Automatic Identification of Cognitive Actions Constituting Speech Genres of Scientific Theoretical Text

Dmitry Devyatkin[a], Ludmila Kadzhaya[b,c], Natalia Chudova[a], Valery Mishlanov[b] and Vladimir Salimovsky[b]

[a] *Federal Research Center "Computer Science and Control" RAS, 9 60-let Oktyabrya ave, Moscow, Russia*
[b] *Perm State University, 15 Bukireva str., Perm, Russia*
[c] *School of Translation Studies, Shandong University, 27 Shanda Nanlu, Jinan, P.R.China*

Abstract

This paper presents speech genres as forms of social and cultural activity at the stage of its objectification via a system of speech actions in a text as a communication unit. We implement the speech genre typology of a scientific text due to projecting information on the structure of the research and cognitive process onto the text that implements this structure. Moreover, we consider each of the studied speech genres as a cognitive-communicative action system characterized by a specific linguistic marker set. Those markers have been used as a linguistic base for the proof-of-concept implementation of a cognitive action parser for theoretical scientific text. Namely, we applied that linguistic knowledge to build compact high-level features, allowing the parser to be reliably trained on a small manually annotated corpus. The experiments on the corpus show the parser can accurately identify the actions.

Keywords

Speech genre, genre form, cognitive-speech action, scientific text, relational-situational model, sequence labeling.

1. Introduction

The theoretical foundation for most modern research on speech genres is the ideas of M.M. Bakhtin [1]. According to his well-known definition, speech genres are relatively stable thematic, stylistic, and compositional types of utterances. M.M. Bakhtin understands an utterance as a speech unit in which margins are determined by the change of speech subjects and completeness. In a similar sense, nowadays, the term text is being used.

M.M. Bakhtin considers the dependency of speech genres upon various types of «ideological creativity» — scientific, artistic, legal, political as a fundamentally important statement. He also emphasizes the importance of considering the research data that study these types of socio-cultural activities in linguistics [2].

Based on those statements, we define speech genres as relatively stable forms of spiritual socio-cultural activity at the stage of objectification via a system of speech actions in a text as a communication unit [3].

In addition to the ideas of M.M. Bakhtin, we are guided by A.N. Leontiev's study of human activity structure. In particular, his assertion that «activity is usually carried out by a certain set of actions, subordinated to special objectives, which can be singled out from a general goal» [4].

This paper uses the research results on the logic and methodology of scientific cognition. As is well known, it has empirical and theoretical levels. The empirical activity consists of applying the conceptual apparatus of science to the studied objects in observation and experiment, whereas

Proceedings of the Linguistic Forum 2020: Language and Artificial Intelligence, November 12-14, 2020, Moscow, Russia
EMAIL: devyatkin@isa.ru (1); kadzhaya@psu.ru (2); nchudova@gmail.com (3); vmishlanov@yandex.ru (4); salimovsky@rambler.ru (5)
ORCID: 0000-0002-0811-725X (1); 0000-0003-1275-9463 (2); 0000-0002-3188-0886 (3); 0000-0002-4925-2490 (4); 0000-0003-0417-8255 (5)

theoretical activity involves transforming and developing conceptual apparatus [5]. We regard the stages of empirical and theoretical cognition as the extralinguistic basis of appropriate speech genres.

The goal of this paper is to study the methods and develop proof-of-concept algorithms to identify cognitive-communicative actions in scientific texts. The objectives are:

1. Characterize the proposed linguistic and psychological bases for automatic identification of cognitive actions in scientific texts.

2. Develop and adapt the linguistic and software tools for detecting descriptions of cognitive actions in theoretical scientific publications.

3. Study the developed method and algorithms for detecting descriptions of cognitive actions in a theoretical scientific text.

The study material includes 160 scientific theoretical texts — papers and monograph chapters — in physics, biology, psychology, and linguistics (40 for each science).

The linguistic method is implemented in two stages. The first stage implies projecting the structure of the cognitive process [6] onto a text array and then correlating the primary goal of a particular text with one or another section of this structure. Therefore, the obtained speech genres typology of scientific empirical and theoretical texts looks as follows (Table 1 and 2):

Table 1

Extralinguistic determinancy of scientific empirical text's speech genres. (Empirical study).

Cognitive actions	Formation of basic empirical knowledge	Distribution of experimental data into groups	Working out the empirical laws
Speech genres	Description of a new scientific phenomenon	Classification text	Report on the empirical law of a cause-and-effect type

The table shows the results of our previous work [3]. The latter proves that the main stages of empirical scientific research, such as the formation of basic empirical knowledge, distribution of experimental data into groups, and working out the empirical laws, correspond to the specific speech genres: description of a new scientific phenomenon (mineral, plant, animal), classification genre, and report on the empirical law of a cause-and-effect type. A similar pattern is found in the study of theoretical texts (Table 2).

