Computer Science Journal of Moldova, vol.**2**, no.1(4), 1994

# Linguistic Automaton Today

M.Ignatieva V.Chizhakovsky A.Popescu
R.Piotrowski

**Abstract**

Over the last thirty years the long range goal of the research project of the Speech Statistics (Sp.St.) international group is to develop and implement in computer simulation psycholinguistically realistic model of language behavior. This model, named the linguistic automaton /LA/, is a complex of hardware, software lingware /linguistic means of support/ and CALL–ware. The Sp.St. group was developing many programs to model different aspects of natural language /NL/ processing by man. The multilevel and multilingual LA–architecture integrates in a flexible way these programs aand provides a control strategy capable of adapting itself to the requirements of a particular task. It permits a modular implementation of LA, making the modification of existing processes and the integration of new capacities in the LA–system easier.

## 1 Introduction

The Sp.St. group, a leading in NLP, through its branches in Russia /St.Petersburg/, Belorus /Minsk/, Moldova /Kishinev/, Central Asia /Chimkent, Samarcand/, has over the past thirty years done also research in Computational Linguistics /CL/ and Artifical Intelligence /AI/ /Tambovtsev, 1985, 1986/. Its industrial NLP–Systems satisfy the requirements of our present day information technology and are competitive on the world market.

What is it that helped Sp.St. to overcome the difficultes of the first few years and carry out successful machine experiments at the time

©1994 by M.Ignatieva, V.Chizhakovsky, A.Popescu, R.Piotrowski

when dozens of Soviet and Western teams failed to realise their widely publicized theories and promises to develop Machine Translation (MT) and other NLP–systems?

First of all it was the direct link Sp.St. had with the actual users of the informational technology being developed: the user financed and controled the research work and this stimulated the development of industrial linguistic product of high quality. Secondly, the correct understanding of the existing differences betweeen the thought /speech activity of man and the intellectual/ linguistic capabilities of the computer. Being aware of these differences, Sp.St. did concentrate its efforts towards development of an analogue of the thought/speech activity of man, e.g. towards creation of such a model that could be not only formulated and mathematically represented, but also, turned into a real and functioning LA. The main part of the present paper will be devoted to the theory and the method of creation of such a LA.

## 2 Presentation

The realisation of LA implies the combination of digital computer hardware, an operating computer program for NLP (software and lingware), and a vast linguistic database.

### 2.1 Theoretical Research in the sphere of text analysis and generation

The history of MT shows that it is impossible to create operative NLP without having theoretical explanation of the way the human mind functions in principle and without modeling its work on the computer. In this respect Sp.St. started in the early 60's to:

a) study the informational and the statistical nature of the text;

b) create semiotic models of text generation and understanding.

The results obtained from the statistical and the informational measurements of the text were of great practical importance as from the very start. Therefore we'll describe first the informational and statistical measurements of the text.

### 2.1.1 Informational Structure of the Text

The aim of the informational measurements was to estimate the distribution of the syntactical information in the text and in its constituent lexical units. It was envisaged to use the results to weigh the informational load of grammar in relation to that of lexical semantics in different languages and sublanguages, also to measure the semantical and pragmatic information carried by the wordforms, the phrases and by their constituent elements (morphemes, words).

A revised version of C.Shannons letter by text prediction method was used. This method was applied to Roumanian, Russian, English, German, French, Spanish, Polish, Estonian, Azerbaidzan (Turkish), Kazakh and Uzbek language texts, for all major styles (belles–letters, scientific and technical/administrative, and conversational) ([30]; [28], pp.320-362; [14]; [32], pp.151-213). The properties of the text, revealed by the data, prove to be of great importance for the construction and further developement of NLP. All above mentioned languages, irrespective of their typology, are 65 to 96 p.c. redundant, and therefore best suited for NLP are the so very explicit administrative and patent documents; second in line come all scientific/technical and social/political texts (they are about 80 p.c. redundant). Least redundant are belles–letters and conversation texts. From about 18 p.c. (for the agglutinative synthetic Turk languages), up to 35 p.c. (for the analytical English language) of the syntactical semantical information is concentrated in the contextual (semantico–syntactical) links (the auxiliary words). The rest of the information (inf.) (65 to 82 p.c.) is contained in the lexical units (wordforms, phrases). Taken out of context, the wordforms have a U shaped inf. structure: the bulk of inf. goes to the 'lexical' begining and to the morphological ending of the wordform; in most cases the middle of the wordforms is redundant.

