Morpho-Syntactic Parsing Based on Neural Networks and Corpus Data

Roman Rybka, Alexander Sboev, Ivan Moloshnikov, Dmitry Gudovskikh
NRC "Kurchatov Institute"
Moscow, Russia
RybkaRB@gmail.com, sag111@mail.ru

Alexander Sboev
NRNU MEPhI
Moscow, Russia
sag111@mail.ru

***Abstract*—this article presentsmethodstoconstruct procedure of morpho-syntactic parsing based on corpus dataset analyzes. It contains 1) the method to eliminate morphological ambiguities using existing morphological parsers and then converting the results of parsing into the format of the language corpus used;2) a method of selecting parameters for syntactic parsing and assessment of the achievable accuracy of parsing, whichcanbe provided by the data of the usedcorpus; 3) a method of parsing sentences on the basis of neural network algorithms and a selected set of parameters in the format of used corpus. The basis for this study are sentences with unambiguous morpho-syntactic marking from the Russian National Corpus.**

I. Introduction

The growth in the intensity of information exchange in the virtual environment and in particular in social networks makes it necessary to use mathematical methods and software complexes for researchinto thevirtual environment and complex social phenomena displayed in it.In this area typical tasksare: automated annotation of text, content analysis of business information, sentiment analysis, analysis of emotive text identification of threats on social networks.

Solving these problems is often based on the usage of BigDatatechnology for the processing and analyzing of very large-scale data. This makes the computerization of the main stages of the analysis of unstructured data a critically important task.

The basis for the training and verification of systems of morphological and syntactic analysis arenational language corpora, in particular,Czech[1], British[2], Russian[3] and so on. Determining the syntactic structure of the sentence by linguists is based on the rules of natural language, for which a rigorous mathematical theory cannot be fully built.An effective means of solving above indicated problems are artificial intelligence techniques based on machine learning and generalizing ability,in particular, neural networks and probabilistic methods in the presence of sufficient base of morphological and syntactic parsing examples.Some of morphological analysis systems contain morphological homonyms removal functionality, such as the choice of the part of speech and features inside a part of speech (such as choosing the right number and case). For their implementation developers use simplified or transformed set of morphological tags which are different from that used in the corpus of dataset.Thus, comparison of analysis results with tagged data in corpus is difficult, and there is need to use a composite function. Also, there is a problem of joint use ofa syntactic parser,which is trained on examples from the corpus with morphological tags,and morphological analyzersusingsimplified or transformed set of morphological tags, which do not always reflect an unambiguous analysis.Aneffective solution of syntactic parsing using methodsofcorpus linguistics necessitates the complex study, including, on the one hand,which accuracy of parsing can be achieved with the data from language corpora, and on the other hand, which methods can be used to achieve such accuracy.Currently various methods are used to establish syntactic relations, in particular, the methods of probabilistic grammars (PCFG, Link Grammar), and methods of artificial neural networks (SVM, SRN, and RAAM, and others).

At the same time different types of corpus information are used:

- grammatical characteristics of individual words with the addition of features that characterize the properties of writing words, the presence of separators, their place in the sentence, and so on;
- word forms;
- other features.

In all cases, the application of one or other combinations of methods and information of corpus has its disadvantages and advantages for specific language. In particular, the methods of formal grammars are mainly used for languages with projective connections, while parsers based on recurrent neural networks [4] lose their analysis accuracy with increasing the number of words in a sentence.On the whole, the approach based on neural network models has some advantages due to the fact that neural networks show known generalizing properties. The use of these properties combined with parametric description of words along withmodern methods of data compression (to reduce the dimensionality of the address space of the task attributes) helps to build universal techniquesfor different languages.

In [5]the authors consider the method of forming features with classification convolution neural networks [6]. Extraction of the features can be performed either from whole sentence, or fromits window of words. Structure construction of the sentence is done using HMM. As shown in the article [7], approach based on deep learning [8] has demonstrated good

ISBN 978-952-68397-0-7 (paperback), ISBN 978-952-68397-1-4 (PDF)

results for English texts in solving problems of POS-tagging and chunking on base of IOBES format. An alternative approach uses a language model with features extraction of words based on the probabilities of co-occurrence of words in the training corpora presented in the works [9]. Syntactic (dependency) parser for some languages (English, Chinese, German, Arabic) [10] based on this approach has been built using a hybrid neural network.