Table 2

Extralinguistic determinancy of scientific theoretical text's speech genres (Theoretical study)

Cognitive actions	Formation of theoretical ontology	Construction of theory on the found grounds	Explanation of facts by theory
Speech genres	Problem-stating theoretical text	Explication of its main concept	Verification text (experimental theory testing)

As can be seen from the diagram, the formation of theoretical ontology corresponds to the problem-stating theoretical text. Meanwhile, the construction of a theory is carried out in explicating its central concept by a system of less general ones. Explanation of facts by theory serves as its confirmation, verification.

We study each speech genre as a set of interrelated cognitive-communicative actions subordinated to a common goal. For example, the speech genre « Problem-stating theoretical text» is constituted by the actions:

1.1. presentation of theories that form available knowledge,

1.2. author's assessment of available knowledge,

The speech genre «Explication of its main concept » involves:

2.1. definition of the concept,

2.2. emphasizing an important thought,

2.3. explanation and clarification of the author's idea.

The speech genre «Verification text (experimental theory testing)» is comprised of the actions:

3.1. definition of a hypothesis to test,

3.2. description of the experimental methodology,

3.3. analysis and explanation of experimental data,

3.4. conclusion on the confirmation or refutation of the checked hypotheses.

Since the results of automatic analysis of empirical texts were presented earlier [7], we consider only the texts that embody the main stages of theoretical knowledge.

The second stage of linguistic analysis describes the multi-level language markers of utterances that realize the studied cognitive actions. At the same time, we become aware that it is crucial to create software tools that model human cognitive functions to solve problems in artificial intelligence. In our case, the point is that the perception of the utterance language form necessarily specifies the author's intention that ensures the adequacy of utterance understanding.

It should be pointed out that the indicator of a particular cognitive action is not the presence of particular linguistic means in the text itself, but the special nature of the speech system [8, 9], generated by peculiarities of choice, recurrence, placement, combination, and modification of multi-level language units.

We use the approach which differs from the numerous pieces of research, based on the genre-study concept of John M. Swales [10, 11] and applied in reviewing scientific texts [12, 13, 14]. While in those works, the object of analysis is generally a scientific paper Introduction (the authors have referred to the «Discussion» section recently), we study the whole text, not the set of rhetorical moves (cognitive and communicative actions) found in a paper's typical compositional part, but realization a substantial part of the cognitive process structure in a text as an integral communicative unit. Thus, our research is cognitive-oriented.

The work, perforce, is of a research outline: only the primary units of the scientific and cognitive process are considered, and within each unit – the most regularly implemented cognitive actions.

From the technical side, considering the speech system means that the cognitive action parser should consider multi-level linguistic features from each analyzed text fragment's wide context. Besides, as far as we analyze natural language, the lexis related to the particular actions can vary intensively. The typical approach to consider all those context-dependent features is to train deep neural models with recurrent and attention layers or to utilize a language model, such as BERT [15] or GPT-2 [16].

However, cognitive actions tagging is a new one, which means a lack of annotated corpora. Together with imbalanced classes, all those deep-learning models tend to over-fit. We propose a multi-step tagging approach to tackle this problem. Firstly, we extract various linguistic features for every clause; then, we apply the templates to generate a compact feature-set and fit a sequence labeling model with this reduced feature-set. The final step is the disambiguation of the clauses with multiple labels.

Such the approach also has reasoning from cognitive science. Namely, perception implies selectivity concerning the properties of the environment. It begins with a particular aim to receive such information on the objects' properties that contributes to the most appropriate behavior [17]. According to the ideas of development psychology about the existence of cultural standards of perception and thought [18], as well as the concept of objectification of mental reflection [19], information received by the analyzer system inputs, is interpreted considering objects' application practice in society (for animals – following the practice prescribed by instinctive behavior programs). As a result, the sensory-perceptual system processes not the isolated signals but the information about the whole objects. The objective nature of perception, defined by the organizing action criterion, allows recognizing the essential objects and events for an individual, their actions, and the others, intentions of subjects' interaction. Instead of tracking the whole set of parameters measured by neurophysiological detectors, a psychic subject operates with non-random structure discrimination rules – structures reflecting something meaningful in their life. After that, the application of the objectification principle in the suggested method for recognizing cognitive-communicative actions

enables us to move from statistical comparisons for a wide range of parameters to handling linguistic features relevant for identifying a subject's intentions.

The rest of the paper is organized as follows. Section 2 briefly provides the results of the recent studies in the closest research topic, which is RST (Rhetorical Structure Theory) parsing [20]. Section 3 contains a high-level description of the proposed approach to identifying cognitive actions and detailed descriptions of each approach's essential step. Eventually, Section 4 has the results of the experimental evaluation on a manually annotated corpus of scientific texts and some speculations regarding the reasons for the drawbacks identified.