When the wordform becomes a running word inf. scheme assumes an L shaped structure: the morphological surge of information towards the end of the wordform is gone. This phenomena is characteristic not only for the English and French, but also for the Roumanian, Russian and German wordforms ([32], pp.217-234). Parallel psycholiguistic tests showed that the user needs to have approximately 70 p.c. of the lexical inf. contained in the text in order to be able to understand the general meaning of text ([2], p.220). On the basis of the above data one could assume that the automatic processing of the lexical units (wordforms, phrases) without resorting to syntactical analysis of the sentence may possible yield sufficient inf. to unable the user understand the general meaning of the text. This conclusion applies, in the first place, to the synthetic languages, because their wordforms carry 20-30% more information then the wordfoms of an analytical language like English ([36], pp.234-245). In addition, the U/L shaped inf. profiles of the wordforms suggested posssibilities to compress the textual and the lexical data base (to save memory) — cf. [25], pp.73-77. Finally, the redunodant morphological endings of the running words offered better opportunities for the construction of algorithmes for grammatical analysis of the text ([13], p.30).

### 2.1.2 Statistical Structure of the Text

Research started in this direction in the late 50's with more emphasis on the morphology and the phraseology of the text, rather than on its syntax. The aim was:

- to study the rank distribution of the frequencies of wordforms, phrases, morphemes and of syntactical structures, using the law of Zipf ([1], pp.4-123; [32], pp.77-151);
- to analyse the laws that determine what is to be the share of certain frequent words in the text in order to formulate selection criteria and choose those words that are informationally loaded and grammatically indispensable for the Automatic Dictionary (AD);
- to compare the distribution of the relative frequency of a certain word (phrase), in specified portions of the text, with the theo-

retical laws of distribution (normal, lognormal, the Poisson law, the law of Čebanov–Fucks, of Markov–Kolmogorov, etc.) ([20], pp.110-131; [19], pp.345-360; [35], pp.247-282; [3]; [4]);

- to discover the dinamics with which new words and phrases appear in the fast developping sublanguages (SL) ([6], pp.171-178).

As a result of this research, towards the end of the 60's it was evident that:

A. The new and important notions, in all contemporary languages, are often expressed by means of phrases. Therefore it is essential to have a frequency dictionary (FD) of phrases for every Application Domain (DM) alongside with the traditional FD of wordforms and words ([10], pp.255-260). The DM–FD of phrases can be used to compile huge AD of phrases and words.

B. The most frequent non terminological words and phrases (including the Auxilary Words) have relatively stable frequencies in chronologically distant text (texts written at different times), also in texts belonging to different styles and SL, whereas the use of the terminological words and phrases is strongly linked to the time of writing and to the SL. As a result, the AD should be constantly update, adding all new terminological words and phrases met by NLP–systems.

C. There are no, up to now, universal laws of distribution that could help us determinr unmistakably the Part of Speech (PS) of a running word in the text. Still, a study, carried out recently by T.Kokočašvili ([21], pp.11-14), involving great lengths of English and Georgian texts, showed, that the agreement/disagreement of the empirically obtained distribution of words/phrases with the expected theoretical model, can be used as a probability filter for words with unknown meaning and grammar. As a result, the unknown (to the AD) word can be attributed to the Auxilary Words, or to the Adverb, or to the Numeral (the distribution of the Numerals is usually binominal or that of Poisson), or, finally, to the terminological Nouns (the empirical distribuition of the terms usually does not coincide with the theoretical models just mentioned).

D. The use of statistical data improved considerably the performance of the lexico–grammatical algorithms. Many Western colleagues recently came to the same observations ([44], pp.95-116; [43], pp.104-203; [42], pp.16-29).