Methods based on the model of transition [11], [12], [13], use a combination of features in a context of words in a sentence, and information on the previous analysis. This type of transition (or parsing rules) for the current state of analysis is calculated using the SVM.

In view of the above, the study from this angle morphological and syntactic methods that presented in this paper is relevant. It includes in particular:

- the method for eliminating morphological ambiguities using existing morphological parsers and then converting results of parsing into the format of corpus used;
- the method of selecting parameters for syntactic parsing and assessment the achievable accuracy of parsing, which the data of used corpus can provide;
- the method of sentence parsing on the basis of neural network algorithms and a selected set of parameters in a format of used corpus.

The basis for our study are sentences with unambiguous morpho-syntactic marking fromSynTagRus[14] (part of Russian National Corpus (RNC).

II. MATERIAL AND METHODS

A. SynTagRus

SynTagRus is an independent corpus implemented in RNC It includes electronic collection of tagged texts and can be used for research of vocabulary and language grammar.The each word in the texts of SynTagRusmatched to exactly one morphological structure, and each sentence is corresponded to theonly one syntactic structure.Themorpho-syntactic format of RNC (list of the various parts of speech, grammatical characteristics of words, type of syntactic relations) were the base for:

- implementing methods of morphological parsing in RNC format and decreasing results ambiguity,
- creating procedure of syntactic parsing of sentence.

B. Method of analyses morphological ambiguty and translate to corpus format

Open solution like an AOT, Mystem[15], etcare most often used for morphological parsing of Russian text. We chose Mystem for the implementationof the prototype due to ease of use in python. The accuracy of the parser was evaluated on the sentence of the SynTagRus. All of its sentences were parsed using Mystemand translated to format of morphological tags of RNC. We define as tag a full unambiguous set of morphological features of the word.AmbiguousMystemparsesaredivided into several tags in the NRC format. Then we compared the Mystemtags converted in RNC format and tags from marked sentences.Evaluation of the completeness of all possible tags from Mystem demonstrated that system gives the correct options for 94% of the words from theSynTagRus sentence. Comparison with unambiguous Mystemtags in the morphological NRCformat showed only 47% of matches, hence it needed improvement for the highest result. As a result of Mystemwork, all of possible morphological features of words puts in list that is contain weights based on they frequency. We have implemented an additional tag classifier based on support vector machines to reduce morphological ambiguityusing[16] and to increase portion of unambiguous tag coinciding with the parses from exemplary sentences. For this reason, all sentences are represented as a sequence of words $\{w_1 \ldots w_2\}$,where each word is a vector with some characteristics (see below). Punctuation marks are also counted as separate words and are replaced by common tag"PUNC". Sequential processing algorithm involves processing sentences from right to left, i.e. from the end of sentence. On each *i*-step it takes into description all the known features including the words which are already parsed on previous steps. A vector is generated for each word. The vector includes the features of the nearest neighbors in the window with size *W* that is chosen empirically. In this work we use eight words with next indexes:*i-3,i-2,i-1, i, i+1,i+2,i+3,i+4*,where *i* is the word that analyzed on step *i*. The feature vector includes the following information for each word:

- All word forms from the window *W*;
- Tags for those words of *W* that have been analyzed on previous steps;
- Classes of ambiguities for all words from *W* (+ their bigrams and trigrams). Class of ambiguity is the set of all possible tags for a word. We represent it as a string of concatenation of the tags. For example, for the Russian equivalent of the word "These" from the sentence-example class ambiguity looks like this: *adjective|nominative_case|plural_adjective|accusative_case|plural|inanimate*;
- Possible tags for each word;
- Subtags, i.e. individual morphological features for those words of *W* that the tags were fixed;
- Possible subtags for each unparsed word from the *W*.