2. Related work

The closest NLP-field to cognitive action identification is RST parsing. RST parsers are widely applicable for text summarization and information extraction. For example, paper [21] utilizes the hierarchical discourse-level structure of fake and real news articles to distinguish them. Such an approach has hardly been applied to the fake news detection problem because of the following issues. First, there is a lack of labeled corpora to train the methods for capturing the discourse-level structure for fake news. The second issue is how to extract useful information from those identified structures. To tackle these problems, they propose a hierarchical discourse-level structure for the fake news detection approach. This structure learns and builds a discourse-level structure for fake or real news pieces in an automated and data-driven manner. The researchers also revealed structure-related properties that can describe the revealed structures and increase detection accuracy. Another application of the RST is presented in paper [22], which proposes an approach to analyzing scientific discourse structures and the extraction of “evidence fragments” from a corpus of biomedical experimental research papers. The researchers trained and validated a scientific discourse tagger on two scientific discourse tagging corpora and checked if it can be transferred to a new dataset. They also show the benefit of utilizing scientific discourse tags for claim-extraction and evidence fragment detection. The experiment results show the applicability of evidence fragments derived from image spans for improving the quality of scientific claims by cataloging, indexing, and reusing evidence fragments as independent texts. In paper [23], researchers propose a new three-step approach to automatic text summarization. First, vector space modeling is used to compute coverage and fidelity scores. Then they apply fuzzy logic to evaluate an aggregated fidelity-coverage score. The last step is applying a discourse analysis on top of sentences, which have the highest fidelity-coverage scores to achieve coherence. The experiments on a labeled dataset show that the approach outperforms the state of the art extractive summarization models.

Although RST is a well-known theory, there is still a lack of human-annotated corpora to train parsers with machine-learning techniques. Therefore, the main research efforts are focused on unsupervised and semi-supervised approaches, such as methods to build dense contextualized vector representations (embeddings) of texts. These representations allow using a simpler machine-learning model to tackle the problem. For example, paper [24] presents an unsupervised automatic text summarization approach that combines rhetorical structure theory, deep neural model, and domain knowledge. This approach contains three crucial parts: domain knowledge base construction with representation learning, attentional autoencoder network for rhetorical parsing, and a subroutine-based network for text summarization. They use domain knowledge to increase the quality of unsupervised rhetorical parsing and utilize the concept of translation to tackle the lack of data to train the rhetorical parsing module. The summarization model hardly depends on the revealed discourse structure and can generate content-balanced results. They also present an unsupervised metric to evaluate the obtained results. The experiments show that the proposed approach has the same accuracy as other modern approaches. In [25], researchers propose a method that evaluates the applicability of language models for rhetorical analysis. Namely, they test their abilities to encode a set of linguistic features obtained from RST. The experiments show that BERT-based models [15] outperform others because they identify richer discourse features in their intermediate layer representations. It has also been shown in other studies [26] that the same BERT layers are responsible for holding syntax dependencies. This may be a clue to automatic revealing the

dependence between syntax and discourse. However, GPT2 [16] and XLNet [27] encode less rhetorical knowledge. They suggest this is because BERT considers context from both the left and right sides, which seems to be crucial for the task. It also does not permute context elements; therefore, it does not distort the original meaning of analyzed texts. The performance of the RST parser is also of research interest. Lin proposes an approach for sentence-level discourse analysis [28]. The process is two-step. First of all, they use a discourse segmenter to detect the elementary discourse units in a text and then apply a discourse parser that builds a discourse tree in a top-down manner. Both the components use Pointer Networks [29] and work in linear time.

Summarizing, the main focus of the current research in genre and discourse research lies in obtaining compact high-level feature sets, allowing one to train the parsers, and on the approaches to extend annotated corpora or transfer the training results. In this study, we also tackle this problem; however, to do so, we tried to consider the linguistic and cognitive background of speech perception, which is not assumed in pure statistical methods.

3. Cognitive action parser

3.1. Overall parser schema

Fig. 1 shows the overall scheme of the parser. Because of the small size of the training corpora, it is impossible to use end-to-end machine learning models to analyze clauses characterized by raw high-dimensional feature sets. Therefore, we separate the extraction of cognitive actions into several following steps: obtaining raw features with the full linguistic analysis of the clauses, matching the clauses with the context-free templates to generate a high-level feature-set with a smaller dimension, training of a model for sequence labeling with this compact feature-set, then disambiguation of the clauses, labeled with several cognitive actions.

