Text indexation rate is about 1 Mb per min (the average indexation rate 100 Mb)

Time of index opening is not more than 1 min.

Search time is about 1 sec.

It should be noted that the technology being developed is not language dependent and can be adjusted to any language systems. Development of ideas put in searching the similar allows one to solve such problems as search of plagiarism, *rubrication* and *text clusterization*, Internet content filtration and anti-spam system creation.

Bibliography

[Dmitriev etc., 1991] Dmitriev A.S., Panas A.I., Starkov S.O. Storing and recognition information based on stable cycles of 1-D map // Phys. Lett. A.–1991–. V.155, –№8–pp.494-499.

[Dmitriev, 1991] Dmitriev A.S. Record and recognition of information in one-dimensional dynamic systems // РЭ. –1991–V.36– №1–p. 101-108.

[Andreyev etc., 1992] Andreyev Y.V., Dmitriev A.S., Chua L.O., Wu C.W. Associative and random access memory using one-dimensional maps // IJBC–1992–V.3–№3–pp.483-504.

[Andreyev etc., 1997] Andreyev Yu. V., Dmitriev A.S., and Starkov S.O. Information Processing in 1-D Systems with Chaos, *IEEE Transactions on Circuits and Systems*, 1997, vol. 44, No. 1, pp. 21-28.

[Andreyev etc., 1994] Andreyev J.V., Belskij J.L., Dmitriev A.S. Record and restoration of the information with the application of steady cycles of two-dimentional multivariate displays // РЭ–1994–V.39–№1–p.114-123.

[Andreyev etc., 1999] Andreyev Yu. V., Dmitriev A.S., and Ovsyannikov A.V. Information searching system "Forget-Me-Not" based on complex dynamics of nonlinear systems, *Proc. 7th Int. Workshop NDES-99*, 1999. Ronne, Denmark. pp.273-276.

Authors' Information

Kalafati Yuri Dmitrievich – Senior Scientist, PhD, Institute or Radio Engineering & Electronics, RAS, Moscow; e-mail: kalafati@controlchaostech.com

Moiseyev Konstantin Vladimirovich – Director, Controlling Chaos Technologies, Moscow; e-mail: moiseyev@controlchaostech.com

Starkov Sergey Olegovich – Professor, Obninsk State Technical University of Nuclear Power Engineering, Kaluga region, Obninsk; email: Starkov@jate.obninsk.ru

Shushkova Svetlana Alexandrovna – Undergraduate, Obninsk State Technical University of Nuclear Power Engineering, Kaluga region, Obninsk; e-mail: shushkovasvetlana@ramler.ru

COMMON SCIENTIFIC LEXICON FOR AUTOMATIC DISCOURSE ANALYSIS OF SCIENTIFIC AND TECHNICAL TEXTS *

Elena Bolshakova

Abstract: The paper reports on preliminary results of an ongoing research aiming at development of an automatic procedure for recognition of discourse-compositional structure of scientific and technical texts, which is required in many NLP applications. The procedure exploits as discourse markers various domain-independent words and expressions that are specific for scientific and technical texts and organize scientific discourse. The paper discusses features of scientific discourse and common scientific lexicon comprising such words and expressions. Methodological issues of development of a computer dictionary for common scientific lexicon are concerned; basic principles of its organization are described as well. Main steps of the discourse-analyzing procedure based on the dictionary and surface syntactical analysis are pointed out.

* The work is supported by the grant № 06-01-00571 of Russian Fond of Fundamental Researches (RFFI).

Keywords: scientific and technical prose, common scientific words and expressions, discourse markers, scientific discourse operations, discourse-compositional analysis.

ACM Classification Keywords: I.2.7 [Artificial Intelligence]: Natural language processing – Text analysis

Introduction

Intensive use of scientific phraseology is admittedly the most distinctive feature of functional style of scientific and technical prose [10]. The prose comprises documents of various genres and types: scientific papers, annotations, reviews, technical reports, etc. The phraseology includes both scientific and technical terms and various word expressions of common nature, such as English expressions *to test hypothesis, key concept, to be more precise, mentioned above, for this reason* etc. and Russian: *принятая гипотеза, по указанной причине, обосновать вывод, описанный ниже* and so on. Such lexical items are usually called ***common scientific words and expressions*** [5].