### 2.1.3 Generation and comprehension of text

Without denying or doubting the importance of informational and statistic as well as traditional linguistic studies, it must be still admitted that they are not sufficient to serve as a basis upon which one could build the architecture of the computer analogue of the speech/thought activity of man. The theoretical basis was to be sought in the semiological hypothesis of man to man and man–computer communication. Machine semiosis, as a science started in the late 60's ([33]) with an aim to find an adequate computer models of text generation and comprehension (perception). On one hand, the efforts were directed towards reassesing and adjusting to the needs of man–computer communication ([7]; [12], pp.124-128, 131-132; [47]) the well known already linguistic, psicholinguistic and cognitive models. On the other hand, attempts were made to start independent investigation in psychiatric linguistics ecologically based and free from machine metaphor ([29]). All this led to the construction of a broader Saussurean model of the linguistic sign ([33]; Singariova, 1987) upon which a semiotic scheme of communication was developed, comprising various hypothesis of level (strata) generation and comprehension of information ([36], pp.220-221). This scheme (see Fig.1) serves as a psycholinguistic basis for the construction of an "intelligent" NLP–system and describes the formation of a message, starting with its denotative conception ($Dn$) which reflects a fact from the external reality, then going through the comment–topic (designative) image and then, finally, to its lexico–grammatical and phonetic/graphemic encoding. The transition from one level (strata) to another is monitored by the communicative pragmatic operator (CPO),