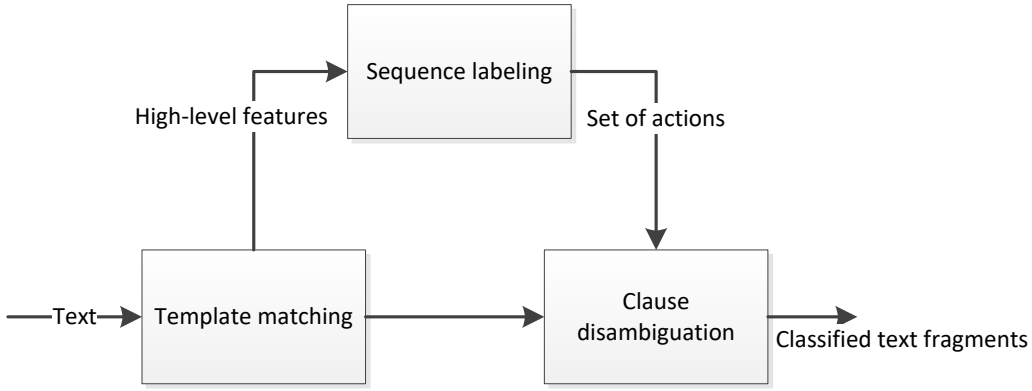

Figure 1: The approach to cognitive actions extraction

As we noted earlier, another reason for such an approach lies in psychology and cognitive science. Namely, an individual perceives information considering its own experience. As a result, the sensory-perceptual system processes not the isolated signals but the information about the whole objects. This schema allows recognizing the essential items and events for an individual, their actions, intentions, and interactions.

We gathered a corpus of 688 text fragments of research papers from various fields (Biology, Math, and Geology) to train and evaluate the parser. Each fragment has from 1 to 10 clauses and from 10 to 1K tokens [30].

3.2. Markers of the cognitive-communicative action

Considering the very limited scope of this paper, we describe only the most significant speech markers of the Russian texts that realize the given action. We use those markers as the linguistic

background to build the set of context-free templates. As an example of utterance markers' description, let us examine the means of implementing one of the main cognitive and communicative actions of a scientific theoretical text — definition of a concept (as shown in [31], in a particular text, definitive constructions can be considered as constituents of its mental space.). As a main syntactic model for defining a scientific concept, a construction with two Nominative cases and a quasi-copula *это* is used, and also the linking verbs *есть*, *суть (is)* (in constructions with plural of both N^1), including a zero one. E.g.: *N^1 — это …N^1; Фонем-**а** — это функциональная фонетическая единиц-**а**…; Фонем-**а есть** функциональная фонетическая **единица**…; Фонемы **суть** функциональные фонетические единиц-**ы**…; Фонем-**а** [ø] – функциональная фонетическая единиц-**а**…*(**dash** is an important formal marker). If the machine processes the corpora of only scientific texts, it is sufficient to determine the case of the nominal components of this structural model to automatically identify the given genre (utterances like *Убийца и есть дворецкий* in scientific texts are possible only as illustrations). While analyzing text arrays that are not discursively defined, the number of markers includes semantic characteristics of both $\mathbf{N^1}$ specified in pre-formed lists, and also in the morphemic structure of the construction's nominal components.

In the subject position (defined notion), it is expected either a substantive term from the object class (*животное*, *кристалл*) or feature names (*значение*, *коммуникация*), or — more often — phrases with a coordinated or uncoordinated attribute (usually in the form of $\mathbf{N^2}$) with a reference name belonging to the class of generalized nominations (*функциональн-**ый стиль**-ø — **это** [разновидность-ø язык-**а**…]*. Nomination terms in subject position ($N^{терм}$) are identified by the word-formation markers, mainly suffixes: *-ит-* (*лигнит*, *хромит*), *-ант-* (*инвариант*, *адресант*), *-ид*, (*хлорид*, *пестицид*), *-оид- (коллоид, суффиксоид)*, *-ин-* (*экзотоксин*, *папаверин*), *-ем-а* (*синтаксема*, *фразема*), *-а-ци-я-*, *а/ ени-е*, *-изм*, *-ость*, *-ств-о*, etc. Occurrence of commonly used nouns with a similar morphemic structure (or sound composition) is very rare (names like *апельсин*, *маргарин* in such syntactic models are unlikely).

The predicative position in this model is replaced by highly generalized categories of objects, attributes, processes (*unit, type, class, type, variety, variant; category, concept; process, action, construction, construct, education; attribute, property, nomination, word, material, substance; component, part, section, fragment* [+ $\mathbf{N^2}$], etc.), which subordinate the differentiating components required in this model — coordinated or uncoordinated attributes. For example, *Кванторн-**ые** местоимени-**я** — **это** языков-**ые** эквивалент-**ы** логическ-**их** оператор-**ов** – [кванторов существования и общности]*.

Variations of the studied model are the structures with the anaphoric pronoun ***это** (this, that)* in the subject position (its antecedent is contained in the immediately preceding sentence; E.g., ***Это — класс** позиционно чередующ-**ихся** звук-**ов***) and the structures with relative clauses, subordinate to $\mathbf{N^1}$ at the predicative position which is the antecedent of the relative pronoun ***который*** ($\mathbf{N^1}$ — ***это … [так-ой]*** $\mathbf{N^1}$ [Prep] ***котор***…; E.g.: *Дискурсивн-**ая деятельность**-ø — **это такая разновидность**-ø речев-**ой** деятельност-**и**, [Prep] **котор**-…)*.

