Journal of Theoretical and Applied Information Technology
31st May 2020. Vol.98. No 10

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

LINK GRAMMAR AND FORMAL ANALYSIS OF PARAPHRASED SENTENCES IN A NATURAL LANGUAGE

[1]AISLU B. KASEKEYEVA, [2]TATIANA V. BATURA, [3]LYUBOV V. EFIMOVA, [4]FEODOR A. MURZIN, [5]JAMALBEK A. TUSSUPOV, [6,7]AIGERIM S. YERIMBETOVA, [7]KUNTUGAN ZH. DOSHTAYEV

[1]Doctoral student, L.N.Gumilyov Eurasian National University, specialty Information Systems, Kazakhstan
[2]PhD, Senior researcher, A.P. Ershov Institute of Informatics Systems SB RAS, Russia
[3]PhD student, A.P. Ershov Institute of Informatics Systems SB RAS, Russia
[4]PhD, Deputy director, A.P. Ershov Institute of Informatics Systems SB RAS, Russia
[5]Doctor of Physical and Mathematical Sciences, Professor, Department "Information Systems", Faculty of Information technologies, L.N. Gumilyov Eurasian National University, Kazakhstan
[6]PhD, Senior researcher, Institute of Information and Computing Technologies CS MES RK, Kazakhstan
[7]PhD, Kazakh Academy of Transport and Communications named after M.Tynyshpayev, Kazakhstan.

E-mail: [1]aibike_7474@mail.ru, [2]tbatura@iis.nsk.su, [3]efimova_l@ngs.ru, [4]murzin@iis.nsk.su,[5]tussupov@mail.ru, [6]aigerian@mail.ru, [7]kuntu@inbox.ru

ABSTRACT

The aim of the work is to create a knowledge base containing information on paraphrased sentences, including containing temporal and spatial concepts found in natural language texts. The content of the base: the most important concepts from the explanatory dictionary S.I. Ozhegov; rephrased sentences; the results of the analysis of vocabulary articles (interpretations) and examples from a literature by means of Link Grammar Parser and DIALING software systems, etc. Results of the work can be used in intelligent information retrieval systems. The diagrams obtained at the output of Link Grammar Parser and DIALING systems are extremely interesting material for further research. It is advisable to investigate the possibility of using a number of constructions and concepts of mathematical logic in computer linguistics, such as: the construction of L. Henkin, realizability and omitting of types, model completeness, forcing, as well as a number of non-classical logics. Researchers are executed at financial support of the Ministry of Education and Science of Republic Kazakhstan (the grant 2018-2020 MES RK No. AP05133550, the grant 2018-2020 MES RK No. AP 05133546), the Integration project of the Siberian Branch of the Russian Academy of Science (AAAA-A18-118022190008-8) and the study was partially supported by Russian Foundation for Basic Research (the project N 19-07-01134).

***Keywords*:** *computational linguistics, semantics, temporal and spatial concepts, Link Grammar Parser, DIALING.*

1. INTRODUCTION

The main purpose consists in investigating constructions in a natural language, containing different concepts, in the semantic plan.

results of the analysis of entries (interpretation) and use examples in fiction of corresponding concepts from S.I. Ozhegov dictionary by means of the program systems Link Grammar Parser and DIALING [3].

The knowledge base containing the information on on paraphrased sentences, including containing temporal and spatial concepts is created. The knowledge base contains: the most important concepts from S.I. Ozhegov explanatory dictionary [10]; the paraphrased variants of some sentences;

Further it is supposed to study the diagrams received on an exit of systems Link Grammar Parser and Dialing by means of the mathematical logic. It is a question of research of possibilities of application in computer linguistics of some

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

constructions and concepts of mathematical logic, such as: Henkin construction, realization and omitting of types, model completeness, forcing, and also some non-classical logics.

It is supposed to carry out the analysis by means of mathematical logic of properties of lexical functions of I.A. Melchuk [1] in a context of a definition of a sense and the set-theoretical models of language offered by S. Markus.

In the future the work will include also researches of combinatory properties of linguistic definitions and constructions, possibilities of application in computer linguistics of concepts from the set-theoretical and algebraic topology and researches of geographical concepts.

Results of the work can be used in intellectual systems of information search. Namely, for a definition of relevance of a text to a search request and for a definition of themes of texts. It also is of interest for scientists working in theoretical linguistics.

2. FORMAL METHODS OF RESEARCH OF SEMANTICS OF TEXTS

Semantics is the section of linguistics studying a semantic value of units of language: separate words, word-combinations, sentences, text fragments. At present there is a number of the machine-oriented methods of a representation of a sense of sentences [1–9].

For example, I.A. Melchuk has entered a concept of lexical functions, developed concepts of syntactic and semantic valences and considered them in a context of the sensible-combinatory dictionary [1]. V.S. Rubashkin and D.G .Lakhuti [5, 6] have entered an hierarchy of syntactic links for more effective work of the semantic analyzer. I.A. Melchuk's approach is supported in the program system DIALING [3].

There was appeared a notion of the universal language for knowledge representation. It can be a convenient tool to obtain new knowledge from already available. Quite probably in the future, researches will develop in a direction of a creation of such semantic languages. For example, now the system called Knowledge Vault contains 1,6 billion facts. System NELL developed within the frameworks of the project ReadTheWeb by Carnegie-Mellon University contains more than 50 million of the statements in addition characterized by various degrees of trust.

One more approach is the use of the system Link Grammar Parser [11–13] developed at Carnegie-Mellon University, based on some special syntax theory. We can notice that the given theory, generally speaking, differs from the classical theory of the syntax. Having the received sentence, the system attributes to it some syntactic structure which consists of set of the marked connectors (links) connecting words.

Received diagrams, as a matter of fact, are analogues to the so-called trees of submission of sentences. In trees of submission, it is possible to ask a question from the main word in the sentence to the minor. Thus, words are built in a treelike structure. The main reason on which the analyzer is called a semantic system, it is possible to mention the unique on completeness set of links. There is available about 100 basic links, thus some of them in addition have 3-4 variants.

Our researches are carried out according to following points.

1. The choice of the most important concepts concerning by time and space from S.I. Ozhegov explanatory dictionary.
2. Creation of a set of the paraphrased variants of various sentences and methods of estimating their similarities.
3. The analysis of entries (interpretation) and using examples in fiction containing corresponding concepts from S.I. Ozhegov dictionary by means of the program systems Link Grammar Parser and DIALING.
4. The analysis of the diagrams received on an exit of systems Link Grammar Parser and DIALING by means of the mathematical logic. The research of possibilities of an application to the analysis of geographical concepts.
5. Integration of results of researches into the knowledge base.