The formation of a full set ofmorphological features of the words was based on all possible variants from Mystem. We allocated 43 morphological features in accordance with the RNC format.We have trained SVM classification models for each morphological feature. As a result of SVM model can give a particular class or the actual number - the so-called decision value, which gives a single score per sample in the binary case. Thus for each word we get 43 positive and negative decision values respectively. We evaluate all possible tags for each word as in (1),the tag that hasthe highest value T will be the winner.

$$T = \sum_{i=1}^{N_i} dv(y_i) + \sum_{j=1}^{N_j} dv(y_j) \quad (1)$$

where *N* as the number of subtags, *y* - subtag or morphological feature, *dv* - classification decision value, *i*as number of $y \in x$,*j* as number of $y \notin x$, where *x* is the set of morphological features included in the tag x.

C. Selection of the set of parameters for syntactic parsing and assessment of syntactic parsing accuracy

We have investigated four groups which consist of a set withcommon parameters and another set supplemented by us. The former includes: morphological characteristics of words; additional ones, such as indicator of punctuation after the word and indicator of capital letter; the distance between words. The latter includes potential syntactic relationships that are established on the basis of morphological characters of two words, further called p_sinto.

TABLE I.DESCRIPTION OF SETS OF PARAMETERS

Parameters \ set№	1	2	3	4
Morphologicalcharacters	+	+	+	+
Additional		+	+	+
The displacement of the main word to the dependent word of the pair in sentence			+	+
Potential syntactic relations (p_sinto) between main and dependent words (pair of words)			+	+
Potential syntactic relations from the words of the pair to other words in expression				+

The effectiveness of a set of parameters was evaluated by determining the degree of ambiguity in establishing syntactic relations described by this set. For this purpose the special procedure has been created [17], based on count of pairs of words in sentences from RNC and results of their parsing. Efficiency was evaluated by the number of ambiguous relationships: the higher is the number of ambiguous relationships, the worse is the efficiency of a set of parameters.

D. Methods of syntactic parsing

1) Approachestobuildinga syntactic parsetreeandto formationoftrainingexamplestodeterminesyntacticrelations

Two approaches were investigated to construct a syntactic parse tree:

- the first one is based on exhaustive enumerating of all possible options for establishing Sinto between words in a sentence;
- the second one is based on the Covington scheme [13] of incremental parsing.

In the first case, training set consists of pairs of words on all sentences RNC divided into two classes those which:

- form Sinto;
- do not form a Sinto.

The number of examples in the second class is much larger than in the first. Therefore we build method of filtering the pairs of words that form sinto. After that we determine syntactic relations in the filtered set of examples.

The essence of the second approach lies in building a model of transitions in three lists of words:

- R – right list consisting of all unparsed words,
- L – left list comprisingthe word for finding the relationship between R[0] and L[0],
- M – intermediate list. If relation between R[0] and L[0] has not been found, then the list M will be replenished with the word L[0].

Thus 4 classes of actions are used:

- No-Arc – transferring L[0] to the top of the list of M,
- Shift – moving the R[0], and all the words from M to L, so that the word of R[0] was the apex of L[0],
- Right-arc – setting a relation between R[0] and L[0],
- Left-arc – setting a relation between L[0] and R[0].

If the action is Right-arc or Left-arc, the word L[0] moves to M-list and becomes its apex. The training set in this case consists of examples corresponding to these four actions according to the type of sinto.

2) Determining syntactic relations

Methods MLP, SVM classifier based on strategy one-vs-all [18] and SGD[19], PNN[20], GNT[21], ensembles of decision trees (RFC)[22] in combination with the methods of reducing the dimension of the input space (Nystroem)[23] were investigated to determine the syntactic relations.

MLP neural network is trained by Error Back-Propagation algorithm. The number of neurons in the hidden layers is selected using a genetic algorithm. Neurons of hidden layers withactivationfunctions:suchas sigmoidal $f_1(x) = 1/(1 + e^{-\alpha x})$tangential $f_2(x) = \tanh(x/\alpha)$, where α– slope parameter of activation function were used.