3.3. Context-free template matching

In the first step, we use MyStem [32] and UDPipe [33] to detect lexemes, morphological features, and syntactic dependencies in the analyzed text clauses. Then we obtain predicate-argument structures and semantic roles with the parser developed at the FRC CS&C RAS [34]. Finally, we combine all those results to form relational-situational models of each clause. G. Osipov defines a relational-situational model as a heterogeneous semantic network (HSN) with the following structure [35]:

$$H =< D, N, S, R, F >, \quad (1)$$

Where

$D = \{D_1, D_2, \dots, D_m\}$ is a set of feature sets.

S – a set of tuples like $< n_j, \Delta_j >$ and n_j– a value of a syntaxeme from a name set N, $\Delta_j \subseteq D^k = D_l \times D_t \times \dots \times D_q$ – features of the syntaxeme for each $j = 1,2, \dots, |S|$ and $k \leq m$.

R is a set of relationships on N^2.

F is a set of functions$D^k \rightarrow D_j$, $j=1,..,m$.

In other words, the relational-situational model is an HSN, in which vertices are syntaxemes and edges define semantic relationships between the vertices. These syntaxemes are minimal indivisible semantic-syntactic structures of language. In our case, set *D* contains several morphological features (POS tags, grammatical cases, and moods, etc.) and embeddings of the lexemes. We use pre-trained FastText *(ruscorpora_none_fasttextskipgram_300_2_2019*) [36, 37] to build the character-level embeddings to deal with lexical richness and potential misspellings.

Let every clause of the analyzed text can be represented with this model, therefore all the text is the following sequence of HSNs: $\boldsymbol{H} =< H_1, H_2, .., H_t >$.

Define the context-free templates as a tuple $\boldsymbol{T} =< H_1, H_2, .., H_n >$ of HSNs. We have built more than 100 such templates based on the cognitive-communicative action markers presented in the previous section. Since the HSNs from $\boldsymbol{T}$ are templates, they hold only the feature descriptions essential for the identification. Here we presume that all the HSNs from $\boldsymbol{T}$ and $\boldsymbol{H}$ have the same feature sets *D*.

Eventually we define a reflection: $\varphi: \boldsymbol{H^t} \times \boldsymbol{T^n} \to \{0,1\}^{tn}$ in the following manner. For every text model $< D, N_i, S_i, R_i, F_i > \in \boldsymbol{H}$ and every template $< D, N_j, S_j, R_j, F_j > \in \boldsymbol{T}$ we set $\varphi_{ij} = 1$ if$N_i \cap N_j = N_j$, and$R_i \cap R_j = R_j$, otherwise we set$\varphi_{ij} = 0$. Fig. 2 represents an example of such matching, where the whole network represents a clause, and the red fragment is the part that matches a template.

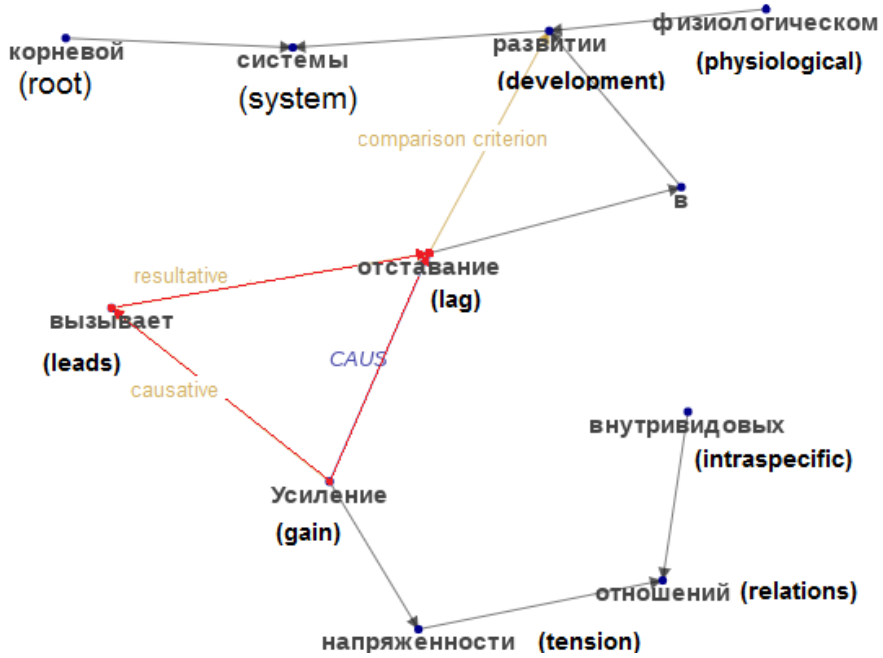

Figure 2: A template example and the HSN of the text that matches it

Although the vital question remains is how to intersect syntaxeme values. In our implementation, the analyzer matches two syntaxemes from different HSNs if they have the same morphology features and syntax dependencies (defined in the template's HSN), and the cosine distance between their lexis embeddings is less than the empirically-defined threshold. It is worth noting that this template-matching process is context-free. That means this process can be implemented with the algorithms, which have good performance and applicable for analyzing large texts.