3. PROGRAM SYSTEM LINK GRAMMAR PARSERLINK

Grammar Parser is a syntactic analyzer of English language developed in 1990th at the Carnegie Mellon University, USA, based on some theory. We notice that in general the given theory differs from the classical theory of syntax. Having received a sentence, the system attributes it with a syntactic structure, which consists of a set of the marked links connecting the pairs of words. The detailed description of the system can be found in [11–13]

Link Grammar Parser includes about 60000 dictionary forms. He allows us to analyze a huge part of syntactic constructions, including numerous rare expressions and idioms. The parser work is stable; it can skip a part of a sentence, which it cannot understand and define some structure for the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

rest part of a sentence. It is capable to process an unknown lexicon, and do reasonable assumptions about the syntactic category of unknown words from the context and writing. It has data about various names, numerical expressions, and various punctuation marks.

Rules of connecting words are described in a set of dictionaries. For each word in the dictionary there is fixed, what connectors it can be connected with other words in a sentence. The connector has a name, with which considered unit (word) can enter in a sentence. For example, the mark S corresponds to communication between a subject and a predicate, O is a connector between an object and a predicate. Only the basic most important connectors, there are more than hundred. For a designation of a direction of a connector, the sign "+"is used to indicate a right connector, and the sign "–" to indicate a left connector. Left-directed and right-directed connectors of the same type make up a connection (link).

For example, if to word W1 it is attributed a connector And +, and to word W2 – connector A – in the syntactic sentence structure, consisting of two words W1 W2, then the link A will be establish between words W1 and W2. The sentence W2 W1 will not receive any interpretation as to W2 is attributed connector p A–, which forms a link only to the left, and to word W1 is attributed And + which forms a link only to the right. We will notice that there can be some variants of analysis of the same sentence.

Received diagrams, as a matter of fact, are analogues to the so-called trees of submission of sentences. In trees of submission, it is possible to raise a question from the main word in the sentence to the minor one. Thus, words are built in some treelike structure. The syntactic analyzer can give out two or more schemes of analysis of the same sentence. This phenomenon is called as a syntactic synonymy.

The main reason why the analyzer is named by a semantic system, it is the unique (taking into account a completeness) set of connectors (about 100 basic ones, and some of them have 3-4 variants). In some cases, a careful work on different contexts has led authors of the system to a transition to almost semantic classifications constructed exclusively on syntactic principles.

For example, following classes of English adverbs are allocated: situational adverbs which concern to all sentence in whole (clausal adverb); time adverbs; introduction adverbs which stand in the beginning of the sentence and are separated by a comma (openers); the adverbs modifying adjectives etc.

From advantages of the system, it is necessary to notice, that the organization of the procedure of a finding of variants of the syntactic representation is very effective. The construction goes not from top to down (top-down) and not from below upwards (bottom-up), but all hypotheses about relations are considered in parallel: at the beginning all possible connections by dictionary formulas are constructed, and then possible subsets of these communications are allocated.

Of course, it leads to some algorithmic opacity of the system, because it is very difficult to track all relations at once. Secondly, it leads to a nonlinear dependence of a speed of the algorithm depending on a number of words, but to exponential one, because a set of all variants of syntactic structures on the sentence containing N words at worst equivalent by a cardinality to a set of all spanning trees of the full graph with N nodes.

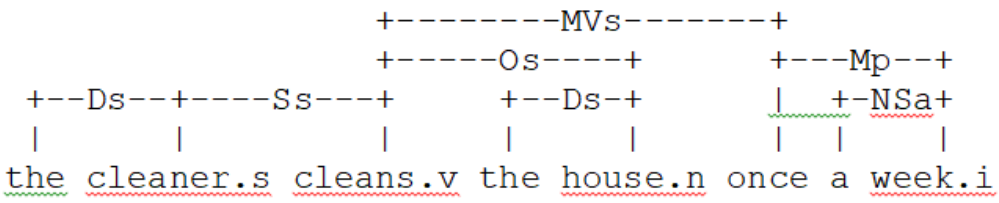
The last feature of the algorithm forces developers to use the timer to stop the procedure, which works too long. However all these lacks are compensated by a linguistic transparency of the system in which rather simple valences of words may be registered, and the order of gathering of valences in an algorithm essentially is not fixed, i.e. connections are constructed as though in parallel, that completely corresponds to our language intuition.

Let's note also the negative moments.

1. The practical testing of the system shows, that at the analysis of complicated sentences, which length exceeds 25-30 words, a combinatory explosion is possible, and a result of a work of the analyzer becomes the "panic" graph, as a rule having a casual variant of a syntactic structure, which is inadequate from the linguistic point of view.

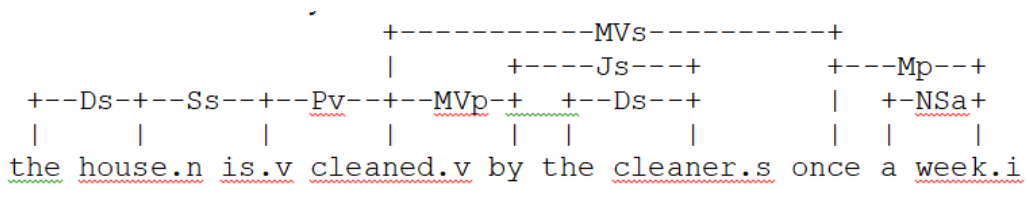
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

2. Application of ideas described above is complicated for inflective languages as Russian, in view of considerably increasing volume of dictionaries, which arises because a morphological development of inflective languages. Each morphological form should be described by the separate formula where the bottom index of a name of a connector should provide a coordination procedure. It leads to increasing a number of connectors. For agglutinative languages (for example, Turkic), the system becomes even more difficult.



2. The house is cleaned by the cleaner once a week.

$Ss(cleaner, cleans) \wedge Os(cleans, house) \leftrightarrow$
$\leftrightarrow Ss(house, is) \wedge Pv(is, cleaned) \wedge$
$\wedge MVp(cleaned, by) \wedge Js(by, cleaner)$



Below examples of parsing two sentences are represented.

1. The cleaner cleans the house once a week.

4. PARAPHRASED SENTENCES

It is known, that natural languages possess variety of ways of a sense expression, it is possible to transfer the same thought by different words. This feature considerably complicates the analysis of texts in a natural language, and the problem of a similar sense detection in different statements is difficultly formalized. Formalization of the analysis of time and spatial concepts of texts allows us to come nearer to the decision of this problem partially at least. In the given section, a process of detection of the paraphrased sentences containing some spatial and time concepts is formally described.

