МЕТОДЫ УСТАНОВЛЕНИЯ СЕМАНТИЧЕСКИХ РОЛЕЙ ДЛЯ ТЕКСТОВ НА РУССКОМ ЯЗЫКЕ

Шелманов А. О. (shelmanov@isa.ru),
Смирнов И. В. (ivs@isa.ru)

Институт системного анализа РАН, Москва, Россия

Ключевые слова: семантический анализ текста, семантико-синтаксический анализ, построение деревьев зависимости с помощью машинного обучения, семантически размеченный корпус

METHODS FOR SEMANTIC ROLE LABELING OF RUSSIAN TEXTS

Shelmanov A. O. (shelmanov@isa.ru),
Smirnov I. V. (ivs@isa.ru)

Institute for Systems Analysis of the Russian Academy of Sciences, Moscow, Russia

The paper introduces two methods for semantic role labeling of Russian texts. The first method is based on semantic dictionary that contains information about predicates, roles and syntaxeme features that correspond to the roles. It also uses heuristics and integer linear programming to find the best joint assignment of roles. The second method is data-driven semantic-syntactic parsing, which was implemented using MaltParser. It performs transition-based data-driven parsing simultaneously building a syntactic tree and assigning semantic roles. It was trained with various feature sets on SynTagRus Treebank, which was automatically enriched with semantic roles by the dictionary-based parser. We managed to automatically alleviate mistakes in the training corpus using output of the data-driven parser. We evaluated the performance of the parsers on the subcorpus of SynTagRus, which we manually annotated with semantic information. The dictionary-based parser and the data-driven semantic-syntactic parser showed close performance. Although the data-driven parser did not outperform the dictionary-based parser, we expect that it can be beneficial in some cases and has potentials for further improvement.

Keywords: semantic role labeling, semantic-syntactic analysis, data-driven dependency parsing, parser, semantic dictionary, semantically annotated corpus

1. Introduction

Semantic Role Labelling (SRL), sometimes also called shallow semantic parsing, is one of the best-known approaches to computational semantic analysis of natural languages. It supposes a simple model of text semantics, which treats a sentence or a clause as a situation (event or action) and entities represented in text as participants that play different roles in the situation. The model is widely used for information retrieval tasks, namely, information extraction, question answering, text summarization, machine translation and others.

Semantic role labelling consists of detection of sentence predicates expressing situation, identification of predicate arguments denoting participants related to the situation, and labeling the arguments with semantic roles. There is no conventional agreement about the set of semantic roles. However, there is a set of frequent roles that in some way are represented in most of the semantic theories. This set includes "AGENT"—the instigator of an action; "PATIENT"—participant that is affected by an action; "INSTRUMENT"—the mean by which an action is performed; "LOCATION"—the place of an action; "TIME"—the time of an action; "CAUSE"—cause of an action; "GOAL"—the entity, to which an action is directed, and others. The significant feature of semantic roles that distinguishes them from syntax relations is that they are extra-lingual concepts. The set of semantic roles and their granularity depend on domain and their application in solving high-level tasks.

We are developing the system that performs SRL-like semantic analysis of Russian texts. The early versions of the SRL system was already applied to solving many tasks of information retrieval [Osipov et al., 2008]. In this paper, we present two methods for semantic role labeling of Russian texts. The first one is a modification of the dictionary-based method implemented in the early versions of the system. The second method is using a data-driven transition-based parser for joint semantic-syntactic parsing of text. To train the data-driven parser we used SynTagRus—the Russian treebank [Apresjan et al., 2005], which we automatically enriched with semantic roles using the dictionary-based parser. We implemented technique to alleviate mistakes in the training corpus using output of the data-driven parser. Performance of the parsers was evaluated on the test corpus, which we manually annotated with semantic information.

The rest of the paper is structured as follows. Related work is reviewed in Section 2. Section 3 discusses the basic principles of our semantic model and two methods for semantic role labeling of Russian texts. Section 4 describes the experimental results for the described methods. Section 5 presents the analysis of errors, compares the described methods, and outlines future work.