Eventually, each clause is represented with a binary vector, which encodes if the clause matches the templates. Since these vectors are sparse, it is reasonable not to use them for further training directly but to build some dense embeddings before. We applied the implementation of SVD transformation from Scikit-Learn [38] to generate them.

3.4. Sequence labeling

In the second step, we apply machine learning to train the models to classify the clauses. Since one clause can represent several actions, we tackle this classification problem as a multi-label one. We have tested the following machine-learning methods with the sliding window approach:

1. Decision tree ensembles, such as a Random forest [39] and Gradient boosting on decision trees (XGBoost implementation) [40].
2. Linear SVM based classifier with L_2 regularization.

The training corpus is relatively small; therefore, in this study, we did not use recurrent networks or CRF for sequence labeling because they tend to over-fit even with our small artificial feature set, as we revealed earlier for the empirical text parser [7].

Since both SVM and tree ensembles are single-label classifiers, we have trained an independent binary classifier for each cognitive action. We selected all the training hyperparameters and the window size for the classifiers with a grid search on three-fold cross-validation.

3.5. Clause disambiguation

In the last step, we have to clarify the type of clauses, labeled with several cognitive actions. In more detail, the task is to build an *mXn* mapping between clause's tokens and labels. Since the annotated corpus is small, we have automatically labeled an additional dataset with the first two components of our approach and define this task as the multi-instance learning (MIL) problem [41]. According to the MIL, each clause (bag) is marked with a class label if this clause contains at least one token corresponding to that class. However, the problem definition, in our case, is non-classical. Instead of label ranking, we train a model to score a clause's tokens for each given label. Namely, we train Pointer Network-based regressor [29] to distinguish token labels inside the clause. The templates are again used to build features in this step, but now we apply them in the level of distinct tokens (Fig. 2).

4. Experiments and Results

The evaluation of classification scores was carried out using the statistical procedure of cross-validation [42]. Table 3 shows the normalized confusion matrix (Here, we hold the same numeration as in the introduction). Each cell of the table *(i, j)* contains the ratio of text fragments with action i, incorrectly classified as an action j. The table shows that classification errors are mainly associated with the recognition of actions "1.1", "1.2" ("*Presentation of theories that form available knowledge*", "*Author's assessment of available knowledge*") and "2.1" ("*Definition of a concept*"). In contrast, the rest of the cognitive actions are detected with an insignificant level of errors. The reason for that may lie in the simplicity of models we use, which catch quite a narrow context of a clause; therefore, this issue can be fixed if we label more data and use more complex models.

Table 3.
Confusion matrix for the sequence labelling step (XGBoost)

	1.1	1.2	2.1	2.2	2.3	3.1	3.2	3.3	3.4
1.1	-	*0,05*	**0,10**	0,01	*0,05*	*0,05*	0,00	0,03	0,00
1.2	*0,05*	-	**0,11**	0,01	0,02	0,01	0,02	0,01	0,00
2.1	*0,05*	0,04	-	0,03	0,02	0,03	0,00	0,01	0,01
2.2	0,01	0,01	0,02	-	0,00	0,01	0,01	0,00	0,00
2.3	*0,05*	0,02	0,01	0,01	-	0,01	0,00	0,00	0,00
3.1	0,03	0,01	0,03	0,01	0,02	-	0,02	0,01	0,01
3.2	0,00	0,03	0,05	0,01	0,01	0,01	-	0,00	0,01
3.3	0,03	0,02	0,02	0,00	0,00	0,00	0,00	-	0,02
3.4	0,02	0,00	0,00	0,00	0,00	0,02	**0,08**	0,01	-

Table 4 presents the classification scores for the sequence labeling step. It is worth noting that the best classification quality (by F1-score with macro-averaging) is achieved by the ensembles of decision trees with the gradient boosting method (XGBoost). The best recall scores are obtained with a linear support vector machine (SVM) because it is the simplest model in the test.

Table 4.
Performance of the sequence labelling step

	XGBoost			Random Forest			Linear SVM		
Code	F_1-macro	P	R	F_1 -macro	P	R	F_1 -macro	P	R
1.1	**0,78**	**0,92**	0,68	0,71	0,67	0,75	0,71	0,57	**0,93**
1.2	**0,95**	**0,99**	0,92	0,91	0,92	0,90	0,89	0,83	**0,97**
2.1	**0,98**	**0,99**	0,97	0,64	0,52	0,84	0,96	0,93	**1,00**
2.2	**0,94**	**0,99**	0,88	0,88	0,88	0,88	0,64	0,49	**0,91**
2.3	**0,99**	**0,99**	0,99	0,77	0,72	0,83	0,93	0,87	**1,00**
3.1	**0,97**	0,99	**0,94**	0,91	1,00	0,84	0,78	0,68	0,90
3.2	**0,98**	**0,99**	0,97	0,83	0,76	0,91	0,90	0,82	**1,00**
3.3	**0,93**	**0,93**	0,93	0,90	0,85	**0,96**	0,76	0,65	0,92
3.4	**0,98**	**0,99**	0,98	0,88	0,80	**0,99**	0,88	0,80	**0,99**

We do not provide the evaluation results for the clause disambiguation step because there is no representable manually labeled corpus for that, and this is out of bonds of our proof-of-concept study; therefore, that is a topic of further research.