In some cases in a database it is expedient to have the paraphrased variants of sentences and methods of estimating their affinity, or, in other words, similarities. It is especially important for sentences containing the casual verbs of movements and changes of a position in a space, in view of complexity of sentences containing them. The verbs enter into this class, expressing such concepts, as: move, approach, delete, bear, give, take, put, lift, lower, throw, catch, send, etc. Certain interest represents the sentences containing some adverbs: ahead, behind, sideways, earlier, later, still, already; from the point of view of their rephrasing.

In [14–21] it is given the description how it is possible to compare the paraphrased sentences, for a case of using the analyzer Link Grammar Parser. We will assume, that L is a set of words of some natural language. For any word $x \in L$ we will designate $Norm(x)$ its normalized form. The record $Syn(x, y)$ designates, that x, y - synonyms.

There are two forms of the equivalence:

1) $x_1 \approx x_2 \leftrightarrow x_1 = x_2 \vee Syn(x_1, x_2)$

2) $x_1 \equiv x_2 \leftrightarrow Norm(x_1) = Norm(x_2)$.

The sentence may be considered as a vector with components, which are words $\bar{x} =< x_1, \ldots, x_n >$. The function $Norm$ can be naturally extended onto sentences $Norm(\bar{x}) =< Norm(x_1), \ldots, Norm(x_n) >$.

The text $T =< \bar{x}_1, \ldots, \bar{x}_n >$ is the sequence of sentences.

Let the formula $\bar{x} \models P(x_i, x_j)$ means that in the scheme of analysis of the sentence $\bar{x} =< x_1, \ldots, x_n >$ by means of the analyzer Link

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Grammar Parser, there is a connector of a type P going from a word x_i to a word x_j. The sign $|=$ means that actually we consider a model. The basic set of the model is the set of pairs $\{<1,x_1>,\ldots,<n,x_n>\}$. Because the same word can enter into the sentence two and more times, it leads to a necessity of a consideration of pairs, instead of separate words. In view of told above, the designation $\bar{x} |= \varphi$, where φ is a formula, for example, of the first order logic, is correct. Actually, simultaneously $\bar{x}$ is a designation both for a vector and for a model.

Let us assume that two sentences are given. $\bar{x} =< x_1,\ldots,x_n > \bar{y} =< y_1,\ldots,y_m >$. It is interesting to consider functions f such that $dom(f) \subseteq \{1,\ldots,n\},\ range(f) \subseteq \{1,\ldots,m\}$ with additional properties of a form: $f(i) = j \to x_i \approx y_j$ and $f(i) = j \to x_i \equiv y_j$, and other similar.

At a comparison of two sentences, more exact, at their nearness analysis, a verification of some logic properties is carried out. For example, let $f(i_1) = j_1,\ f(i_2) = j_2$. Examples of such properties are given below.

1. The invariance of a connector
$\bar{x} |= P(x_{i_1}, x_{i_2}) \to \bar{y} |= P(y_{j_1}, y_{j_2})$.
2. The replacement of a connector by a disjunction of others
$\bar{x} |= P(x_{i_1}, x_{i_2}) \to \bar{y} |= \bigvee_t Q_t(y_{j_1}, y_{j_2})$.
3. The splitting of connector on two connectors
$\bar{x} |= P(x_{i_1}, x_{i_2}) \to \exists k\ (\bar{y} |= Q(y_{j_1}, y_k) \wedge R(y_k, y_{j_2}))$
.
4. The splitting of connector on two connectors with an inversion
$\bar{x} |= P(x_{i_1}, x_{i_2}) \to \exists k\ (\bar{y} |= Q(y_{j_2}, y_k) \wedge R(y_k, y_{j_1}))$

Taking into consideration that $\bar{y}$ is a designation for a corresponding model, the formula from the third point can be rewritten in the form
$\bar{x} |= P(x_{i_1}, x_{i_2}) \to \bar{y} |= \exists y Q(y_{j_1}, y) \wedge R(y, y_{j_2})$

Analogously the formula from the fourth point can be written in a similar form.

Summarizing it is possible to say that there are rules of the form
$R_i : \bar{x} |= \varphi_i(x_1, x_2) \to \bar{y} |= \psi_i(y_1, y_2)$.

Let us notice that for English language, we fixed more than thirty of such rules. For Russian and other languages, this question is less studied. For system DIALING, also may by formulated similar rules, but it is more a complicated question.

Further a function f is constructed and the analysis is carried out, whether there are indexes $i_1, i_2,\ j_1 = f(i_1),\ \ j_2 = f(i_2)$ such that the rule R_i is fulfilled on concrete words from sentences $\bar{x}, \bar{y}$, i.e. $\bar{x} |= \varphi_i(x_{i_1}, x_{i_2}) \to \bar{y} |= \psi_i(y_{j_1}, y_{j_2})$. For simplicity it is possible to say, that the rule is fulfilled on a pair $< i_1, i_2 >$.

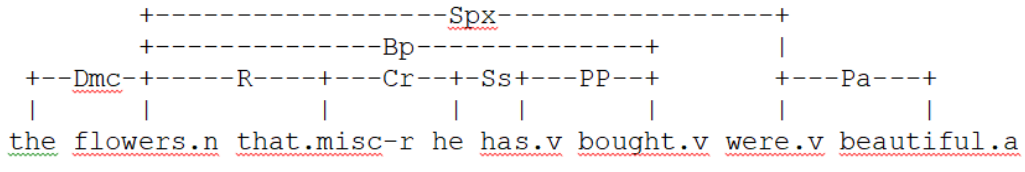
Let us consider a set of all such pairs $< i_1, i_2 >$ on which one of rules is fulfilled. We will designate this set by I and let its cardinality $|I| = n$. Let us notice that the analyzer Link Grammar Parser assumes a presence of only one connector between two words. Therefore no more than one rule will be fulfilled.

Analogously as earlier, let n_1, n_2 be number of connectors obtained as a result of the analysis of sentences $\bar{x}, \bar{y}$ respectively. As a measure of similarity of two offers, it is possible to enter $\mu_0(\bar{x}, \bar{y}) = n / \max(n_1, n_2)$ or $\mu_1(\bar{x}, \bar{y}) = 2n/(n_1 + n_2)$.

5. Examples of paraphrased sentences in English language

1. Complicating the expanded definition.

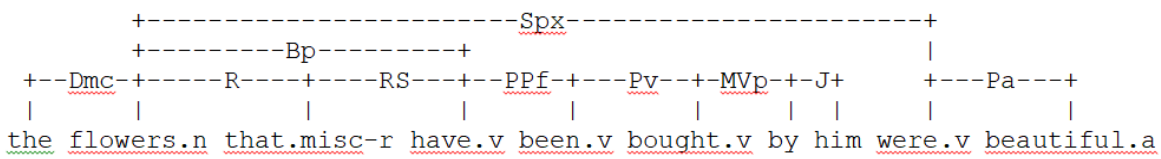
1.1. The flowers that he has bought were beautiful.