2. Related Work

Semantic role labeling has been attracting attention of many researchers for the last 12 years. The mainstream approach to solve this task is to treat it as classification problem and to use supervised machine-learning techniques. This approach was pioneered by [Gildea and Jurafsky, 2002] and developed in many other works, which enhanced feature set [Xue and Palmer, 2004], applied different machine learning methods [Pradhan et al., 2004], utilized inference procedures (e.g. based on integer linear programming [Punyakanok et al., 2008]), and applied other methods for global scoring (e.g. reranking [Toutanova et al., 2005]). Due to large interest in semantic parsing of natural language, several shared tasks devoted to the problem were conducted [Carreras and Marquez, 2004], [Carreras and Màrquez, 2005].

One of the tendencies in natural language processing consists in treating semantic role labeling as establishing labeled dependencies between predicate and its arguments and applying approaches developed for dependency parsing (e.g. transition-based parsing [Choi and Palmer, 2011]). We consider this technique promising for semantic-syntactic parsing of Russian and apply it in our research.

Simultaneously with SRL, ideas of combining syntactic and semantic parsing together has been developing. Reason for combining this procedures lies in a hypothesis that parsing would benefit from tight interaction between syntactic and semantic layers. Several researchers tried to exploit this idea using various techniques. [Gildea and Jurafsky, 2002] used reranking on a set of diverse syntax parses and semantic structures. [Musillo and Merlo, 2006], [Merlo and Musillo, 2008] created joint constituency based syntax parser that simultaneously with syntactic analysis assigns semantic roles to constituents. Great contribution to the development of syntactic-semantic parsing was made by participants of CoNLL shared tasks 2008 and 2009 [Surdeanu et al., 2008], [Hajic et al., 2009]. The setup of the tasks considered dependency-based approach for both syntactic and semantic parsing and encouraged participants to create parsing techniques that combine them. Six research teams in 2008 and four teams in 2009 implemented systems that combined syntactic and semantic parsing in some way. In our research we implemented the technique that is quite similar to the one proposed by [Samuelsson et al., 2008], which utilizes MaltParser [Nivre et al., 2007] for joint syntactic-semantic parsing. However, [Samuelsson et al., 2008] did not apply this technique to Russian texts and did not use an automatically generated training set.

Although SRL is considered as well-researched problem for many languages (e.g., the participants of CoNLL Shared Task 2009 successfully solved it for seven languages with the same framework) the progress of SRL for Russian is relatively slow. There are several research groups working on the problem [Anisimovich et al., 2012], [Ermakov and Pleshko, 2009], [Kuznetsov, 2012], [Smirnov et al., 2014]. However, it seems that evaluation results of SRL systems for Russian language have not been published. One of the major reasons for this is absence of semantically annotated corpora for training and objective evaluation of SRL systems for Russian. Although Framebank project [Kashkin and Ljashevskaja, 2013] addresses this issue, it is still in development and cannot be used for machine learning techniques and evaluation so far.

3. Methods for Semantic Role Labeling

This section discusses semantic model and two methods for semantic role labeling developed to process Russian texts: dictionary-based parser and data-driven syntactic-semantic parser.

3.1. Principles of Semantic Analysis of Russian Language

We use relational-situational model of text [Osipov et al., 2008], [Osipov et al., 2013] as underlining model of text semantics. The model is based on the theory of Communicative grammar of Russian language [Zolotova et al., 2004]. The core concept of the Communicative grammar is syntaxeme—minimal indivisible semantic-syntactic structures of language, which possesses atomic semantic meaning. We interpret nominal syntaxemes (expressed by heads of noun phrases or by main nouns in prepositional phrases) as participants of situations and semantic meanings of syntaxemes as semantic roles. Therefore, in our case, SRL consists of detection of predicate words, identification of nominal syntaxemes, and labeling them with meanings that depend on the predicate words.

According to the relational-situational model, meaning of a syntaxeme in Russian texts is determined by preposition, grammatical case, and categorial semantic class (CSC) of the head noun of the syntaxeme. Categorial semantic class is a generalized meaning of a word. We distinguish the following categorial classes: “concrete” (material entities), “abstract” (immaterial entities, states, processes), “personal” (person that able to act purposefully), “location”, “time”, “measure”, “measurement parameter”, and “quantity”. Two syntaxemes of identical morphological form may have different meanings if they belong to different CSCs.

