Doran, C., and Solorio, T., Eds., *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL 2019), Volume 1 (Long and Short Papers)*, pages 4171–4186. `https://doi.org/10.18653/v1/N19-1423`.

Ehrmann, M., Cecconi, F., Vannella, D., Mccrae, J. P., Cimiano, P., and Navigli, R. (2014). Representing multilingual data as linked data: the case of babelnet 2.0. In Calzolari, N. et al., Eds., *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC'14)*, pages 401–408. European Language Resources Association. `https://www.aclweb.org/anthology/L14-1628/`.

Galieva, A., Kirillovich, A., Khakimov, B., Loukachevitch, N., Nevzorova, O., and Suleymanov, D. (2017). Toward domain-specific russian-tatar thesaurus construction. In *Proceedings of the International Conference IMS-2017*, page 120–124. Association for Computing Machinery. `https://doi.org/10.1145/3143699.3143716`.

Gildea, D. and Jurafsky, D. (2002). Automatic labeling of semantic roles. *Computational linguistics*, 28(3):245–288. `https://doi.org/10.1162/089120102760275983`.

He, L., Lee, K., Lewis, M., and Zettlemoyer, L. (2017). Deep semantic role labeling: What works and what's next. In Barzilay, R. and Kan, M.-Y., Eds., *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL 2017), Volume 1: Long Papers*, pages 473–483. Association for Computational Linguistics. `https://doi.org/10.18653/v1/P17-1044`.

Kirillovich, A., Nevzorova, O., Gimadiev, E., and Loukachevitch, N. (2017). Ruthes cloud: Towards a multilevel linguistic linked open data resource for russian. In Różewski, P. and Lange, C., Eds., *Proceedings of the 8th International Conference on Knowledge Engineering and Semantic Web (KESW 2017)*, Communications in Computer and Information Science, vol. 786, pages 38–52. Springer. `https://doi.org/10.1007/978-3-319-69548-8_4`.

Kuratov, Y. and Arkhipov, M. (2019). Adaptation of deep bidirectional multilingual transformers for russian language. In *Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference "Dialogue"*, pages 333–339. http://www.dialog-21.ru/media/4606/kuratovyplusarkhipovm-025.pdf.

Kuznetsov, I. (2013). Semantic role labeling system for russian language. In Joho, H. and Ignatov, D. I., Eds., *ECIR 2013–Doctoral Consortium*, pages 15–18.

Lample, G., Conneau, A., Denoyer, L., and Ranzato, M. (2017). Unsupervised machine translation using monolingual corpora only. *arXiv:1711.00043*. http://arxiv.org/abs/1711.00043.

Larionov, D., Shelmanov, A., Chistova, E., and Smirnov, I. (2019). Semantic role labeling with pretrained language models for known and unknown predicates. In Angelova, G., Mitkov, R., Nikolova, I., and Temnikova, I., Eds., *Proceedings of Recent Advances of Natural Language Processing (RANLP 2019)*, pages 620–630. Incoma Ltd. https://doi.org/10.26615/978-954-452-056-4_073.

Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P. N., Hellmann, S., Morsey, M., van Kleef, P., Auer, S., and Bizer, C. (2015). Dbpedia – a large-scale, multilingual knowledge base extracted from wikipedia. *Semantic Web*, 6(2):167–195. https://doi.org/10.3233/SW-140134.

Loukachevitch, N. and Rusnachenko, N. (2020). Sentiment frames for attitude extraction in russian. *Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference "Dialogue"*, pages 526–537.

Lyashevskaya, O. and Kashkin, E. (2015). Framebank: A database of russian lexical constructions. In Khachay, M. Y., Konstantinova, N., Panchenko, A., Ignatov, D., and Labunets, V. G., Eds., *Revised Selected Papers of the 4th International Conference on Analysis of Images, Social Networks and Texts (AIST 2015)*, Communications in Computer and Information Science, vol. 542, pages 350–360. Springer. https://doi.org/10.1007/978-3-319-26123-2_34.

Lyashevskaya, O. (2012). Dictionary of valencies meets corpus annotation: A case of russian framebank. In Fjeld, R. V. and Torjusen, J. M., Eds., *Proceedings of the 15th EURALEX International Congress (Euralex 2012)*, pages 1023–1030. University of Oslo. https://euralex.org/publications/dictionary-of-valencies-meets-corpus-annotation-a-case-of-russian-framebank/.

McCrae, J. P., Fellbaum, C., and Cimiano, P. (2014). Publishing and linking wordnet using lemon and rdf. In Chiarcos, C., McCrae, J. P., Osenova, P., and Vertan, C., Eds., *Proceedings of the 3rd Workshop on Linked Data in Linguistics (LDL-2014)*, pages 13–16. European Language Resources Association.

McCrae, J. P., Chiarcos, C., Bond, F., Cimiano, P., Declerck, T., de Melo, G., Gracia, J., Hellmann, S., Klimek, B., Moran, S., Osenova, P., Pareja-Lora, A., and Pool, J. (2016). The open linguistics working group: Developing the linguistic linked open data cloud. In Calzolari, N. et al., Eds., *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC'16)*, pages 2435–2441. European Language Resources Association. https://www.aclweb.org/anthology/L16-1386.