Journal of Theoretical and Applied Information Technology
31st May 2020. Vol.98. No 10

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

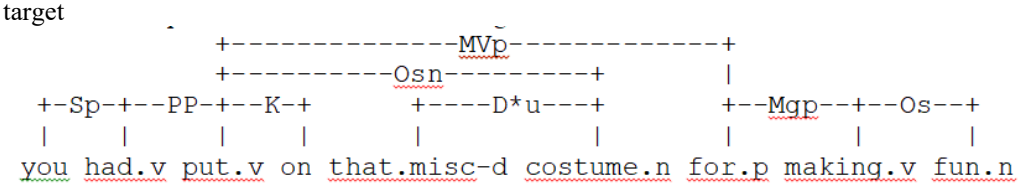
1.2. The flowers that have been bought by him were beautiful.



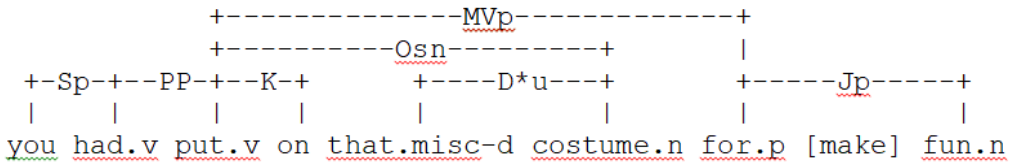
Cr(*that, he*) ∧ *Ss*(*he, has*) ∧ *PP*(*has, bought*) ↔
↔ *RS*(*that*,*have*) ∧ *PPf*(*have*, *been*) ∧
∧ *Pv*(*been*, *bought*) ∧ *MVp*(*bought*,*by*) ∧ *J*(*by*,*him*).

2. A synonymic series expressing target relationships (*MVp* + ((*Mgp* + *Os*) *or Js*), etc.

2.1. You had put on that costume for making fun.

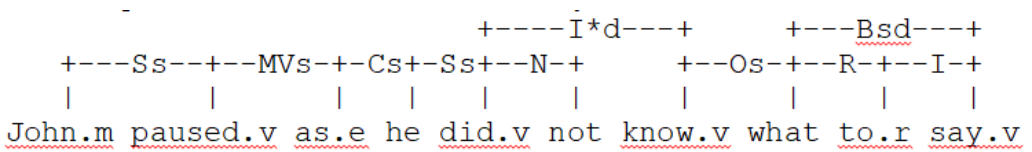


2.2. You had put on that costume for make fun.



Mgp(*for, making*) ∧ *Os*(*making*, *fun*) ↔
↔ *Jp*(*for*, *fun*).

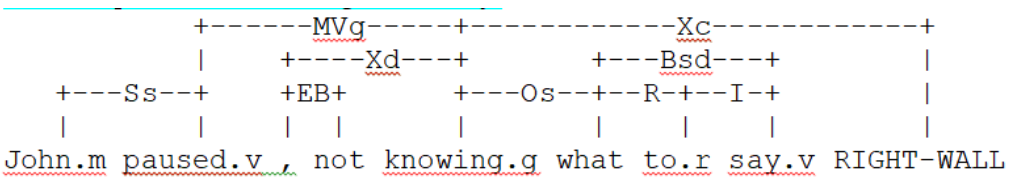
2.3. You had put on that costume so that you could make fun.



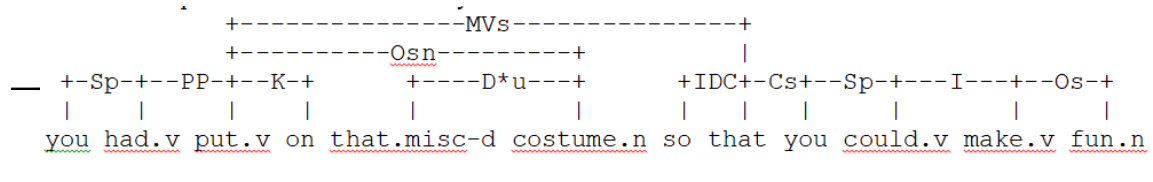
Mgp(*for*, *making*) ↔ *IDC* (*so*, *that*) ∧
∧ *Cs*(*that*, *you*) ∧ *Sp*(*you*, *could*) ∧ *I*(*could*, *make*).

3. A synonymous series expressing a causal relationship (what to say).

3.1. John paused, not knowing what to say.



3.2. John paused as he did not know what to say.
MVg(*paused*, *knowing*) ∧ *Xd*(*not*, *knowing*) ∧
∧ *EB* ↔ *MVs*(*paused*, *as*) ∧ *Cs*(*as*, *he*) ∧
∧ *Ss*(*he*, *did*) ∧ *N*(*did*,*not*) ∧ *I*d*(*did*,*know*).

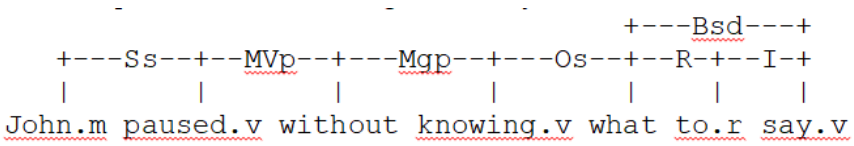


ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

From a comparison of the second and third sentences, we obtain, respectively.
$Jp(for, fun) \leftrightarrow IDC\ (so, that) \wedge Cs(that, you) \wedge$
$\wedge Sp(you, could) \wedge I(could, make) \wedge Os(make, fun)$.

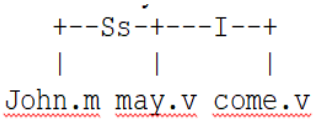
3.3. John paused without knowing what to say.
$MVg(paused, knowing) \wedge Xd(not, knowing) \wedge$
$\wedge EB \leftrightarrow MVp(paused, without) \wedge$
$\wedge Mgp(without, knowing)$.

Or from a comparison of the second and third sentences:
$MVs(paused, as) \wedge Cs(as, he) \wedge Ss(he, did) \wedge$
$\wedge N(did, not) \wedge I^*d(did, know) \leftrightarrow$
$\leftrightarrow MVp(paused, without) \wedge$
$\wedge Mgp(without, knowing)$.