For built classifier we also use linear scoring function (2).

$$f(x) = w^T x + b \tag{2}$$

where$x$ – input vector, w – parameters of the model, b – coefficient. Classes are predicted on the basis of sign of (2). Selecting parameters based on SGD is by minimizing the regularized training error given by (3)

$$E(w, b) = \frac{1}{n}\sum_{i=1}^{N} L(y_i, f(x_i)) + \alpha R(w) \tag{3}$$

where $y_i \in \{-1,1\}$ – the desired class of input examplex_i, L – loss function, R – regularization term (measures l2 or l1), α – a positive coefficient (hyper parameters), N – number of examples in training samples.

Various loss functions (SGD 1-4) of two arguments y and $p = f(x)$ can be used as L (4-7).

$$L(p, d) = \max(0,1 - pd) = \begin{cases} 1 - pd, pd \le 1 \\ 0, pd > 1 \end{cases} \tag{4}$$

$$L(p, d) = \ln(1 + \exp[-pd])\, pd), \tag{5}$$

$$L(p, d) = \max(0,1 - pd)^2 = \begin{cases} 0,1 \le pd \\ (1 - pd)^2, -1 \le pd \le 1, \\ -4pd \ge pd \end{cases} \tag{6}$$

$$L(p, d) = |d - p|, \tag{7}$$

Classificator at the N classes is based on SVM (2) with One-vs-all strategy. It consists of N binary classifiers. They solve the problem of choosing one of all classes (Cl):$\mathrm{Cl_i v(Cl_1, \dots, Cl_{i-1}, Cl_{i+1}, \dots, Cl_N)}$. The class is selected by the maximum probability of determining by all the independent binary classifiers.

PNN is probability neural network.

$$\mathrm{G(x; x_i) = \exp\left[\frac{-1}{2\sigma_i^2}\sum_{k=1}^{P}(x_k - x_{ik})^2\right]} \quad (8)$$

where$\boldsymbol{x_i}$– vector of values of the *i*-th neuron of network, $\boldsymbol{\sigma_i^2}$– dispersion,$\boldsymbol{P}$ – the size of the input vector.

Computational elements in PNN correspond to the values of input training examples. When testing the probability of the class of the input sample is determined using a Gaussian kernel (8).

GNT belongs to a class of neural networks with self-organizational process of learning that is based on winner-take-all-strategy. On the training phase the neuron that is the closest to the current input example wins and moves to the direction of the current object. On testing the input example gets the cluster name corresponding the name of the winner neuron.

We developed a complex model for determining syntactic relations that have small number of examples in training set. This model is based on the GNT and PNN. The basic idea of this model is to reduce the dimension of the training set on the training phase by replacing clusters of excess examples by centers of mass of this clusters. It is performed only for the clusters with examples not relevant to the considered Sinto. On testing phase examples were parsed by PNN.

RFC is classification method based on an ensemble of decision trees. We explored different quantities of decision trees (10 to 1000). In all cases we uses Nystroem algorithm for reducing the dimension of input space of example for RFC.

The following solutions are used for classification of Sinto and action of transitions:

1) Creating a sequence of classification neural network models to determine the Sinto or action independently (binary classification). The sequence is formed based on the number of examples for each class: from biggest to smallest;
2) Combining Sinto into several groups based on the number of examples for them, and on the accuracy of models for an independent classification of syntactic relations (or actions). Neural network models are created for each group of Sinto;
3) Creating a single model for multiclass classification for all Sinto (or actions).

In the first and second cases, each next classifier is trained on base of a training set which is free from the examples used for learning previous models.

E. Measures of accuracy of the syntactic parsingprocedure

Further construction of the model of syntactic parser for the Russian language is performed on the basis on selected approaches to build the syntactic tree and to the formation of a training sample. Its accuracy is evaluated according to the following values:

- UAS – unlabeled attachment score;
- LAS – labeled attachment score;
- TRD –true root determination. It is the ratio of number sentences with correct root determination to number of sentences;
- TSSP –true structure of syntactic parse without type of syntactic relations. It is the ratio of number sentences with right structure of syntactic parse tree to the total number of sentences;
- TSPT –true syntactic parsing tree. It is the ratio of number sentences with right syntactic parsing tree to the total number of sentences.