Miftahutdinov, Z., Alimova, I., and Tutubalina, E. (2020). On biomedical named entity recognition: Experiments in interlingual transfer for clinical and social media texts. In Jose, J. M. et al., Eds., *Proceedings of the 42nd European Conference on Information Retrieval (ECIR 2020)*, Lecture Notes in Computer Science, vol. 12036, pages 281–288. Springer. https://doi.org/10.1007/978-3-030-45442-5_35.

Nivre, J., Boguslavsky, I., and Iomdin, L. (2008). Parsing the syntagrus treebank of russian. In Scott, D. and Uszkoreit, H., Eds., *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*, pages 641–648. https://dl.acm.org/doi/10.5555/1599081.1599162.

Okamura, T., Takeuchi, K., Ishihara, Y., Taguchi, M., Inada, Y., Iizuka, M., Abo, T., and Ueda, H. (2018). Improving japanese semantic-role-labeling performance with transfer learning as case for limited resources of tagged corpora on aggregated language. In Politzer-Ahles, S., Hsu, Y.-Y., Huang, C.-R., and Yao, Y., Eds., *Proceedings of the 32nd Pacific Asia Conference on Language, Information and Computation (PACLIC 2018)*, pages 503–512. Association for Computational Linguistics. https://www.aclweb.org/anthology/Y18-1058.

Palmer, M., Gildea, D., and Xue, N. (2010). *Semantic Role Labeling*. Morgan & Claypool.

Pradhan, S., Ward, W., Hacioglu, K., Martin, J. H., and Jurafsky, D. (2005). Semantic role labeling using different syntactic views. In Knight, K., Ng, H. T., and Oflazer, K., Eds., *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, pages 581–588. Association for Computational Linguistics. https://doi.org/10.3115/1219840.1219912.

Rospocher, M., Corcoglioniti, F., and Palmero Aprosio, A. (2019). Premon: Lodifing linguistic predicate models. *Language Resources and Evaluation*, 53:499–524. `https://doi.org/10.1007/s10579-018-9437-8`.

Ruder, S., Søgaard, A., and Vulić, I. (2019). Unsupervised cross-lingual representation learning. In Nakov, P. and Palmer, A., Eds., *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL 2019): Tutorial Abstracts, July 28, 2019, Florence, Italy*, pages 31–38. Association for Computational Linguistics. `https://doi.org/10.18653/v1/P19-4007`.

Rusnachenko, N., Loukachevitch, N., and Tutubalina, E. (2019). Distant supervision for sentiment attitude extraction. In Mitkov, R. and Angelova, G., Eds., *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*, pages 1022–1030. INCOMA Ltd. `https://doi.org/10.26615/978-954-452-056-4_118`.

Schneider, N. and Wooters, C. (2017). The nltk framenet api: Designing for discoverability with a rich linguistic resource. In Specia, L., Post, M., and Paul, M., Eds., *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP 2017): System Demonstrations*, pages 1–6. Association for Computational Linguistics. `https://doi.org/10.18653/v1/D17-2001`.

Schwenk, H. (2018). Filtering and mining parallel data in a joint multilingual space. In Gurevych, I. and Miyao, Y., Eds., *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL 2018), Volume 2: Short Papers*, pages 228–234. Association for Computational Linguistics. `https://doi.org/10.18653/v1/P18-2037`.

Segalovich, I. (2003). A fast morphological algorithm with unknown word guessing induced by a dictionary for a web search engine. In Arabnia, H. R. and Kozerenko, E. B., Eds., *Proceedings of the 2003 International Conference on Machine Learning, Models, Technologies and Applications (MLMTA'03), June 23–26, 2003, Las Vegas, Nevada, USA*, pages 273–280. CSREA Press.

Shelmanov, A. and Devyatkin, D. (2017). Semantic role labeling with neural networks for texts in russian. In *Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference "Dialogue"*, volume 2, pages 245–256. `http://www.dialog-21.ru/media/3945/shelmanovaodevyatkinda.pdf`.

Solovyev, V. and Ivanov, V. (2016). Knowledge-driven event extraction in russian: corpus-based linguistic resources. *Computational intelligence and neuroscience*, 2016. `https://doi.org/10.1155/2016/4183760`.

Straka, M. and Straková, J. (2017). Tokenizing, pos tagging, lemmatizing and parsing ud 2.0 with udpipe. In Hajič, J. and Zeman, D., Eds., *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 88–99. Association for Computational Linguistics. `https://doi.org/10.18653/v1/K17-3009`.

Subburathinam, A., Lu, D., Ji, H., May, J., Chang, S.-F., Sil, A., and Voss, C. (2019). Cross-lingual structure transfer for relation and event extraction. In Inui, K., Jiang, J., Ng, V., and Wan, X., Eds., *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP 2019)*, pages 313–325. Association for Computational Linguistics. `https://doi.org/10.18653/v1/D19-1030`.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. In Guyon, I., von Luxburg, U., Bengio, S., Wallach, H. M., Fergus, R., Vishwanathan, S. V. N., and Garnett, R., Eds., *Proceedings of the 31st Conference on Advances in Neural Information Processing Systems (NIPS 2017)*, pages 5998–6008. Curran Associates, Inc. `http://papers.nips.cc/paper/7181-attention-is-all-you-need`.