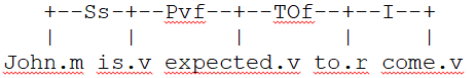


4. There are three types of predicate complication depending on the nature of the complicating element: active-verbal complication; passive-verbal complication; adjective complication. The connection Pa is an important sign of adjective complication. Basically, complication occurs in the predicate, i.e. verbal phrases and additions. The complicating element takes on the function of expressing communication with the subject. Further an example of an active-verbal complication of the predicate is represented. The complicator transmits a modal characteristic of the connection of actions with the subject. TOf and MVi connectors can be used interchangeably.

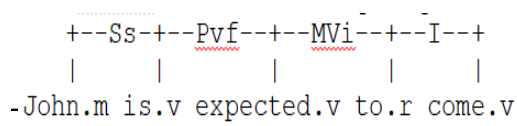
4.1. John may come.



4.2a. John is expected to come.

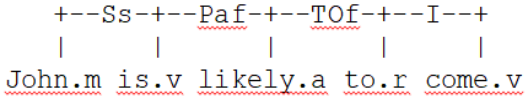


4.2b. Another, also true version of parsing this sentence.
$Ss(John, may) \wedge I(may, come) \leftrightarrow$
$\leftrightarrow Ss(\text{John, is}) \wedge Pvf(is, expected) \wedge$
$\wedge TOf(expected, to) \wedge \text{I}(to, come)$.

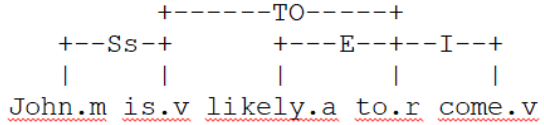


$Ss(John, may) \wedge I(may, come) \leftrightarrow$
$\leftrightarrow Ss(John, s) \wedge Pvf(is, expected) \wedge$
$\wedge MVi(expected, to) \wedge \text{I}(to, come)$.

4.3a. John is likely to come.



4.3b. Another, also true version of parsing this



sentence.

$Ss(John, may) \wedge I(may, come) \leftrightarrow$
$\leftrightarrow Ss(John, is) \wedge Paf(is, likely) \wedge TOf(likely, to) \wedge$
$\wedge I(to, come)$.
From a comparison of the second and third sentences, we obtain, respectively.
$Pvf(is, expected) \wedge TOf(expected, to) \leftrightarrow$
$\leftrightarrow Paf(is, likely) \wedge TOf(likely, to)$.

From a comparison of the second and third sentences, we see that the connectors Pvf and Paf can be interchangeably.
For another version of this sentence, we obtain the following formula.

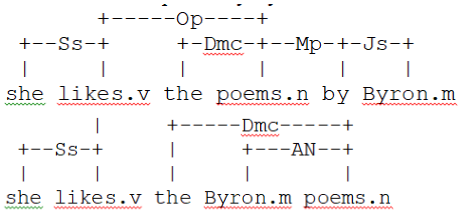
$Ss(John, may) \wedge I(may, come) \leftrightarrow Ss(John, is) \wedge$
$\wedge TO(is, to) \wedge E(to, likely) \wedge I(to, come)$.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

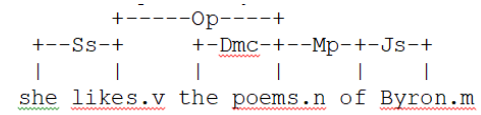
Pvf(*is*, *expected*) ∧ *TOf*(*expected*, *to*) ∧
∧ *I*(*to*, *come*) ↔ *TO*(*is*, *to*) ∧ *E*(*to*, *likely*) ∧
∧ *I*(*to*, *come*).

5. A synonymous series expressing attributive relationships: by Byron, of Byron; displayed through a combination of Mp + Js bonds.

5.1. She likes the Byron poems.



5.2. She likes the poems by Byron.

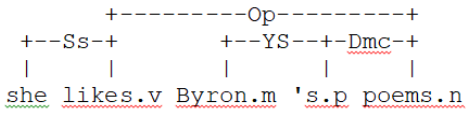


We get the same scheme for another grammatically correct sentence.

She likes the poems of Byron.

AN(*Byron*, *poems*) ↔ *Mp*(*poems*, *by*) ∧
∧ *Js*(*by*, *Byron*).
AN(*Byron*, *poems*) ↔ *Mp*(*poems*, *of*) ∧
∧ Js(*of*, *Byron*).

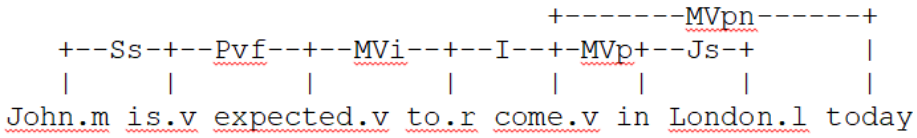
5.3. She likes Byrons poems.
AN(*Byron*, *poems*) ↔ *YS*(*Byron*, '*s*) ∧
∧ *Dmc*('*s*, *poems*), or from a comparison of the second and third sentences
Mp(*poems*, *by*) ∧ *Js*(*by*, *Byron*) ↔
↔ *YS*(*Byron*, '*s*) ∧ *Dmc*('*s*, *poems*).



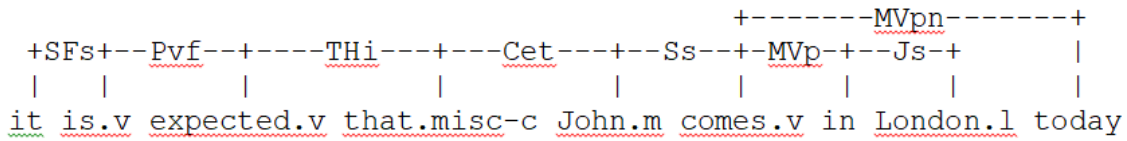
6. Passive-verbal complication of the predicate in the form of a passive voice (Pvf connection). The TOf and MVi connectors are interchangeable in some cases. This type of complication carries a hue of modal meaning, i.e. the speaker, as it were, relieves himself of responsibility for the reliability of the reported fact, the real action. This is confirmed by the use of nominalization transformations, i.e. the transition to the category of nouns of other parts of speech in these sentences. For example: to come → coming.

6.1a. John is expected to come in London today.

6.1b. Another, also true variant of parsing.