We use rich inventory of universal roles (i.e. meanings) that consist of more than 80 roles. They partially coincide with basic semantic roles used in other research. The top ten most frequent roles are: “Subject”, “Object”, “Predicate”, “Locative”, “Deliberative”, “Directive”, “Causative”, “Possessive”, “Result”, “Addressee”. More complete list of semantic roles and their description are represented in [Osipov, 2011]. The Fig. 1 demonstrates an example of a sentence taken from SynTagRus and labeled with semantic roles.

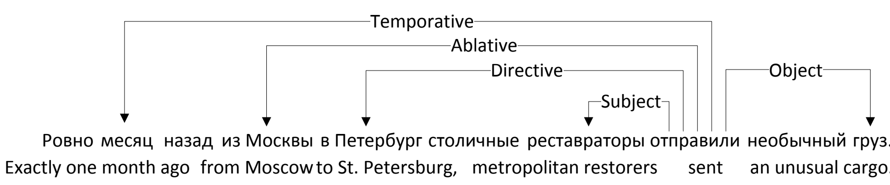

Fig. 1. Example of a sentence labeled with semantic roles. “Temporative”—time or period of time of an action; “Ablative”—starting point of an action; “Directive”—ending point of a motion; “Subject”—initiator of an action; “Object”—something that is affected by an action

3.2. Semantic Role Labeling Using Semantic Dictionary

To perform semantic parsing, we use semantic dictionary, which is being developed in Institute for Systems Analysis of Russian Academy of Sciences [Zav’jalova, 2004], [Osipov et al., 2008]. The dictionary stores frames that provide information about predicate and its semantic roles. The predicate is described by a set of predicate words with their lemmas. Information about roles includes sets of features that syntaxemes should have to obtain a specific role. These features include grammar case, preposition, and categorial semantic class. The developed semantic dictionary contains 2,856 frames and 3,585 predicate words. More detailed description of the semantic dictionary can be found in [Osipov, 2011]. Table 1 illustrates an example of a frame in the semantic dictionary for the situation expressed by the predicate word "отправить" ("send") from the example in Fig. 1. The dictionary-based semantic parser uses the semantic dictionary as the primary knowledge source for text processing.

Table 1. Frame in the semantic dictionary for the situation expressed by predicate words "отправить" ("send"), "направить" ("guide"), "послать" ("send"), "сослать" ("deport")

Semantic role	Categorial class	Preposition	Grammar Case
Ablative	Location, concrete	Из, из-за, из-под, от, с	Genitive
Addressee	Personal	К	Dative
Destinative	Any	Для	Genitive
Directive	Location, concrete	В, за, на	Accusative
Mediative	Concrete	По	Dative
Object	Any	–	Accusative
Objective	Any	За	Instrumental
		На	Accusative
Subject	Personal	–	Nominative

The input of the dictionary-based semantic parser is a list of sentences, which are split into clauses (simple sentences, participle expressions, and other locutions), tokens of each clause are organized in a dependency tree, and the clauses are linked with dependency relations. Tokens in a syntax tree have morphological features; nouns are assigned categorial semantic classes. CSCs are recognized by standalone CSC-processor, which uses additional dictionaries and heuristic rules.

The semantic parsing algorithm consists of the following major steps:

- predicate identification and search for a set of corresponding frames in the semantic dictionary;
- argument identification;
- argument classification;
- postprocessing.

The predicate identification step is performed using predicate words from the semantic dictionary and several pruning conditions that restrict found predicates to have particular part-of-speech (verb, noun, participle, etc.) and not to be modal verbs. For each found predicate, the parser searches for frames in the semantic dictionary by comparing lemma of the predicate with lemmas of predicate words in the dictionary. It is usual that more than one frame is found for a given predicate because of polysemy. Disambiguation of predicate sense and final choice of the frame are carried out during postprocessing step. For each found predicate, argument identification, argument classification and postprocessing steps are performed independently.