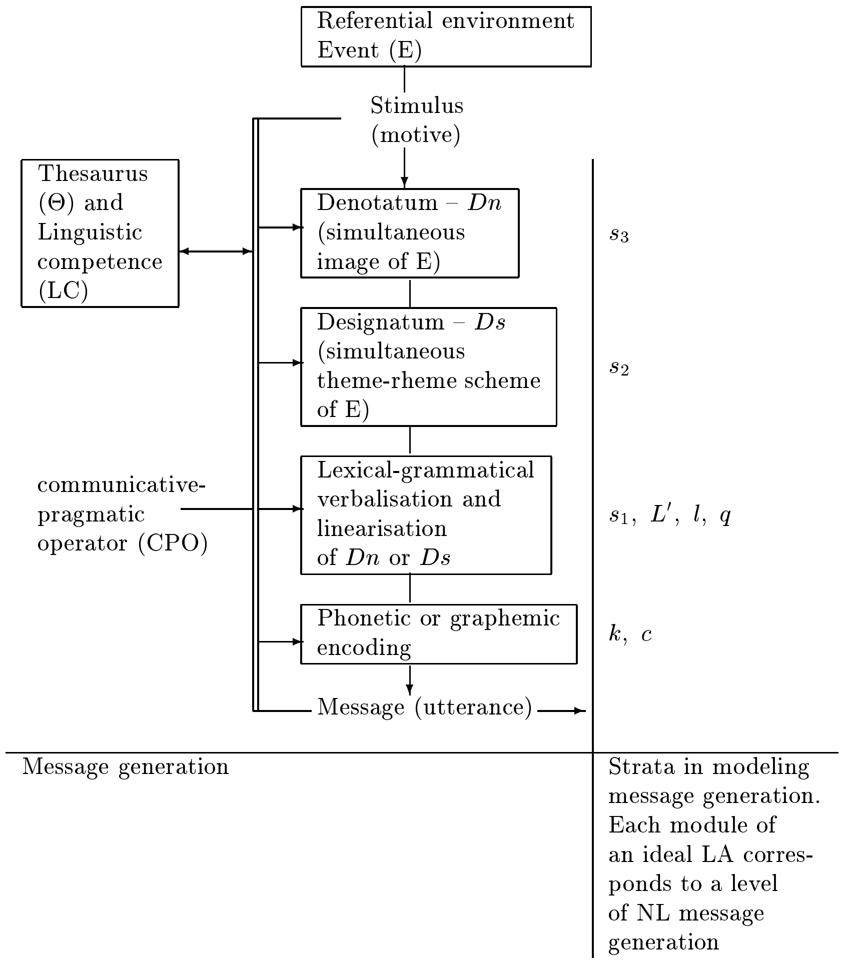


Figure 1: A Stratified Scheme of Message Generation

which selects linguistic knowledge and inf. from the thesaurus ($\theta$) and produces the message by regulating the transition from one strata to another ([34], p.100). Research carried out in Sp.St. has led to the choice of two schemes (see Fig.2) for comprehention and decoding of a message. According to the first hypothetical scheme this idea was expressed earlier by G.Miller and Ph.Johnson–Laird (1976) the recipient of sound or visual (graphic) signals compares those signals with the sensory records kept in the linguistic knowledge (phonemic–phonetic or graphic samples). The finding of a match triggers of a surface lexico–grammatical analysis of the sentence and of the wordforms/phrases comprising it. A deep topic–comment analysis on the designating level is carried out next; this analysis is based on the semantic–syntactic inf. obtained from the encyclopaedic and linguistic knowledge (the data base) and on the study of the context. Finally, the inf. about the message, gathered on all previous levels (strata), is being given a generalised interpretation in the last, denotative, strata. The recipient of inf. performs all the above mentioned operations on the basis of his personal pragmatics, presupposition and familirity with the situation. These operations are destined to help the recipient form a denotation ($Dn_2$), e.g. a generalised, simultaneous image of the fact contained in the message just received. The recipient has understood the message exactly as it was ment by the sender only when $Dn_1 = Dn_2$; the decoding of the message does not correspond to its initial meaning when $Dn_1 \neq Dn_2$.

According to the second hypothesis ([24], pp.217-250) the formation of $Dn_2$ begins within the framework of the sensory and the lexico–grammatical decoding of the message (blocks 4 and 5, Fig.2). The key components of the message particular words, phrases, simple semantico–syntactical constructs are also being established at this initial stage. The recipients performs all this on the basis of his/her pragmatic frame of mind and expectations, also acquaintance with the referential situation and presupposition. then comes the formation of a hypothesis about the message just received. This hypothesis is based on the inf. obtained in blocks 4 and 3. Finally, depending on the pragmatic frame of mind of the recipient and on his/her presupposition