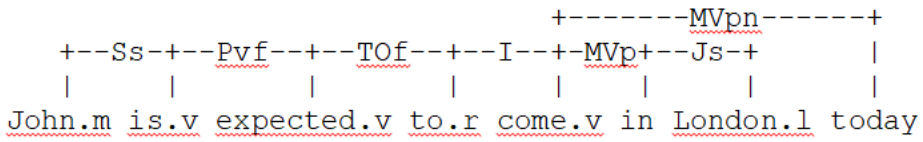


6.2. It is expected that John comes in London today.



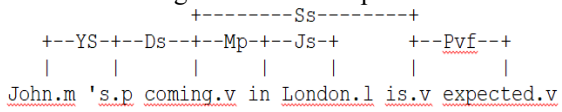
Journal of Theoretical and Applied Information Technology
31st May 2020. Vol.98. No 10

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195



Ss(*John*, *is*) ∧ *Pvf*(*is*, *expected*) ∧
∧ *TOf*(*expected*, *to*) ∧ *I*(*to*, *come*) ↔
↔ *SFs*(*it*, *is*) ∧ *Pvf*(*is*, *expected*) ∧
∧ *THi*(*expected*, *that*) ∧ *Cet*(*that*, *John*) ∧
∧ *Ss*(*John*, *comes*).

6.3. John's coming in London is expected.



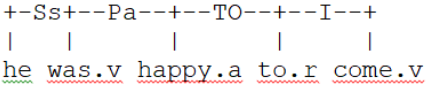
Ss(*John*, *is*) ∧ *Pvf*(*is*, *expected*) ∧
∧ *TOf*(*expected*, *to*) ∧ *I*(*to*, *come*) ↔
↔ *YS*(*John*, '*s*) ∧ *Ds*('*s*, *coming*) ∧
∧ *Ss*(*coming*, *is*) ∧ *Pvf*(*is*, *expected*) ∧
∧ *Mp*(*coming*, *in*) ∧ *Js*(*in*, *London*).

или (из сравнения второго и третьего предложений)

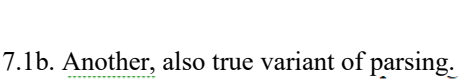
SFs(*it*, *is*) ∧ *Pvf*(*is*, *expected*) ∧
∧ *THi*(*expected*, *that*) ∧ *Cet*(*that*, *John*) ∧
∧ *Ss*(*John*, *comes*) ↔ *YS*(*John*, '*s*) ∧
∧ *Ds*('*s*, *coming*) ∧ *Ss*(*coming*, *is*) ∧
∧ *Pvf*(*is*, *expected*) ∧ *Mp*(*coming*, *in*) ∧
∧ *Js*(*in*, *London*).

7. Adjective complication of the predicate. The complicator element refers to the physical, mental, or other characteristic of the subject. For example: happy (link Pa) associated with the action that is indicated by the subsequent infinitive (link I). Adjectives, participles, and words of the category of state are used as complicators.

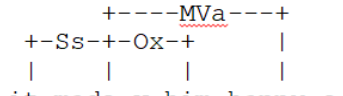
7.1a. He was happy to come.



7.1b. Another, also true variant of parsing.

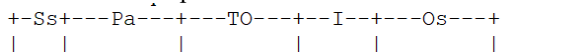


7.2. It made him happy.

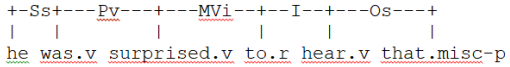


Ss(*he*, *was*) ∧ *Pa*(*was*, *happy*) ∧ *TO*(*happy*, *to*) ∧
∧ *I*(*to*, *come*) ↔ *Ss*(*it*, *made*) ∧
∧ *MVa*(*made*, *happy*) ∧ *Ox*(*made*, *him*).

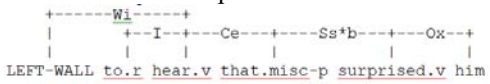
8. Adjective complication of the predicate.
8.1a. He was surprised to hear that.



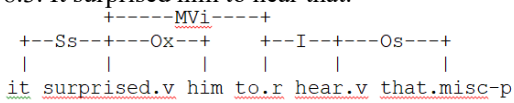
8.1b. Another, also true variant of parsing.



8.2. To hear that surprised him.

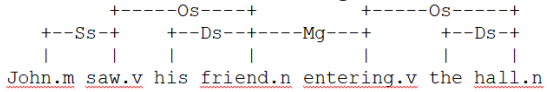


8.3. It surprised him to hear that.



Ss(*he*, *was*) ∧ *Pa*(*was*, *surprised*) ∧
∧ *TO*(*surprised*, *to*) ↔ *Ss*(*it*, *surprised*) ∧
∧ *MVi*(*surprised*, *to*) ∧ *Ox*(*surprised*, *him*).

9. Complication of direct object complementation (Os) is possible after verbs and is achieved by joining the infinitive, participle, adjective, state category word, prepositional group to a noun or pronoun, etc. as a supplement. A distinctive feature of the constructions under consideration is the presence of a semi-predictive relation between direct complement and its complication (Os + Mg).
9.1. John saw his friend entering the hall.

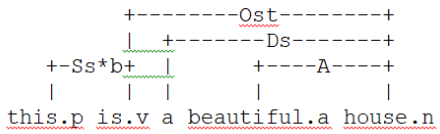


9.2. John saw that his friend entered the hall.

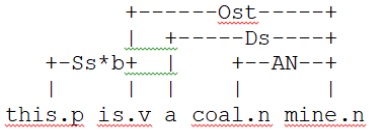
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Os(*saw*, *friend*) ∧ Mg(*friend*, *entering*) ↔
↔ TH(*saw*, *that*) ∧ *Cet*(*that*, *friend*) ∧
∧ *Ss*(*friend*, *entered*).

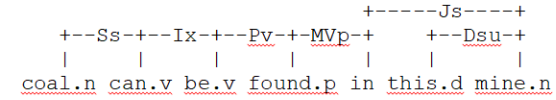
10. An example of a simple sentence modification.
10.1. This is a beautiful house.



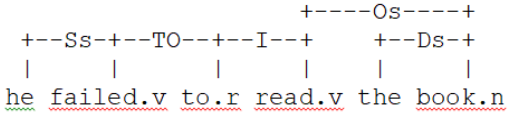
11. The word "Coal" is different parts of speech: in the first case an adjective, in the second case a noun. The relationship between AN and SS is a sign that coal is playing differently roles.
11.1. This is a coal mine.



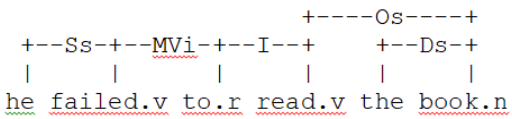
11.2. Coal can be found in this mine.



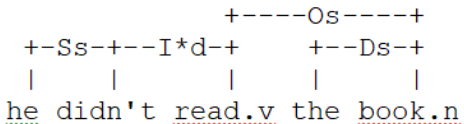
12. Active verb complication. The complicator conveys the reality of the action or denies the reality of the action. In this case, the main element of the predicate takes the form of communion.
12.1a. He failed to read the book.