Both scientific terms and common scientific expressions are equally necessary to create a coherent and consistent scientific text. However, they differ in their functions within scientific discourse (scientific speech). Specific terms denote concepts, objects, and processes of particular scientific domains, whereas common scientific expressions are domain independent, they are used to organize scientific text narrative by expressing the logic of scientific reasoning, by structuring the text under development, by introducing cross-text references.

Lexicon of common scientific words and expressions is a syntactically quite heterogeneous set. It comprises, besides content (autosemantic) words, functional (auxiliary) words. Noun and verb-noun combinations, adverb and participle expressions, compound prepositions and conjunctions are included as well. Certain common scientific words and expressions are known as ***discourse markers*** [14, 15]: for instance, *Eng. in other words, Rus. другими словами*. Some word combinations are stable expressions exploited as ready-for-use colloquial formulas (clichés) [1], such as *Eng. as it was stated above*, *to outline directions of further research; Rus. из вышесказанного следует, как показало проведенное исследование*. It is worth noting that some clichés are common for scientific and technical prose, the others are specific for particular genres.

The paper describes the results of our research on elaboration of a computer dictionary of common scientific words and expressions, as well as on development of a procedure for discourse analysis of scientific and technical texts. The work is done within the overall framework of creating computational models of scientific and technical prose [2], supposing that really effective and deep automatic analysis of scientific and technical texts requires taking into account functional features of scientific and technical prose, in the first place, peculiarities of its phraseology and discourse.

The work started with an empirical study of scientific texts in several fields of exact sciences (mainly in computer science), so that scientific papers as the "core" of the functional style were analyzed. The study was initially performed for Russian scientific texts, and then expanded to English. In both languages the principal features of scientific discourse and lexicon proved to be the same.

As the work progressed, the importance of the common scientific lexicon became increasingly obvious, despite of its relatively small size. So we began to develop a computer dictionary comprising a wide range of common scientific words and expressions and providing a classification of their syntactic and semantic features. For Russian, the dictionary is now partially implemented; for English, only the classification work was done so far.

Our approach considers any lexical device signaling scientific discourse as discourse marker, and we include into our dictionary expressions specific for scientific and technical texts, such as English *by definition* or Russian *по определению*. As a result, the dictionary covers a wider set of lexical units than style-independent DiMLex lexicon of discourse markers [14] developed for German and English and mainly consists of conjunctions and conjunctive adverbs.

Instead of concept of discourse relationship proposed in well-known discourse theory RST [8] for explaining relations between adjacent phrases in text, we rely on the concept of ***scientific discourse operation*** fruitful for recognition of discourse-compositional structures specific for scientific and technical texts.

As units of common scientific lexicon may be served as surface cues, we assume the hypothesis that shallow text analysis based on the lexicon is adequate for detecting discourse structure of scientific text, without a deeper

syntactical-semantic analysis of all its sentences. Our discourse-analyzing procedure is now under development; it differs from the procedures developed for Japanese texts [7, 11] and based on deep syntactical analysis of sentences and a relatively narrow set of style-independent discourse markers.

The paper starts with an overview of specific features of scientific discourse derived from our empirical study, which also determined the set of common scientific words and expressions we consider as discourse markers. Illustrative examples from both languages – English and Russian are given. Then organizing principles for the computer dictionary of common scientific lexicon are described; and basic steps of the procedure for recognition of discourse-compositional structure of a given scientific text are presented. Potential applications of the dictionary and the procedure are pointed out in conclusion.

Scientific Discourse and Common Scientific Lexicon

The global purpose of scientific communication is to convey new ideas and results of scientific research, as well as to explain and rationalize them. Therefore, scientific discourse (speech) involves reasoning that is organized as a sequence of mental operations of informing and arguing. Among these typical operations one can notice assuming hypotheses, defining new terms and concepts, exemplification, resuming and so on. We call such intellectual operations *scientific discourse operations*.

As a rule, these mental-discourse operations are introduced into texts and more or less explicitly marked by authors of the texts with the aid of lexical devices – *common scientific words and expressions*. Therefore the words and expressions have metatext function and are often called *discourse markers* [15]. Accordingly to the applied discourse operations, scientific text is composed of *discourse segments*; in general, each segment comprises several adjacent sentences and includes discourse markers.