The argument identification step is performed using a system of heuristic rules. Clause of the predicate is the main scope of search for its arguments. The parser runs through words in the clause, checks whether they satisfy a number of conditions, and assigns them weights (between 0 and 1). The weight of a word indicates confidence that the word is an argument of the predicate. The weight assignment is driven not by classifier or statistical measures but by the system of heuristics. These rules take into consideration features of the predicate, part of speech of the given word, syntactic links between the predicate and the word. Links between clauses also determine potential arguments. Parser takes into consideration words outside of the predicate clause, which are linked to it with clause relations. The result of this step is a set of potential arguments of the predicate—words that got non-zero confidence weights.

In the argument classification step, for each found frame and for each potential argument parser tries to assign semantic label independently from the other arguments using the semantic dictionary. It compares features of the given argument with features that correspond to a specific role recorded in the given frame in the semantic dictionary. The parser examines grammar case, categorial semantic class and preposition. Depending on features of the predicate, comparison function differently treats grammar cases of the argument and the role in the dictionary frame. For example, if a predicate is in passive voice the instrumental and nominal cases are switched before comparison. Result of this step is a set of arguments that got zero or one label for each frame.

The postprocessing step consist of choosing the best combination of labeled arguments for each dictionary frame and applying domain constraints. Constraints demand argument structure not to have duplicate labels. Arguments also differ by their confidence weights and if two arguments claim the same semantic label it is reasonable to assign label to argument with the greatest weight. The task of choosing the best combination of labeled arguments can be considered as optimization problem and can be solved by means of integer linear programming (ILP) [Punyakanok et al., 2008]. In particular, to solve this task we applied Hungarian method. When distributions of semantic roles for each frame are built, the parser finally chooses the best frame and corresponding distribution of roles as the result. The best frame of the predicate is determined as the frame, which leads to the greatest number of assigned semantic roles in text.

The output of the parser consists of arguments labeled with semantic roles and predicate words labeled with descriptors of the best frames.

3.3. Semantic Role Labeling Using Data-Driven Semantic-Syntactic Parsing

Semantic role labeling can be performed using dependency parser that has ability to label relations between words. The labels of relations can be considered as semantic roles. Therefore, such dependency parser can perform joint semantic-syntactic analysis simultaneously building the syntax tree and labeling tokens with semantic roles. Although semantic relations do not always coincide with syntax relations and parser cannot determine predicates, using such semantic structure and external predicate labeler it is possible to restore semantic dependencies quite well.

To perform syntactic-semantic analysis we used MaltParser [Nivre et al., 2007]—the system for data-driven dependency parsing. MaltParser implements transition-based parsing framework, which includes various parsing algorithms and classifiers for transition prediction. It can build a dependency tree and label dependencies simultaneously. The framework provides ability to create complex features for classifiers including features that are based on partially built syntax tree and labels of established dependencies. Thus, MaltParser can perform joint semantic-syntactic analysis since it consults semantic labels and syntactic features during inference procedure.

To perform semantic-syntactic analysis MaltParser have to be trained on syntactically and semantically annotated corpus. There is only one substantial Treebank of Russian texts—SynTagRus corpus [Apresjan et al., 2005], which is part of National Corpus of Russian Language[1]. The corpus in conjunction with MaltParser was used for data-driven syntactic parsing of Russian texts by several researchers [Nivre et al., 2008], [Sharoff and Nivre, 2011]. However, there is still no substantial semantically annotated corpus for Russian for effective semantic role labeling using machine learning like PropBank [Palmer et al., 2005].

To overcome this problem we automatically annotated SynTagRus using our dictionary-based semantic parser described in 3.2. The dictionary-based parser was fed with gold syntax trees, gold morphology features, but automatically generated categorial semantic classes of nouns. The created corpus was used to train MaltParser to perform joint semantic-syntactic analysis with various features sets.

4. Experiments

This section describes experiments with the dictionary-based semantic parser and the data-driven semantic-syntactic parser.

4.1. Test Corpus and Evaluation Metrics

To evaluate performance of the semantic parsers we manually annotated subcorpus of SynTagRus with semantic information. The created semantic test corpus was annotated with nominal syntaxemes, their semantic roles and categorial semantic classes, predicates, and semantic dependencies between predicates and syntaxemes. However, not every nominal syntaxeme was annotated. The corpus contains annotations only for cases that can be found in the semantic dictionary. This limitation was introduced due to complexity of the task of manual semantic annotation of texts with rich set of semantic roles. It is also worth noting that the created corpus is not error-free and work on it is still in progress. The test corpus currently contains 1,730 sentences (29,041 tokens without punctuation), 3,871 tokens have semantic roles, and 61 roles are unique.