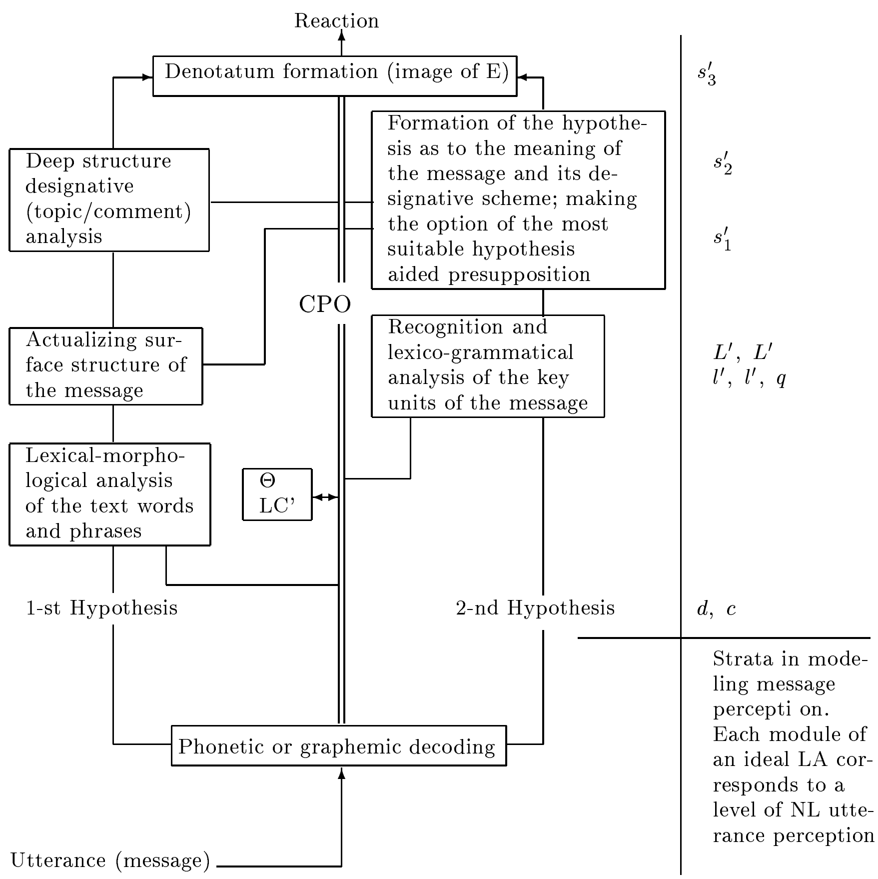


Figure 2: A Stratified Scheme of Message Perception and Analysis

(plus activation of the lexico–grammatical analysis, if such need arises) and by comparing the information, as it is being received, with the semantic–syntactic frames (segments) recorded in the thesaurus and in the linguistic knowledge (linguistic database, comprising all the linguistic competence) comes the choice of the most likely hypothesis. This hypothesis forms the basis of the actual representation of meaning to the denotation ($Dn$) of the message. Further on these two schemes (Fig.1 and 2) will be regarded as originals, identical with reality and the computer models for text analysis and synthesis will be based on them.

## 2.2 The Linguistic Automaton

Relying on the semiotic concept of level generation and comprehension of textual information and on the knowledge of the informational statistical properties of the text, and also taking into consideration the demands of the industry, of the users, towards the end of the 60's Sp.St. began to develop its own NLP conception. The key element in this conception is the idea of Linguistic Automaton. The realisation of this idea is still in progress.

The informational, social and economic situation during the second half of the 20-th century suggested that the Linguistic Automaton should comply with the following requirements:

a) to be multifunctional, i.e. to be able to perform different NLP tasks (indexing, annotation, abstracting, MT, etc.) involving huge amounts of textual information, depending on the user demands;

b) to minimise the loss of inf. and to weaken computer language rejection of natural language;

c) to be capable of regeneration, which means to possess an inbuilt ability to preserve its most essential properties in case of a failure, such as caused by viruses, power, hard disk, RAM failure, or any other failure causing, the distortion of words, etc.;

d) to allow for further development and improvement, by adapting the LA to the communication informational evolution of society and to the changing pragmatic outlook of the actual users of inf.

### 2.2.1 Linguistic Strategy for the construction of LA

The elaboration of a linguistic strategy for the construction of a LA necessitated the acceptance of one of the following two possible solutions.

The first required to decide whether the lexical or the grammatical element should have a priority in the construction of the algorithm for the LA. In view of this Sp.St. had to take into consideration the following facts:

a) the lexical units carry most of the inf. contained in the text as was evidenced by the results from our study of the text.

b) the automatic analysis and synthesis of the lexical units is subjected to a lesser degree to genotypical and phenotypical paradoxes, such as: human being — robot, fuzziness, diachrony — synchrony, etc. ([32], pp.49-61) than the grammatical analysis (parsing) of the input and the generation of syntactical structures at the output.

Therefore, it was thought to be more expedient to begin the construction of the LA by creating first a word/phrase data base and procedures to process the lexical units in the text rather than develop grammatical algorithms like most novices in NLP did.

The second solution required to choose between Chomsky's strictly deductive tradition and a probability based, functional grammar of text. The former still exists in some formalisms, applicable in the contemporary MT systems, such as Goverment Binding Grammar ([8]). Tree Adjoining Grammar ([18]), Phrase Structure Grammar ([40]). The alternative approach was shaped in the 60's in the works of J.Greenberg and his school (1963), cf. [11]; [16].

The results from the informational statistical measurements of texts consistently pointed out to the fact that the generation of text can be viewed as a complex Markov process, whereby strict planning is possible only within short intervals; the lexical units, further appart in the text, have fast fading stochastic ties ([31]). It was possible in the past to model the lexico–grammatical system of the language using traditional methods, such as relation algebra and non fuzzy linguistic sets ([36], pp.16-17), but in order to model the processes of analysis and synthesis of text inside the LA it was absolutely indepensable to use the functional linguistics of speech, relying on valency models of typical situation (frames), on probability models to resolve ambiguity and in the end on formal recognition of semantic patterns. Therefore the linguistic strategy adopted by Sp.St. for the construction of a LA differs from the NLP approaches used by most research teams. The difference lies in the priority given to the lexical studies, adapted to the probability based, functional linguistics of speech.

### 2.2.2 Architecture and Construction of the LA

Before going into details it is necessary to underline two of the construction principles formulated as a result of the linguistic strategy discussed above.

The first principle is the open (module level) stratification, allowing on one hand to remove from the LA some of the modules and replace them with other modules, and on the other hand, to relate each module to a particular level of generation/comprehension of the message (see Fig.1 and 2).