III. EXPERIMENTS

A. Using SynTagRus

Researchinto methods and algorithms was performed on a version of the SynTagRus, which contains 580 thousand of words and their parses, as well as about 43 thousand sentences. An analysis of the sentences using the morphological analyzer Mystemdemonstrated that it can be used to obtain morphological analysis for 38 thousand sentences from SynTagRus (483 thousand words).The other words have no morphological tags after analysis (for example, words written with numbers: 1799, etc.) or nodes in the syntactic structure of exemplary sentences contains several words for which there are no markings on the syntactic relations between them.Therefore, selected set of examples for further studyincludes 38 thousand sentences. 90% of randomly selected sentences were used as training samples and the remaining 10% sentences were used as test samples.

B. Testing the morphological tagger

Table II shows the results of the test identification of parts of speech.

TABLE II.POSCLASSIFICATION RESULTS

Features name	Number in testing set	Precision	Recall	F1-score
S	19689	0.99	0.98	0.99
A	7573	0.94	0.99	0.96
V	7459	1.00	0.99	0.99
PR	5157	1.00	1.00	1.00
ADV	2910	0.99	0.96	0.98
CONJ	2894	0.98	0.99	0.99
PART	2036	0.98	0.99	0.98
NUM	405	1.00	0.69	0.82
COM	21	0.46	0.13	0.20
INTJ	5	0.59	0.83	0.69
Total	48149	0.98	0.98	0.98

Obviously, some classes (COM, INTJ) were poorly represented in the training samples, and hence low result precision was observed for these ones. Comparison with other systems [24] showed that the quality of the implemented

method for solving PoS task is not inferior to the existing methods. The main goal was the decision of complex task: morphological analysis using RNC morphological markup and removal of ambiguity.Testing the algorithm of full morphological analysis showed precision=0.94, recall=0.92,and f1-score=0.93. The achieved accuracy coincides with the previously obtained test value for full set of the words in the corpus and meets our needs.

C. Selecting parameters for syntactic parsing

The results of Table III show that the fourth set of parameters has the best effectiveness. Further we will assess the accuracy of syntactic parsing using the 4-th parameter set.

The result of syntactic parsing is the syntactic tree of a sentence. At that the nodes correspond to the words or to their characteristics. The arcs correspond to the links, and their syntactic types. In accordance with the format used by the RNC, syntactic trees have several properties:

- the parse tree has only one vertex;
- there is only one input connection for all words in a sentence except for vertex;
- syntactic parse tree includes all the words in the sentence.

TABLE III.COMPARING SETS OF PARAMETERS

№ setofparameters	Average number of ambiguous syntactic relations for word of sentences	Percentage of clear syntactic relations determined
1	102,29	58,48
2	56,28	78,9
3	8,84	85,72
4	1,43	98,91

From this perspective, we have formed the criteria for the effectiveness of a selected set of parameters to syntactic parsing:

1) the number of sentences with unambiguous parsing;
2) the number of sentences with ambiguous parsing;
3) the average number of syntactic trees for ambiguous-parsed sentences.

The number of RNC sentences having after parsing properties “1-3” is calculated. If the sentence after parsing does not have the “2” property, then the procedure of normalization is carried out. Its goal is to transform each ambiguous parsing to several syntactic trees. If after normalization procedure the result does not have “1” and “3” properties, then this is an error of parsing.

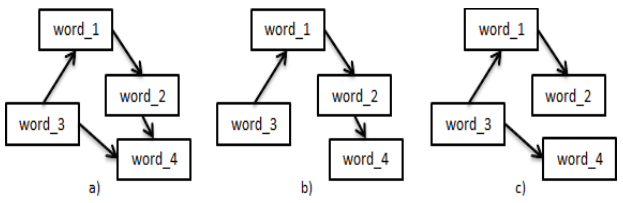

Fig. 2.Example of ambiguous parsing (a) and results of their normalization (b,c)

The evaluation results show that the proportion of uniquely-parsed sentences is 79.9% of the total number of sentences. The average number of parses of ambiguous-parsed sentences (20.1%) is 21.3.Classifications based on neural network PNN, MLP and SVM are designed to establish p_sinto. Positive and negative predictive values (PPV, NPV) of setting p_sinto are very high (see TableIV)

In the case of a MLP neural network, here and below the number of neurons in the hidden layer was chosen using a genetic algorithm.