[1] Available at http://www.ruscorpora.ru

To evaluate performance of the semantic parsers we used three measures: precision, recall and F_1-measure. Evaluation took into account only tokens that have semantic roles in the test corpus. In this case, recall is the percentage of tokens that were properly assigned semantic role by a parser among all tokens with role labels in the test corpus; precision is the percentage of tokens that were properly assigned semantic role among all tokens that were assigned semantic role by a parser and have role labels in the test corpus. F_1-measure is the harmonic mean of precision and recall.

4.2. Experiments with the Dictionary-Based Semantic Parser

Before testing the dictionary-based semantic parser, we evaluated performance of the CSC-processor on the created test corpus. Since categorial semantic classes in the test corpus are assigned only to tokens that possess semantic role, we measured only precision of the processor. Precision is the percentage of tokens assigned proper CSC among all tokens that have categorial semantic class in the test corpus. The precision of the CSC-processor is 88.3%.

We evaluated performance of the dictionary-based semantic parser for the following cases:

- GoldSynt+GoldCSC—the input of the semantic parser consists of gold syntax trees and gold CSCs from the test corpus.
- GoldSynt+CSC—the input of the semantic parser consists of gold syntax trees from the test corpus and CSCs that are automatically generated by the CSC-processor. This case was used for generating the training corpus for the data-driven semantic-syntactic parser.
- Synt+GoldCSC—the input of the semantic parser consists of syntax trees that are automatically generated by MaltParser and gold CSCs from the test corpus. MaltParser for syntactic parsing was trained on the 48,096 sentences (74,665 tokens without punctuation) of the SynTagRus treebank with P3 feature set (see subsection 4.3). The unlabeled attachment score of the syntax parser is 88.0%.
- Synt+CSC—the input of the semantic parser consists of syntax trees and CSCs that are both generated automatically.

In all cases, morphological and lexical features were gold and were taken from the test corpus. To split sentences into clauses we used freely available NLP framework AOT.RU [Sokirko, 2001]. Table 2 shows performance of the dictionary-based semantic parser.

Table 2. Performance of the dictionary-based semantic parser

Case	Recall,%	Precision,%	F_1-measure,%
GoldSynt + GoldCSC	**82.5**	**94.5**	**88.1**
GoldSynt + CSC	70.4	89.3	78.7
Synt + GoldCSC	78.7	94.0	85.7
Synt + CSC	67.3	88.7	76.5

4.3. Experiments with the Data-Driven Semantic-Syntactic Parser

We trained the data-driven semantic-syntactic parser with four different feature sets. The basic feature set was used in [Sharoff and Nivre, 2011] to train syntax parser for Russian[2]. The basic set contains information about word form (FORM), lemma (LEMMA), part-of-speech (POSTAG), morphological attributes (FEATS), types of built relations (DEPREL). All of these features (except for DEPREL) were gold during training and testing. The tested feature sets were the following:

- P1 = Basic = FORM + LEMMA + POSTAG + FEATS + DEPREL;
- P2 = Basic + CSC (automatically generated categorial semantic classes of nouns);
- P3 = P2 + SPLFEATS (in this feature set we treat case, gender, and number as the separate features);
- P4 = P3 + descriptors of predicate frames identified by the dictionary-based semantic parser.

We trained MaltParser with LIBLINEAR library (optimized implementation of SVM without kernel function) [Fan et al., 2008] with "nivreeager" parsing algorithm [Nivre et al., 2007]. The training corpus contains 48,096 sentences (699,708 tokens without punctuation). Table 3 shows performance of the data-driven semantic-syntactic parser.