12.1b. Another, also true variant of parsing

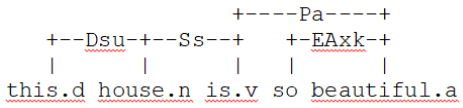


12.2. He didn't read the book.



Ss(*he*, *failed*) ∧ *MVi*(*failed*, *to*) ∧
∧ *I*(*to*, *read*) ↔ *Ss*(*he*, *didn't*) ∧ *I*d*(*didn't*, *read*).

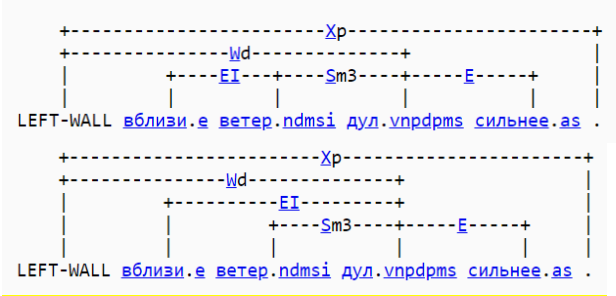
10.2. This house is so beautiful.

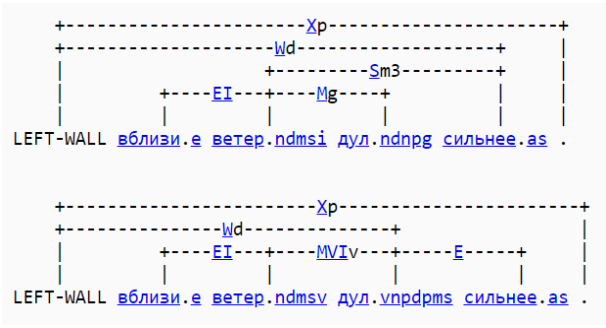


*Ss*b*(*this*, *is*) ∧ *Ost*(*is*, *house*) ∧
∧ *A*(*house*, *beautiful*) ↔ *Ss*(*house*, *is*) ∧
∧ *Pa*(*is*, *beautiful*) ∧ *EAxk*(*beautiful*, *so*).

5. EXAMPLES OF PARAPHRASED SENTENCES IN RUSSIAN LANGUAGE

Let us consider the sentence «Вблизи ветер дул сильнее» (*Near the wind blew stronger*).
There are 4 variants of parsing this sentence by means of Link Grammar Parser





The following notations are accepted here.

Xp – the link for connecting the beginning of a sentence with a point (at the end of a sentence).

Wd – the link points to the top of the sentence.

EI – connection with an adverb denoting a spatial concept.

Sm3 – connects the subject with the verb (lowercase letters supplement information about gender and person).

E – the connection between the verb and the adverb modifying it.

Mg – the connection between a noun and a participle.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

MVIv – combines the predicate with the complement.

We see that the Xp connection is present in all schemes. We can say that it is invariant. The word *вблизи* (near), related to spatial concepts, is considered as a constant in our signature. Regarding other non-invariant relations, using logical formulas, we can write down the existing equivalences.

$$x_1 = ветер$$
$$x_2 = дул$$
$$\varphi_1 = EI(вблизи, x_1) \wedge Sm3(x_1, x_2)$$
$$\varphi_2 = EI(вблизи, x_2) \wedge Sm3(x_1, x_2)$$
$$\varphi_3 = EI(вблизи, x_1) \wedge Mg(x_1, x_2)$$
$$\varphi_4 = EI(вблизи, x_1) \wedge MVIv(x_1, x_2)$$
$$\varphi_1 \leftrightarrow \varphi_2 \leftrightarrow \varphi_3 \leftrightarrow \varphi_4.$$

DIALING system. The DIALING system was developed as a system of Russian-English translation from 1999 to 2002 on the basis of Dialing LLC (Moscow) [3]. At different times, more than 20 specialists took part in the work on the system, most of which were well-known linguistic scientists. Like all modern word processing systems, DIALING includes all the basic stages of text analysis.

Let us consider some examples.

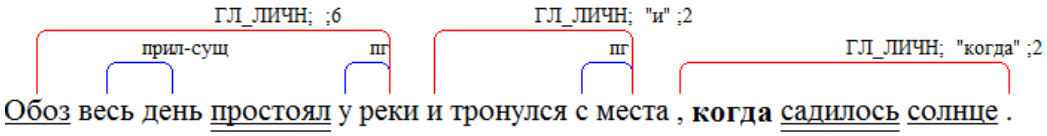
1. Обоз весь день простоял у реки и тронулся с места, когда садилось солнце.

The convoy stood all day by the river and set off when the sun was setting.

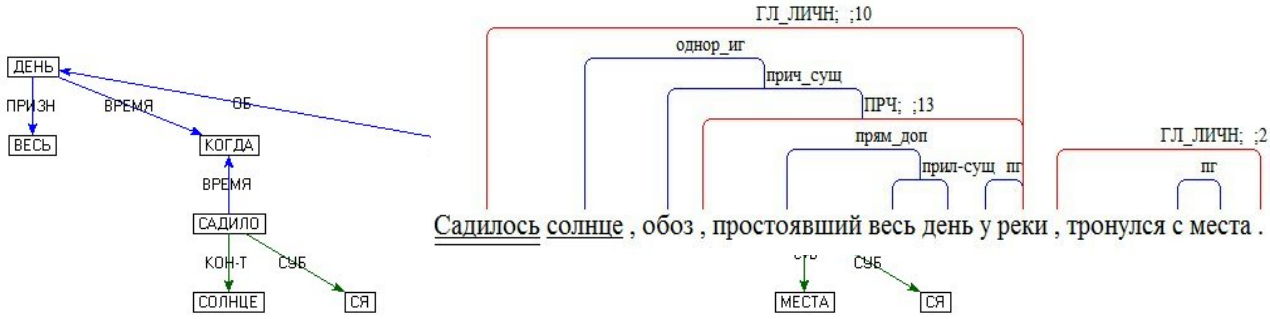
2. Садилось солнце, обоз, простоявший весь день у реки, тронулся с места.