The most evident markers of scientific discourse are *mental performative expressions* like *Eng. we conclude, we would assume* or *Rus. мы докажем, мы предположим*. For Russian, they are described in detail in [12]. Performative expressions are based on "mental" verbs, e.g., *Eng. to conclude, to consider, to admit, to propose; Rus. заметим, рассмотрим, выразим* and so on. As a rule, these verbs explicate particular steps of scientific reasoning and have valences (complementing arguments): *Eng. we consider N*, *we conclude that S*; *Rus. рассмотрим N*, *подчеркивается, что S*. Besides pure "mental" verbs (*to conclude, to assume,* etc.), verbs of physical action (*to see, to show,* etc.) are used as mental.

Various forms of mental performative expressions are identified in scientific texts: *canonical forms, verbal variants, impersonal forms, and descriptive variants*. Canonical forms are based on mental verbs in the second person plural, often with the corresponding pronoun: e.g., *Eng. we resume, let us proceed to, we will proceed; Rus. мы покажем, мы рассмотрим*.

Verbal variants – e.g., *Eng. summing up, strictly speaking; Rus. подводя итоги, строго говоря* – are often used together will canonical forms: e.g., *Eng. refining the definition, wee see that…; Rus. суммируя вышесказанное, укажем…* . Impersonal forms – such as *Eng. it should be added, it was found, it is reasonable to assume; Rus. необходимо/нетрудно заметить, представляется, что…* – often include words of author's estimation (*should, reasonable, necessary*). Verbal and impersonal forms are used in texts to paraphrase canonical forms (e.g., *it was found* instead of *we found*) or to give some cross-text references (e.g., *as it was stated above*). Though they are less explicit forms than canonical, they are functionally equivalent.

One can also find in scientific and technical texts 'hidden' performatives, which we call descriptive variants: e.g., *These data are given in Table 3* stands for *We gave these data in Table 3*. Additional examples of descriptive variants are: *Eng. N is briefly described, N are given in; Rus. N кратко описано*.

Mental-discourse operations might be expressed by various parenthetic words and expressions: indicators of order (e.g., *Eng. first* or *lastly; Rus. во-первых, наконец*), markers of equivalency (e.g., *Eng. in other words; Rus. иными словами*), various connectives between textual parts (e.g., *Eng. nevertheless* or *so far; Rus. тем не менее, благодаря тому, что*) and so on. The metatext nature of these discourse markers is more obvious; they are used in texts of all styles.

As typical just for scientific and technical texts, we should point out abstract nouns, such as *problem*, *analysis*, *model, concept, conclusion* and so on. They are intended to name mental constructs by which scientific information is semantically structured. We call such nouns *common scientific variables*, since they have the

obligatory attributive valence (*problem of N, model of N*). Common scientific variables are mainly used with mental performative verbs, thereby forming stable noun-verb combinations, such as *Eng. to test hypothesis* or *to draw conclusions; Rus. подвергнуть анализу, проводить аналогию, опровергнуть гипотезу* [5]. Meanings of such verbs are close to Mel'čuk's Lexical Functions [9] with corresponding nouns as arguments.

Common Scientific Lexicon and Scientific Discourse Operations

Based on our empirical study, we propose for discourse analysis a particular set of scientific discourse operations, the most significant are presented in the Table 1.

We also propose to use the set of scientific discourse operations for semantical classification of heterogeneous collection of common words and word combinations, i.e., to classify them according to their discourse-organizing functions in scientific texts, irrespectively of their grammatical form and syntactic features.

Table 1. Scientific discourse operations

Operation	Russian Examples	English Examples
Description or statement	*укажем, что; характеризуя*	*let us to describe; we point out that*
Elaboration or adding information	*в частности; в дополнение к*	*to be more precise; in addition*
Expressing relations of causal, conditional, and concession type	*по этой причине; следовательно*	*hence; provided that; however*
Actualization of the topic	*перейдем к; рассмотрим*	*as for; let us consider; regarding*
Emphasizing	*особо подчеркнем; необходимо отметить*	*first of all; it is necessary to emphasize*
Presupposition	*предположим/допустим, что*	*we would assume; it may be admitted*
Definition	*будем называть; по определению*	*by definition; we call it/them,*
Comparison	*по сравнению с*	*as compared with*
Contraposition	*с одной стороны*	*on the one hand; as opposite to*
Illustration or exemplification	*к примеру; например*	*as illustrated below; for example*
Generation or resuming	*суммируя вышесказанное; в общем*	*in general; summing up*
Enumeration or ordering	*во-первых; наконец*	*next; finally*
Labeling with a scientific variable	*идея; модель; результат*	*result; idea; model*
Expressing of author's attitude	*целесообразно считать; по всей видимости*	*in our opinion*; *it seems reasonable*