Table 3. Performance of the data-driven semantic-syntactic parser

Feature set	Recall,%	Precision,%	F_1-measure,%
P1	59.9	86.3	70.7
P2	60.4	86.5	71.1
P3	60.7	86.7	71.4
P4	**64.9**	**87.3**	**74.5**

Since the dictionary-based parser does not work perfectly, the training corpus contains mistakes. Error rate can be estimated by performance of the dictionary-based parser used to prepare training corpus that corresponds to the case "GoldSynt + CSC" in table 2.

To increase performance we suggested automatically enhancing the training corpus. We parsed the training corpus by the semantic-syntactic parser that had been trained with P4 feature set. Then we complemented training corpus with semantic labels that were found in the output of the semantic-syntactic parser but were absent in the initial training corpus. In addition, we removed sentences, which semantic labels seriously diverged from the output of the semantic-syntactic parser. From training corpus, we removed sentences, for which half and more semantic roles diverge from roles generated by the data-driven parser. Divergence was counted only if token was labeled with role both in the output of the parser and in the training corpus. We found that percentage of such sentences is less than 1%. Removing them helps

[2] Available at http://corpus.leeds.ac.uk/mocky/

to eliminate most noisy training examples that appear due to imperfect results of the dictionary-based parser. The enhanced corpus was used for training of new semantic-syntactic parser with the same feature set. We performed three iterations of this procedure. Table 4 shows performance of the semantic-syntactic parser on each iteration.

Table 4. Performance of the parser after enhancing the training corpus

#Iteration	Recall, %	Precision, %	F_1-measure, %
1	65.8	87.5	75.1
2	**66.3**	**87.7**	**75.5**
3	66.1	87.4	75.3

5. Discussion

The dictionary-based semantic parser suffers from mistakes in clause boundary recognition module. Performance of the parser also strongly correlates with quality of categorial semantic class recognition because it uses strict comparison between features of syntaxemes in text and features of roles in the semantic dictionary. However, the parser is less sensitive to errors in syntactic parsing because syntax relations only influence argument identification but not the argument labeling procedure. Evaluation of the dictionary-based parser also revealed mistakes in the semantic dictionary and some conflict situations while choosing proper semantic frame for a predicate.

Performance of the data-driven semantic-syntactic parser depends on performance of the dictionary-based parser and their mistakes notably correlate. The procedure for training corpus enhancement appeared to be beneficial. Using this procedure, we substantially increased performance of the semantic-syntactic parser. After two iterations of the procedure, both recall and precision of the parser increased. Recall increased by 1.3%, precision—by 0.4%, and F_1-measure—by 1.0%. The third iteration of the procedure decreased performance, which can be result of overfitting.

Performance comparison of the data-driven and dictionary-based parses should be done using the case “Synt + CSC” from table 2. Comparison of the best result for the data-driven parser from table 4 with the result of the dictionary-based parser in this case shows that performance of the data-driven parser is slightly lower than performance of the dictionary-based parser (the difference in F_1-measure is 1.3%). However, the used evaluation framework cannot provide purely fair comparison because the test set is not full and covers only cases that are represented in the semantic dictionary. Slightly more than 30% of roles found in the output of the parsers are not covered in the test set and could not be evaluated. The plausible case, in which data-driven parser could outperform the dictionary-based parser, is the case of unknown predicates. To test the ability of the data-driven parser to produce semantic labels for unknown predicates we plan to conduct additional experiments. For example, we could remove some predicates that appear in the test set from the semantic dictionary making these predicates “unknown”, then perform the described experiment and examine the cases with “unknown” predicates in the test set.

We intend to prepare semantically annotated test corpus with finer grained set of semantic roles (from 10 to 20 roles) and evaluate parser on it. The finer grained set would reduce labor of annotators, the amount of mistakes, and would make task of complete semantic annotation of the corpus feasible. We are also planning to include some additional features into the feature set and implement two data-driven parsers to experiment with techniques of co-training.

Good technique for evaluation of natural language components is performance evaluation of the final NLP application. Therefore, we intend to test the created SRL components as a part of IR systems that solve high-level tasks: question answering and information extraction from medical texts for English and Russian languages.

Acknowledgments

We are grateful to Laboratory of Computational Linguistics of Information Transmission Problems of Russian Academy of Sciences for provision of SynTagRus corpus. This work was partially supported by RFBR, project 13-04-12062.