The sun was setting, the convoy, which stood all day by the river, began to move

1.1. Syntactic parsing result



1.2. Semantic analysis result

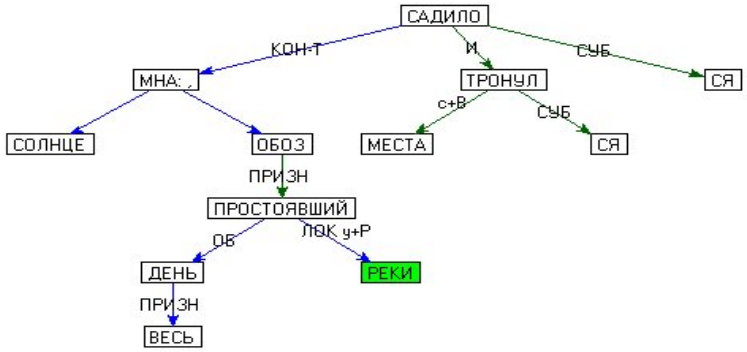


6. THE DIALING SYSTEM

The ideas discussed in the previous sections can be applied to analyze re-phrased sentences, namely, to evaluate their proximity, also based on the diagrams obtained at the output of the

2.1. Syntactic parsing result.
2.2. Semantic analysis result.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195



7. CONCLUSION

In the work, a process of creation of the knowledge base containing the information on time and spatial concepts, appearing in texts in a natural language is described. The process of detection of the paraphrased sentences containing some spatial and time concepts is formally described. Now the work of filling the knowledge base is conducted. As a result of the carried out research the conclusion has been made about the expediency of application of such tools as Link Grammar Parser and DIALING for solving the considering task.

Link Grammar Parser is an unusual enough system, the main reason on which the analyzer names a semantic system, that it is possible to consider unique on completeness set of links. For example, following classes of adverbs are allocated: situational adverbs which concern all sentence in the whole (clausal adverb); time adverbs; introduction adverbs which stand in the beginning of the sentence and are separated by a comma (openers); the adverbs modifying adjectives, etc. From advantages of the system it is necessary to notice, that the organization of the procedure of a finding variants of syntactic representation is very effective.

Let's note also the negative moments. Practical testing the system shows, that at the analysis of the complicated sentences which length exceeds 25-30 words, a combinatory explosion is possible, and the "panic" graph, as a rule, a casual variant of syntactic structure becomes result of work of the analyzer, from the linguistic point of view inadequate. The use of the system is complicated for inflectional languages such as Russian, in view of increasing volumes of dictionaries which arise owing to morphological development of inflectional languages.

For a case of English language a testing [15] has been held. The main goal is a definition of relevance of the text to the search request. In this case, the minimum variant yielding good enough results has been found out, when only 8 links were considered. For English language, the list of the most important links of system Link Grammar Parser has been determined. It contains 35 links. Links were found out, which essentially spoilt a situation in case of incomplete sentences. There are 8 such links. For Russian such detailed analysis of links is not carried out. Possibilities of the system DIALING are even less studied.

Nevertheless, it is necessary to notice, that the diagrams received on an exit of systems Link Grammar Parser and DIALING represent extremely interesting material for the further researches; including for researches of time and spatial concepts; both from the point of a view of applications and for theoretical linguistics.

REFERENCES

[1] I.A. Melchuk "Experience of Theory of Linguistic Models like Sence ↔ Text", *Moscow*: *Nauka*, 1974.

[2] U.D. Apresyan "Experimental Semantic Investigation of a Russian Verb", *Moscow*: *Nauka*, 1967.

[3] A.V. Sokirko "Semantic dictionaries in automatic word processing", *Cand. Diss. (PhD Thesis), MGPII, Moscow*. 2000.

[4] E.V. Paducheva "Dynamic models in semantics of lexic", *Moscow: Languages of Slavic culture*, 2004.

[5] D.G. Lakhuti "The automatic analysis of texts in a natural language", *Information Processes and Systems,* Vol 11, 2003, pp. 18-25.

[6] V.Sh. Rubashkin "Meaning Representation and Analysis in Intelligent Information Systems", *Moscow*: *Nauka*, 1989.

[7] S. Markus "Set-theoretical Models of Languages", *Moscow*: *Nauka*, 1970.

[8] T. Batura and F. Murzin "Logical Methods for Representing Meaning of Natural Language

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Texts", *Proceedings ICCS,* Vol. 3038, 2004, pp. 545-551.

[9] T.V. Batura and F.A. Murzin "The Machine-Oriented Logic Methods of Representation of Semantics of the Text in Natural Language", *Novosibirsk*: *Publishing Company of NGTU*, 2008.

[10] S.I. Ozhegov "Explanatory dictionary of the Russian language", *Moscow*: *Mir i obrasovaniye*, 2014.

[11] D. Temperley, D. Sleator and J. Lafferty "Link Grammar Documentation", 1998, http://www.link.cs.cmu.edu/link/dict/index.html

[12] "Link Grammar Documentation", 2015, http://www.abisource.com/projects/link-grammar

[13] "The CMU Link Grammar natural language parser", 2018, https://github.com/opencog/link-grammar/

[14] T.V. Batura, F.A. Murzin, A.A. Perfliev and T.V. Shmanina "Methods of the Increase of the Efficiency of Information Search on the Basis of Syntactic Analysis", *Novosibirsk*: *Publishing Company of SB RAS*, 2014.

[15] T.V. Batura, A.M. Bakieva, A.S. Yerimbetova, F.A. Murzin and S.K. Sagnayeva "Link grammar, relevance and definition of themes of texts", *Novosibirsk*: *Publishing Company of SB RAS*, 2018.

[16] T.V. Batura, F.A. Murzin, D.F. Semich, A.M. Bakiyeva, A.S. Yerimbetova and S.K. Sagnayeva "Estimation of the Degree of Similarity of Sentences in a Natural Language Based on Using the Link Grammar Parser Program System", *Journal of Theoretical and Applied Information Technology,* Vol.86, 2016, Iss. 1, pp.68-77.

[17] A.S. Yerimbetova, S.K. Sagnayeva, F.A. Murzin and J.A. Tussupov "Creation of tools and algorithms for assessing the relevance of documents", *Proceedings RPC,* 2018, Electronic publication: 9781538675311. Paper № 8482202.

[18] F. Murzin, A. Yerimbetova, M. Tussupova and K. Aleksandrov "Development and analysis of technologies of searching information relevant to the search query using linguistic support", *Proceedings DAMDID/RCDL,* 2018, pp 207-214.

[19] S. Tazhibayeva and B. Yeskeldiyeva. "A comparison of equality in Kazakh and Turkic Languages of Siberia", *Mediterranean Journal of Social Sciences,* Vol.6, Iss. 4S2, 2015, pp. 398-408, http://www.mcser.org/journal/index.php/mjss

[20] S. Tazhibayeva and B. Yeskeldiyeva. "Multilingualism in modern Kazakhstan: New challenges", *Asian Social Science,* Vol. 11, Iss. 6, 2015, pp. 56-64.

[21] J. Tussupov, M.A. Sambetbayeva, A.M. Fedotov, O.A. Fedotova, S.K. Sagnayeva, A.A. Bapanov and S.Z. Tazhibaeva "Classification model and morphological analysis in multilingual scientific and educational information systems", *Journal of Theoretical and Applied Information Technology*, Vol.86, Iss. 1, N 10, 2016, pp.96-111.