Two following text fragments taken from English and Russian texts illustrate the usage of various scientific discourse operations and corresponding common scientific words and expressions (they are underlined):

Использование VBA–функции в данном случае является весьма целесообразным. Поясним это на примере с изменением процентной ставки. Допустим, вместо собственной функции мы использовали бы Len… Однако мы сталкиваемся с серьезной проблемой – как формировать такую структуру? … Именно по этой причине здесь не используется эта функция, хотя, на первый взгляд, она была бы вполне уместна. (1)

For dealing with discourse markers, we do not regard this distinction as particular helpful, though. As we have illustrated above and will elaborate below, these words can carry a wide variety of semantic and pragmatic overtones, which render the choice of a marker meaning-driven, as opposite to a mere consequence of structural decisions. (2)

It should be pointed out that besides units of common scientific lexicon, non-lexical devices are used to organize scientific discourse. In particular, such devices as sections, paragraphs, items, rubrics, and numeration are intended to structure scientific texts and to form their composition. All structuring and discourse-organizing devices present an interconnected system: devices can complement or substitute one another. For example, section headings are really substitutes for performative expression *we proceed to*, whereas numeration often complements performative formulas: e.g. *Let us enumerate main statements: 1)…2)…* . This interconnected system is rather excessive, since for most discourse operations there exist collections of similar lexical markers allowing for flexible paraphrasing.

In general, some discourse operation with its lexical and non-lexical devices can be used to implement another operation. For example, for categorization, a definition of new term is often required. As a result, certain discourse segments are embedded into some others, and in this way hierarchical structure of scientific text is formed.

Compiling the Dictionary of Common Scintific Words and Expressions

To develop the computer dictionary, collections of Russian and English common scientific words and word combinations were compiled from few available text dictionaries of scientific phraseology [4, 5] and from scientific texts in several fields of science (mainly in computer science and artificial intelligence), through their manual scanning. While selecting a word or expression for our collection, we used the following non-formal criteria:

- discourse-organizing function of the word or expression should be evident;
- it should be rather frequently used in texts of several scientific fields.

In addition, inter-language correspondences were used: for Russian expressions English equivalents were looked for, and vice versa.

Each compiled collection of common scientific words and expressions (for Russian and English) was divided into functional classes in accordance with the proposed list of discourse operations. Within each class, all words and word combinations that are semantically close and interchangeable in the texts as discourse markers were gathered into a group, thereby giving a subclass of functionally equivalent markers. Each group of functional equivalence often includes words of different parts of speech and contains from 2 to 9 units, the number depending on the language. For example, the resulted group of the consequence relationship includes for English: *hence, therefore, as a result, consequently, it follows that, we conclude that* etc., and for Russian: *значит*, *итак*, *таким образом*, *тем самым, как видим* etc. For both English and Russian, we obtain 53 groups corresponding to particular discourse organizing functions.

We should note that resulted groups of expressions differ in nature of objects being marked: while expressions of some groups indicate particular relations between text segments, the other mark particular text sentences or text segments.

Semantic and Syntactic Information in the Dictionary

To develop lexical entries of our computer dictionary, we considered requirements for its use by automatic text processing system, first of all, by discourse-analyzing procedure. The dictionary contains both units that correspond to words of common scientific lexicon and units representing word combinations. The former comprises all words of the lexicon, including those encountered in texts only within scientific expressions.

For a particular word, each unit stores adequate morphosyntactic information, including the part of speech and the flexional class (if any), as well as pointers to dictionary units describing available combinations with this word. In turn, unit for a particular word combination represents information about syntactical properties of the combination: stable *vs*. free, continuous *vs*. discontinuous.

For each dictionary unit considered as discourse marker, our computer dictionary provides semantic information that facilitates recognition of underlying discourse operation, namely:

- Functional class and group of the unit within the proposed semantic classification;
- Contextual conditions necessary for unit to be discourse marker within texts;
- Information about size and boundaries of implied discourse segment (one or several sentences; the beginning or the end of discourse segment is to be marked by this unit).