TABLE IV.BEST MODEL TO DETERMINE P_SINTO

Number of syntactic relations type	Bestmodelof NN	PPV/NPV
1,32,33,42,52,54,55,57,58	PNN	99.9/99.8
5,6	SVM	99.9/99.9
2-4, 7-31, 34-41, 43-50, 56, 63-76	MLP 2 layer (40, 20 neurons)	99.8/99.6

Thus, for future work, we have chosen a set of parameters [17], including morphological characteristics, features of capitalization and punctuation, as well as p_sinto established on the basis of morphological characters of the two words in the sentence.

D. Building the procedure of syntactic parsing

1) Enumerating all possible combinations of words in the sentence and the definition of syntactic relations in them

In this case, procedure for determining the syntactic relations included the development of neural network models for two tasks:

*a) Filtering a set of examples that form syntactic relations (650 thousand) from those which do not form (10 million).*The following models demonstrated the best results:

- MLP (2 layers: 22 and 22 neurons), with estimations of PPV and NPV are equal 83.45/88.3%;
- MLP (1 layer: 50 neurons) with estimations of PPV and NPV are equal 82.1/89.12%.

b) Determining syntactic relations for the examples within the filtered set. The following methods were investigated for this:

- constructing separate neural network models to solve the tasks of binary classification for each type of syntactic relations;
- forming groups of syntactic relations based on the accuracy of their definitions.

The models based on MLP (2 layers) and SVM using method of stochastic gradient descent (SGD) [19] have demonstrated best results in solving the problem of independent determination of syntactic relations. Average PPV was 92.52 and NPV was 94.31%.

Average PPV in case of classification of examples of syntactic relations with small number of examples in training

set with use of model based on GNT and PNN is equal 99.1% (NPV is 99.8%).

We achieved the best result when created 7 groups of types of syntactic relations. Last group consisted of syntactic relations with small number of examples in training set. For their classification we also use the complex model (GNT and PNN). For other group best results were demonstrated by methods on base of SVM with training algorithm SGD. The use of grouping increases PPV to 95.6% (NPV to 97.4%).

Thus, the overall PPV of Sinto determination after selecting syntactically significant variants is 79.89%. This estimate was obtained after processing the test sets by model of definition of syntactically significant variants with use of MLP along with further classification of Sinto using methods SVM with SGD, GNT with PNN.

2) An approach based on incremental parsing scheme

The training set based on the Covington parse scheme is about 1.35 million examples, of which:

- 700 thousand relate to classes meaning the type of action without defining syntactic relation;
- 650 thousand relate to classes meaning the type of action with defining syntactic relation.

Several approaches have been analyzed (see TableVI):

- model based on ensembles of decision trees or SVM with strategies such as One-vs-all (see Table VI, variants of descent – "multiclass");
- the decision on the basis of a binary classification action when a classifier is constructed for each action.

The examples for the class actions that have already built classifiers are excluded from the training set. (see Table VI, variants of descent – "binary").

Direct usage of ensemble classification methods such as RFC requires a lot of RAM, so we used Nystroem[23]algorithm for compressing the input data space.

TABLE VI.PPV AND NPV FOR DEFINING ACTIONS IN INCREMENTAL PARSING SCHEME

Variantofdescent	Usingmethods	PPV	NPV
multiclass	SVM (linearkernel)	90.1	91.2
multiclass	Nystroem + RFC	83.8	84.1
binary	SVC+SGD (1-4)	87.3	88.1
binary	Nystroem + RFC and SVC+SGD (1-4)	89.2	90.1

Thus, the approach based on the incremental parsing scheme with the classifier based on SVM with linear kernel was selected for the implementation of the model for parsing the Russian language.

3) Testing

The results of testing the implemented model and comparison with other systems are presented in Table VII.