References

1. *Anisimovich K. V., Druzhkin K. J., Minlos F. R., Petrova M. A., Selegey V. P. and Zuev K. A.* (2012), Syntactic and semantic parser based on ABBYY Compreno linguistic technologies, Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference “Dialog 2012”, Vol. 2, pp. 91–103
2. *Apresjan J. D., Boguslavskij I. M., Iomdin B. L., Iomdin L. L., Sannikov A. V., Sannikov V. G. and Sizov L. L.* (2005), Syntactically and semantically annotated corpus of Russian language: Present state and perspectives [Sintaksicheski i semanticheski annotirovannyj korpus russkogo jazyka: sovremennoe sostojanie i perspektivy], National Corpus of Russian Language: 2003–2005 [Natsional’nyj korpus russkogo jazyka: 2003–2005], pp. 193–214, (In Russian)
3. *Carreras X. and Màrquez L.* (2004), Introduction to the CoNLL-2004 shared task: Semantic role labeling, Proceedings of CoNLL-2004 Shared Task
4. *Carreras X. and Màrquez L.* (2005), Introduction to the CoNLL-2005 shared task: Semantic role labeling, Proceedings of the Ninth Conference on Computational Natural Language Learning (CoNLL-2005), pp. 152–164
5. *Choi J. D. and Palmer M.* (2011), Transition-based semantic role labeling using predicate argument clustering, Proceedings of the ACL 2011 Workshop on Relational Models of Semantics, pp. 37–45
6. *Ermakov A. E. and Pleshko V. V.* (2009), Semantic interpretation in text processing computer systems [Semanticheskaja interpretatsija v sistemah komp’juternogo analiza teksta], Information Technologies [Informatsionnye tehnologii], Vol. 6, pp. 2–7, (In Russian)

7. *Fan R.-E., Chang K.-W., Hsieh C.-J., Wang X.-R. and Lin C.-J.* (2008), LIBLINEAR: A library for large linear classification, The Journal of Machine Learning Research, Vol. 9, pp. 1871–1874
8. *Gildea D. and Jurafsky D.* (2002), Automatic labeling of semantic roles, Computational Linguistics, Vol. 28, pp. 245–288
9. *Hajic J., Ciaramita M., Johansson R., Kawahara D., Mart M. A., Màrquez L., Meyers A., Nivre J., Padó S., Štepánek J. et al.* (2009), The CoNLL-2009 shared task: Syntactic and semantic dependencies in multiple languages, Proceedings of the Thirteenth Conference on Computational Natural Language Learning: Shared Task, pp. 1–18
10. *Kashkin E. V. and Ljashevskaja O. N.* (2013), Semantic roles and construction net in Russian FrameBank [Semanticheskie roli i set' konstruktsij v sisteme FrameBank], Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialog 2013" [Komp'iuternaia Lingvistika i Intellektual'nye Tehnologii: Trudy Mezhdunarodnoj Konferentsii "Dialog 2013"], Vol. 1, pp. 325–343, (In Russian)
11. *Kuznetsov I. O.* (2012), Automatic identification of arguments of verbs: Theoretical background and state-of-the-art techniques [Avtomaticheskoe vydelenie glagol'nyh aktantov: teoreticheskaja osnova i aktual'nye podhody], Sci-tech information. Series 2: Information Processes and Systems [Nauchno-tehnicheskaja informatsija. Serija 2: Informatsionnye protsessy i sistemy], (12), pp. 2–7, (In Russian)
12. *Merlo P. and Musillo G.* (2008), Semantic parsing for high-precision semantic role labelling, Proceedings of the Twelfth Conference on Computational Natural Language Learning, pp. 1–8
13. *Musillo G. and Merlo P.* (2006), Accurate parsing of the proposition bank, Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers, pp. 101–104
14. *Nivre J., Boguslavsky I. M. and Iomdin L. L.* (2008), Parsing the SynTagRus Treebank of Russian, Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008), pp. 641–648
15. *Nivre J., Hall J., Nilsson J., Chanev A., Eryigit G., Kübler S., Marinov S. and Marsi E.* (2007), MaltParser: A language-independent system for data-driven dependency parsing, Natural Language Engineering, Vol. 13, pp. 95–135
16. *Osipov G.* (2011), Methods of artificial intelligence [Metody iskusstvennogo intellekta], FIZMATLIT, Moscow, (in Russian)
17. *Osipov, G., Smirnov, I. and Tikhomirov, I.* (2008), Relational–situational method for search and analysis of texts and its applications [Reljatsionno-situatsionnyj metod poiska i analiza tekstov i ego prilozhenija], Artificial Intelligence and Decision Making [Iskusstvennyj intellekt i prinjatie reshenij], (1), pp. 3–10 (in Russian)
18. *Osipov G., Smirnov I., Tikhomirov I. and Shelmanov A.* (2013), Relational–Situational Method for Intelligent Search and Analysis of Scientific Publications, Proceedings of the Workshop on Integrating IR technologies for Professional Search, in conjunction with the 35th European Conference on Information Retrieval (ECIR'13), Vol. 968, CEUR Workshop Proceedings