As most word combinations of common scientific lexicon have syntactic valences, we propose to represent information about valences with the aid of special ***lexicosyntactic patterns*** [3]. Each lexicosyntactic pattern fixes lexemes (constituent words of the particular combination) and their grammatical forms, as well as specifies syntactic conditions necessary for filing its empty slots (valences of the fixed lexemes). An example of such a pattern is *"let us consider" NP* with *NP* denoting a noun phrase; this pattern describes English expression that corresponds to discourse operation of topic actualization. Another example is *NP "we will call" T*, where *T* denotes an author's term and *NP* is a noun phrase explaining its meaning; the pattern represents the typical English expression for definition of new terms.

Lexicosyntactic patterns proved to be a convenient device for describing stable colloquial expressions comprising both phrasal formulas (like Eng. *the paper describes main features of*, *argument can be made against*) and predicative constructs (such as *to take as starting point for*). So a formal language for specifying lexicosyntactic patterns was elaborated, as well as a methodology for acquiring new patterns for the particular discourse operation from scientific and technical texts. Based on the acquiring methodology, a collection of patterns was created, which describes typical Russian single-sentence definitions of new terms. An example of lexicosyntactic pattern for discourse operation of defining a new term is given below:

«под» NP1 <case=ins> V<пониматься; tense=pres, person=3> NP2 <case=nom> < NP1.numb=V.numb>

where particular lexemes of the pattern are quoted, letter *V* denotes the verb, NP1 and NP2 denote noun phrases, and grammatical conditions are written within angle brackets – they specify values of grammatical parameters (tense, person, case, number) or establish their equality. The pattern describes typical Russian expression often encounted in scientific and technical texts, e.g. *Под графемной конструкцией понимается графическая форма, построенная из базисных, проблемно-ориентированных и/или графических конструкций* (fixed lexemes of the pattern are underlined) or *Под данными при такой формализации понимаются последовательности символов в некоторых алфавитах*.

Automatic Recognition of Discourse Structure

We consider discourse-compositional structure of scientific text as hierarchical structure of sequenced and embedded discourse segments, which corresponds to applied discourse operations and applied structuring devices. The structure may be represented as a tree, with tree nodes corresponding to discourse segments, and tree links fixing semantic (e.g., causal) and structural (e.g., embedding) relations between segments.

Our study of scientific discourse and common scientific lexicon showed that the lexicon has its own functional semantics, which makes it possible to superficially read scientific texts, i.e. to identify underlying discourse operations, to derive discourse-compositional structure of the texts, and thereby to comprehend logic of scientific reasoning, without deep understanding of these texts. So we are developing our procedure for recognition of discourse-compositional structure of scientific texts on the basis of shallow text analysis and the described computer dictionary. In order to reconstruct discourse-compositional tree for a given text, the recognition procedure takes the following steps:

1. Grapheme analysis of words, delimiting of sentences, and detecting of text composition elements: section headings, paragraphs, items, rubrics, and numeration.
2. Morphologic analysis of words and identification of all occurrences of common scientific words and word combinations.
3. Recognition of dictionary discourse markers in the given text through matching text fragments with those dictionary lexicosyntactic patterns that contain identified common scientific words.
4. Delimiting of discourse segments, based on recognized discourse markers and semantic information presented in the dictionary for functional groups and classes. In general case, the result of the segmentation is ambiguous: several plausible discourse trees fit the sequence of recognized markers.
5. Estimation of plausible discourse-compositional trees resulted at the previous step and selection of the most plausible one. A number of heuristic rules are used for this purpose (in particular, an exemplifying segment is rather embedded into another segment than includes it).

To implement steps 3 and 4, surface syntactical analysis of sentences is required, which takes into account: *i)* agreement and coordination of words; and *ii)* overall grammatical structure of sentences.

It should be noted that reliability of recognition of discourse-compositional structure depends on various factors, among them are the number and types of discourse markers encountered in the text. In order to increase the reliability, the other linguistic devices, such as anaphoric links and repetitions of lexical units in adjacent sentences are to be considered.

Conclusion

In the paper we have overviewed the features of scientific discourse and the wide spectrum of common scientific words and expressions, with their role in scientific discourse. We concerned the methodological issues of development of the computer dictionary comprising common scientific lexicon and providing semantic and syntactic information valuable for automatic discourse analysis of scientific and technical texts. We have also outlined heuristic multi-step procedure intended to recognize discourse-compositional structure of a given scientific text, with the aid of the dictionary and surface syntactical analysis of sentences in the text.