TABLE VII.ESTIMATION OF SYNTACTIC PARSING AND COMPARISON WITH ESTIMATION FROM LITERATURE SOURCES

Taskdescription	UAS (%)	TRD (%)	LAS (%)	TSPT (%)	TSSP (%)
Our estimations 1	85.81	82.23	79.33	14.05	29.47
Our estimations 2	91.73	88.84	89.39	35.91	52.38
Estimation from literature sources 1	94.3	---	92.3	29.7	37.4
Estimation from literature sources 2	89.4	---	83.4	21.8	33.3

"Estimation 1" was obtained by using method presented on previous section. It contains SVC for the classification of activities and the selected set of parameter. For obtaining "Estimation 2" we added word forms in set of parameters.

"Estimation from literature sources 1"was based on the use of ETAP-3 [25]. The basis of this system arerules [26] developed by linguists for analyzingRussian sentences. "Estimation from literature sources 2" based on incremental parsing scheme of Nivre-eager is presented in [12]. In this case set of parameters is based on word forms, parts-of-speech, and morphological features.The work [27] presents results of comparison of several syntactic parsers for Russian sentences. The average precision of UAS for all parsers was 88,8%.

Testing our models to determine syntactic relations on the corpus sentences showed that the accuracy of determining the type of syntactic relations is equal to 79.33%, and the accuracy of forming syntactic tree is equal to 14.05%. Adding word forms to the selected set of parameters increases the accuracy of the determination of syntactic relations by 10% and the one of forming syntactic tree by 20.7%.

IV. CONCLUSION

Methods for eliminating morphological ambiguities using Mystem and convertingmorphological features into the format of Russian National Corpus were developed on base of SVM classifier.Also ,a new method of selecting parameters for syntactic parsing that provide opportunity to determine syntactic relation with minimal ambiguity was proposed.. Achievable accuracy of parsing on base of RNC sentences with morpho-syntactic markup was calculated. Calculational evaluations show that the proportion of uniquely-parsed sentences is 79.9%. Some approaches to determine syntactic relation and to form a syntactic parsing tree were investigated. The procedure of syntactic parsing, based on using the combination of neural networks SVM, PNN, MLP for extraction potential relations, SVM for defining syntactic relations, andthe set of selected parameters with added word forms was developed.Syntactic parse of sentences (TSPT) of corpus dataset using this procedure demonstrates precision of 35.91%.This is a few higher than data from the literature sources give (see [28], [12]) but far lower than the evaluation ofachievable accuracy.The use of a model of syntactic parsing based on a system of sequential transitions leads to the accumulation of error resulting from inaccuracy of the definition of each syntactic relations (or actions). The increasing of quality of each classifier for determining the syntactic relations essentially improves the parsing in whole.

In the most of current systems, morphological analysis and syntactic analysis are carried out independently in two stagesOn the first stage, removing ambiguities is performed. On the second stage, parsing is performed on the number of optionsthat already have been reduced, as result of 1st stage. However, it should be noted that the feature space mostly is the same for both stages. This makes a division of the total morpho-syntactic problem in two stages artificial. Thus, the decision of the total morpho-syntactic task without division into separate stages is the aim of our future work.

REFERENCES

[1] "Czech National Corpus," Institute of the Czech National Corpus, [Online]. Available: https://www.korpus.cz/. [Accessed 28 09 2015].

[2] "British National Corpus," Oxford University Press, [Online]. Available: http://corpus.byu.edu/. [Accessed 28 09 2015].

[3] "RussianNationalCorpus," [Online]. Available: http://ruscorpora.ru/en/.

[4] Wong Chun Kit, "Recursive Auto-Associative Memory as Connectionist Language Processing Model-Training Improvements via Hybrid Neural-Genetic Schemata," *City University of Hong Kong*, Hong Kong, 2004.

[5] Collobert R. et al., "Natural Language Processing (Almost) from Scratch," *Journal of Machine Learning Research*, vol. 12, pp. .2493-2537, 2011.