19. *Osipov G., Smirnov I., Tikhomirov I. and Zavjalova O.* (2008), Application of Linguistic Knowledge to Search Precision Improvement, Proceedings of 4th International IEEE conference on Intelligent Systems, Vol. 2, pp. 17-2–17-5
20. *Palmer, M., Gildea, D. and Kingsbury, P.* (2005), The proposition bank: An annotated corpus of semantic roles, Computational Linguistics, Vol. 31, pp. 71–106
21. *Pradhan S. S., Ward W., Hacioglu K., Martin J. H. and Jurafsky D.* (2004), Shallow semantic parsing using support vector machines, Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics: HLT-NAACL 2004, pp. 233–240
22. *Punyakanok V., Roth D. and Yih W.-t.* (2008), The importance of syntactic parsing and inference in semantic role labeling, Computational Linguistics, Vol. 34, pp. 257–287
23. *Samuelsson Y., Täckström O., Velupillai S., Eklund J., Fišel M. and Saers M.* (2008), Mixing and blending syntactic and semantic dependencies, Proceedings of the Twelfth Conference on Computational Natural Language Learning, pp. 248–252
24. *Sharoff S. and Nivre J.* (2011), The proper place of men and machines in language technology: Processing Russian without any linguistic knowledge, Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference “Dialog 2011”, pp. 591–604
25. *Smirnov I. V., Shelmanov A. O., Kuznetsova E. S. and Hramoin I. V.* (2014), Semantic-syntactic analysis of natural languages. Part II. Method for semantic-syntactic analysis of texts [Semantiko-sintaksicheskij analiz estestvennyh jazykov Chast’ II. Metod semantiko-sintaksicheskogo analiza tekstov], Artificial Intelligence and Decision Making [Iskusstvennyj intellekt i prinjatie reshenij], (1), pp. 95–108 (in Russian)
26. *Sokirko A.* (2001), A short description of Dialing Project, available at: http://www.aot.ru/
27. *Surdeanu M., Johansson R., Meyers A., Màrquez L. and Nivre J.* (2008), The CoNLL-2008 shared task on joint parsing of syntactic and semantic dependencies, Proceedings of the Twelfth Conference on Computational Natural Language Learning, pp. 159–177
28. *Toutanova K., Haghighi A. and Manning C. D.* (2005), Joint learning improves semantic role labeling, Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, pp. 589–596
29. *Xue N. and Palmer M.* (2004), Calibrating features for semantic role labeling, Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, pp. 88–94
30. *Zav’jalova, O. S.* (2004), About principals of creating dictionary of verbs for automatic text processing [O printsipah postroenija slovarja glagolov dlja zadach avtomaticheskogo analiza teksta], Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference “Dialog 2004” [Komp’iuternaia Lingvistika i Intellektual’nye Tehnologii: Trudy Mezhdunarodnoj Konferentsii “Dialog 2004”], (In Russian)
31. *Zolotova, G. A., Onipenko, N. K. and Sidorova, M. J.* (2004), Communicative grammar of Russian language [Kommunikativnaja grammatika russkogo jazyka], Institute of Russian language named after V. V. Vinogradov [Institut russkogo jazyka RAN im. V. V. Vinogradova], (in Russian).