Potential applications for the proposed discourse-analyzing procedure include:

- Eliciting of knowledge represented in scientific and technical texts, in particular, extraction of new terms and their definitions introduced into texts by authors [13];
- Text abstracting, which may be based on processing of detected markers, e.g. the expression *we illustrate our approach with N* transforms into *the approach is illustrated with N* [6];
- Document browsing and intra-document information retrieval, which are especially topical for large-size technical documents;
- Computer-aided writing and editing of scientific and technical texts [2].

Some applications will supposedly be investigated after implementation, testing, and refinement of the dictionary of common scientific lexicon and the recognition procedure.

Bibliography

1. Bolshakova, E.I. Phraseological Database Extended by Educational Material for Learning Scientific Style. In: ACH/ALLC 2001: The 2001 Joint Int. Conference. Conf. Abstracts, Posters and Demonstrations, New York, 2001, p. 147-149.
2. Bolshakova, E.I. Towards Computer-Aided Editing of Scientific and Technical Texts // International Journal on Information Theories & Applications. V. 10, N 2, 2003, p. 204-210.
3. Bolshakova, E.I., Vasilieva N.E., Morozov S.S. Lexicosyntactic Patterns for Automatic Processing of Scientific and Technical Texts. In: Proc. of 10th National Conference on Artificial Intelligence with International Participation 2006. Moscow, Fizmatlit, Vol 2, 2006, p. 506-514 (in Russian).
4. Dictionary of Word Combinations Frequently Used in English Scientific Literature. Nauka Publ., Moscow, 1968.
5. Dictionary of Verb-Noun Combinations of the Common Scientific Speech. Nauka Publ., Moscow, 1973 (in Russian).
6. Ellouze, M., Hamadou A.B. Relevant Information Extraction Driven with Rhetorical Schemas to Summarize Scientific Papers. In: Advances in Natural Language Processing. Third International Conference, PorTAL 2002. E. Ranchhod and N.J. Mamede (Eds.). Lecture Notes in Computer Science, N 2389, Springer-Verlag, 2002, p. 111-114.
7. Kurohashi, S., Nagao M. Automatic Detection of Discourse Structure by Checking Surface Information in Sentences. In: COLING 94 Proceedings of the 15th Int. Conf. On Computational Linguistics. Vol. II, Kyoto, Japan, 1994, p. 1123-1127.
8. Mann, W.C., Thompson S.A. Rhetorical Structure Theory: Toward a Functional Theory of Text Organization. Text, 8 (3), 1988, p. 243-281.
9. Mel'čuk, I. Dependency Syntax: Theory and Practice. SONY Press, NY, 1988.
10. Mitrofanova, O. Language of Scientific and Technical Literature. Moscow University Press, 1973 (in Russian)
11. Ono, K., Sumita K, Miike S. Abstract Generation Based on Rhetorical Structure Extraction. In: COLING 94 Proceedings of the 15th Int. Conf. On Computational Linguistics. Vol. II, Kyoto, Japan, 1994, p. 344-348.
12. Ryabtseva, N.K. Mental Performatives in Scientific Discourse. Voprosy yazykoznaniya, V 4, 1992, p. 12-28 (in Russian).
13. Paice, C., Jones P. The Identification of Important Concepts in Highly Structured Technical Papers. Proc. of 16th Annual Int. ACM SIGIR Conference on Research and Development in Information Retrieval, Pittsburg, 1993, p.69-78.
14. Stede, M., Umbach C. DiMLex: A Lexicon of Discourse Markers for Text Generation and Understanding. Proceedings of Int. Conf. On Computational Linguistics COLING-ACM'98, Vol. 2, 1998, p. 1238-1242.
15. Wierzbicka A. Metatext w tekscie. In: O spojnosci tekstu. Wroclaw-Warszawa-Krakow-Gdansk, 1971, p.105-121.

Authors' Information

Elena I. Bolshakova – Moscow State Lomonossov University, Faculty of Computational Mathematics and Cybernetic, Algorithmic Language Department; Leninskie Gory, Moscow State University, VMK, Moscow 119899, Russia; e-mail: bolsh@cs.msu.su