5. Conclusion

The experiments show that the proposed cognitive actions can be accurately identified in scientific texts. We suppose that considering the patterns of speech perception, which is not assumed in pure statistical methods of the broadest possible set of linguistic features, can be the step towards developing intelligent systems focused on human cognitive functions. At the same time, it is essential that the object of our study, in contrast to others, is not only the introductory part of a scientific text or its largely standardized sections but the entire text.

The implemented parser can also be useful for solving applied problems related to science development. These problems include generating article's abstracts, detecting promising research areas, subject and methodological gaps, or denoting interdisciplinary interests, etc.

Although we got quite accurate results, it is clear that the further extension of the training corpus is necessary if we are building a full cognitive action parser. It seems that active learning would be an appropriate framework for that. It should be noted that the most complex and time-consuming part of the study is obtaining the linguistic markers. However, we believe there is the possibility to tackle this issue. It has been shown that linguistic models, such as BERT [25], can identify syntactic and discourse features in their intermediate layer representations. Therefore, the required linguistic markers could be extracted in a (semi-) automated manner if one had a reliable approach to the extraction of those features from the models.

6. Acknowledgements

This study is supported by Russian Foundation for Basic Research, grant No 17-29-07049 ofi_m.

7. References

[1] M.M. Bakhtin, Holquist M., McGee V., and Emerson C. The Problem with Speech Genres. Speech Genres and Other Late Essays. Austin: University of Texas Press, 1986, 43 pages.

[2] M.M. Bahtin, Pod maskoj. Maska vtoraya. [Behind the mask. Second mask] (In Russian), P.N. Medvedev, M.M. Bakhtin, A.J. Wehrle, "The formal method in literary scholarship: A critical introduction to sociological poetics", Harvard University Press, 1985.

[3] V.A. Salimovsky Zhanry rechi v funktsional'no-stylisticheskom osveshchenii (nauchnij akademicheskij text) [Speech genres in functional stylistic perspective (scientific text)]. Perm, PSU, 2002. 236 p. (in Russian).

[4] A. N. Leont'ev. Activity, consciousness and personality. Englewood Cliffs, NJ: Prentice Hall, 1978.

[5] V.S. Shvyrev Teoreticheskoe i empiricheskoe v nauchnom poznanii [Theoretical and empirical in scientific knowledge]. Nauka, 1978. p. 382 (in Russian).

[6] A.S. Maidanov Metodologiya nauchnogo tvorchestva [Methodology of scientific creativity]. Moscow, LKI Publ., 2008. 512 p. (in Russian).

[7] D. Devyatkin Extraction of Cognitive Operations from Scientific Texts. In: Proceedings of Russian Conference on Artificial Intelligence, Springer: Cham, 2019, pp. 189-200.

[8] M.N. Kozhina Rechevedenie: teoriya funkcional'noj stilistiki: izbrannye trudy [Speech studies: theory of functional stylistics: selected works]. ed. Flinta: Nauka, 2014. 624 p. (in Russian).

[9] B. N. Golovin Osnovy kultury rechi [The basics of speech culture]. Moscow, Vyssh. sk. Publ., 1988. 320 p. (in Russian).

[10] J.M. Swales Genre analysis: English in academic and research settings. Cambridge: Cambridge University Press, 1990. 261 p.

[11] J. Swales Research Genres: Explorations and Applications. Cambridge: Cambridge University Press, 2004. 314 p.

[12] S. Teufel, J. Carletta, M. Viens An annotationscheme for discourse-level argumentation in researcharticles. in: Proceedings of EACL'99: Ninth Conference of the European Chapter of the Association for ComputationalLinguistics, 8–12 June 1999. University of Bergen, Norway,1999, pp. 110–117.

[13] M. Liakata, S. Teufel, A. Siddharthan, C. Batchelor Corpora for conceptualisation and zoning of scientificpapers. in: Proceedings of the 7th International Conference onLanguage Resources and Evaluation. Paris, France: ELDA.LREC, 2010, pp. 2054–2061.

[14] A.I. Moreno, J.M. Swales "Strengthening move analysis methodology towards bridging the function-form gap" Journal of English for Academic Purposes. 50 (2017): 40–63.