[6] LeCun Y. et al., "Convolutional Networks and Applications in Vision.," *Courant Institute of Mathematical Sciences. ComputerScienceDepartment*., New York, 2010.

[7] R. Collobert. , "AISTATS," in Deep Learning for Efficient Discriminative Parsing, 2011.

[8] LeCun Y., BengioY.,Hinton G., "Deep learning," *Nature* 521, p. 436–444, 2015.

[9] Bengio Y. et al., "A neural probabilistic language model," The Journal of Machine Learning Research, vol. 3, pp. 1137-1155, 2003.

[10] Chen D., Manning C.D., "A fast and accurate dependency parser using neural networks," *Proc. EMNLP*, p. 740–750, 2014.

[11] Kübler S. et al., Dependency Parsing Synthesis Lectures on Human Language Technologies, 2009.

[12] Sharoff S., Nivre J., "The proper place of men and machines in language technology Processing Russian without any linguistic knowledge," in Proc. Dialogue 2011, *RussianConference on Computational Linguistics*., Moscow, 2011.

[13] Nivre J., "Incrementality in Deterministic Dependency Parsing," *School of Mathematics and Systems Engineering, Vaxjo*, Sweden, 2004.

[14] J. Apresjan, I. Boguslavsky, B. Iomdin, L. Iomdin, A. Sannikov, V. Sizov , "A Syntactically and Semantically Tagged Corpus of Russian: State of the Art and Prospects," *in 5th edition pf the International Conference Ressources and Evaluation*, Genoa, Italy, 2006.

[15] "Mystem," [Online]. Available: https://tech.yandex.ru/mystem/.

[16] V. V. Petrochenkov, A. O. Kazennikov, "A statistical tagger for morphological tagging of Russian language texts," *Avtomat. iTelemekh.,* no. 10, p. 154–165, 2013.

[17] Rybka R. et al., "Statistically selected set of parameters for definitions of syntactic relations in Russian language sentences," System analysis and information technologies. *Journal of the Voronezh State University.*, no. 2, pp. 117-124, 2014.

[18] Rocha A., Goldenstein S., "Multiclass from Binary: Expanding One-vs-All, One-vs-One and ECOC-based Approaches," in IEEE Transactions on Neural Networks and Learning Systems, 2013.

[19] Zhang T., "Solving Large Scale Linear Prediction Problems Using Stochastic Gradient Descent Algorithms," in ICML 2004, 2004.

[20] D. F. Specht, "Probabilistic neural networks," Neural Networks, vol. 3, no. 1, p. 109–118, 1990.

[21] Sboev A., et al, "Neuronetwork package Neurotree adapted for the segment of the Russian grid network," *Informatization and Communication*, 2012.

[22] L. Breiman, "Random forest," University of California, Berkeley, 2001.

[23] Kumar S., Mohri M., "Ensemble Nystrom Method," *Courant Institute of Mathematical Sciences,* New York, 2009.

[24] O.N. Lyashevskaya, I. Astaf'yeva, A. Bonch-Osmolovskaya et al, "NLP evaluation: russian morphological parsers," *Papers from the Annual International Conference"Dialogue",* vol. 9, no. 16, pp. 318-326, 2010.

[25] Kazennikov A., "A comparative analysis of machine learning dependency tree-based parsing algorithms," *Papers from the Annual International Conference "Dialogue",* no. 9 (16), pp. 157-163, 2010.

[26] JurijApresjan et al., "ETAP-3 Linguistic Processor: a Full-Fledged NLP Implementation of the Meaning <>Text Theory," Laboratory of *Computational Linguistics Institute for Information Transmission Problems,* Russian Academy of Sciences, Moscow.

[27] Anastasia Gareyshina, Maxim Ionov, Olga Lyashevskaya, Dmitry Privoznov, Elena Sokolova, Svetlana Toldova, "RU-EVAL-2012: Evaluating dependency parsers for Russian," *Proceding of COLING* 2012, Posters, , pp. 349-360, 2012.

[28] Kazennikov A., "A comparative analysis of machine learning," *Papers from the Annual International Conference "Dialogue"*, no. 9 (16), pp. 157-163, 2010.