[15] J. Devlin, M.-W.Chang, K.Lee, K.Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, in: Proc. 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. NAACL-LT 2019, Minneapolis, MN, USA, June 2-7, 2019, Vol. 1 (Long Short Pap., 2019: pp. 4171–4186.https://aclweb.org/anthology/papers/N/N19/N 19-1423/.

[16] A. Radrof, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever, Language Models are Unsupervised Multitask Learners, 2018.

[17] D. Bruner Psihologiya poznaniya. Za predelami neposredstvennoj informacii [Psychology of cognition. Beyond the immediate information]. Progress, 1977, 413 p. (in Russian).

[18] A.V. Zaporozhec, L.A. Venger, V.P. Zinchenko, A.G. Ruzskaya Vospriyatie i dejstvie [Perception and action]. Prosveshchenie, 1967, 323 p. (in Russian).

[19] A.N. Leontyev O putyah issledovaniya vospriyatiya (vstupitel'naya stat'ya) [On ways to study perception (introductory article)]. Vospriyatie i deyatel'nost'. / ed. by Leontyev A.N.. MSU, 1976, pp. 3-27. (in Russian).

[20] W. Mann, S. Thompson, Sandra A. "Rhetorical structure theory: toward a functional theory of text organization" Text: Interdisciplinary Journal for the Study of Discourse. 8(3) (1988): 243–281.

[21] H. Karimi, J. Tang Learning hierarchical discourse-level structure for fake news detection //arXiv preprint arXiv:1903.07389, (2019).

[22] X. Li, G. Burns, N. Peng Discourse tagging for scientific evidence extraction //arXiv preprint arXiv:1909.04758, (2019).
[23] A. B. Ayed, I. Biskri, J.G. Meunier Automatic Text Summarization: A New Hybrid Model Based on Vector Space Modelling, Fuzzy Logic and Rhetorical Structure Analysis. in: International Conference on Computational Collective Intelligence, Springer, Cham, 2019, pp. 26-34.
[24] S. Hou, R Lu. "Knowledge-guided unsupervised rhetorical parsing for text summarization" Information Systems, 94 (2020): 101615.
[25] Z. Zhu et al. Examining the rhetorical capacities of neural language models //arXiv preprint arXiv:2010.00153, (2020).
[26] C. D. Manning et al. "Emergent linguistic structure in artificial neural networks trained by self-supervision" Proceedings of the National Academy of Sciences (2020).
[27] Z. Yang et al. "Xlnet: Generalized autoregressive pretraining for language understanding" Advances in neural information processing systems (2019): 5753-5763.
[28] X. Lin et al. A unified linear-time framework for sentence-level discourse parsing //arXiv preprint arXiv:1905.05682, (2019).
[29] O. Vinyals, M. Fortunato, N. Jaitly "Pointer networks" Advances in neural information processing systems (2015): 2692-2700.
[30] Mental actions dataset. http://nlp.isa.ru/mental_actions, last accessed 26/11/2020/.
[31] D.A. Devyatkin, Y. M. Kuznetsova "Mentalnye deistviya i predmety v prostranstve nauchnogo discursa [Mental Actions and Mental Objects in the Space of Science Discourse]" Iskusstvennyi intellect I privyatie reshenyi, 1 (2020): 50-69. (In Russian).
[32] Mystem analyzer, https://tech.yandex.ru/mystem/doc/index-docpage, last accessed 26/11/2020
[33] M. Straka, J Straková. Tokenizing, pos tagging, lemmatizing and parsing ud 2.0 with udpipe, in: Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies, 2017, pp. 88-99.
[34] A. Shelmanov, D. Devyatkin Semantic role labeling with neural networks for texts in Russian, in: Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference" Dialogue, 2017, 1, pp. 245-256.
[35] G. S. Osipov, I. V. Smirnov, I. A. Tikhomirov "Relational-situational method for text search and analysis and its applications "Scientific and Technical Information Processing, 37 6 (2010): 432-437.
[36] A. Kutuzov, E. Kuzmenko Building web-interfaces for vector semantic models with the webvectors toolkit, in: Proceedings of the Software Demonstrations of the 15th Conference of the European Chapter of the Association for Computational Linguistics, 2017, pp. 99-103.
[37] T. Mikolov et al. Advances in pre-training distributed word representations, arXiv preprint arXiv:1712.09405, (2017).
[38] F. Pedregosa et al. "Scikit-learn: Machine learning in Python" The Journal of machine Learning research. 12 (2011): 2825-2830.
[39] L. Breiman "Random forests" Machine learning 45 1 (2001): 5-32.
[40] J. H. Friedman "Stochastic gradient boosting" Computational statistics & data analysis 38 4 (2002): 367-378.
[41] T. G. Dietterich, R. H. Lathrop, T. Lozano-Pérez "Solving the multiple instance problem with axis-parallel rectangles", Artificial intelligence, 89 1-2 (1997): 31-71.
[42] P. Flach Machine learning: the art and science of algorithms that make sense of data, Cambridge University Press, 2012.