The second principle lies in the constant interaction between man and computer during the construction, performance and improvement of the LA. This means that we must use not only the traditional human knowledge about natural language when compiling Automatic Dictionaries and grammars, or when educating further the LA, but we should use also computer results obtained during the processing of large amounts of contemporary text belonging to a certain style or sublanguage (SL). The total amount of processed texts from certain style/SL should be used as a data base upon which to build a computer grammar for that particular style/SL. Besides, the texts should be processed not only in batch mode free from human interference but also in interactive mode through a constant man–computer dialogue. The computer tutorial should also be made in interactive mode.

### 2.2.3 Two Possible Schemes to represent the LA

The LA is a complex system, therefore a multiform representation is needed to describe it. The multiform representation includes models and schemes involving hardware, software, lingware, etc. approaches. Two representation schemes are of primary importance:

a) Structural — functional scheme;

b) Control and decision making scheme.

**2.2.3.1. Structural — Functional Representation.** This representation is made irrespective of the physical construction of the LA. It is level based system, having four layers (strata):

**1.** Lower stratum. The lower stratum is taken by the Linguistic Information Database (LIDB). This database is analogous to the lingvistic competence (knowledge) contained in the speech/thought apparatus of the man. The LIDB includes input and output dictionaries of the lexical units (the basic forms of the word, wordforms, the dictionary forms of the words and of the phrases), lists of morphemes and other relevant grammatical inf.

**2.** Middle stratum. It comprises a set of functional modules each of which performs a specified linguistic task, modeling a certain function of the speech/thought activity of man (Fig.1 and 2). This set of modules falls into two subsets. The first subset models the following analytical modules:

module for decoding of the text ($d$);

error correcting module ($c$);

module for lexical analysis of the key lexical units contained in the text ($l_k$);

module for lexical (word/phrase) analysis of all lexical units in the text ($l$);

module for independent morphological analysis of the running words in the text ($q$);

module for lexical and morphological analysis of the key lexical units in the text ($L_k$);

module for lexical and morphological analysis of all lexical units in the text ($L$);

module to analyse the surface structure of the text ($s_1$);

module for analysis of the deep (topic — comment) structure of the text ($s_2$);

module for semantic pragmatic analysis of the text ($s_3$).

The second subset unites synthetising modules, such as:

module for graphemic or phonemic representation (coding) of the text ($k$);

error correcting module ($c$);

module for lexical synthesis (choosing of lexical equivalents from the AD for the input wordforms and phrases) ($l'$);

module for lexico–morphological synthesis of wordforms and phrases ($L'$);

module for synthesis of the surface structure of the output text ($s'_1$);

module for synthesis of the topic — comment structure of the output text ($s'_2$);

module for synthesis of the semantic — pragmatic meaning of the output text ($s'_3$).

At the upper stratum some concrete systems and subsystems (SS) of NLP are placed. Several of which are worth mentioning here:

- SS defining the analysed text as belonging to a certain language ($L$),
- SS of word–by–word and phrase–by–phrase MT ($T_1$),
- SS of rough lexical and morphological MT ($T_2$),
- SS of semantic–syntactic MT ($T_3$),
- SS of topic–comment MT for headlines of articles and books ($T_4$),
- SS of text fragmentation ($Fr$),
- SS of indexing the text and its pragmatics ($I$),
- SS for forming a retrival pattern of a document as well as for annotating and text abstracting ($A$),
- an expert SS for a man–machine dialogue processing, aimed at checking and controlling the correctness of target text ($E$),
- SS systematizing text elements (letters, combination of letters, words, phrases) according to the alphabetic or frequncy order.

**3.** The upper stratum (F) may be presented as a set of generating functions:

$$F = \{F_i\}, i = \overline{1, m},$$

where $F_i$ — the function of generating a concrete NLP system or subsystem defined on a set of functional modules $M$.

In a "generalised form" $F_i$ is a logical function having the following form:

$$F_i = m_i^1 \wedge m_i^2 \wedge \ldots m_i^k, \qquad \text{where } m_i^j \in M, \quad j = \overline{1, k}$$

Then the SS of lexical and morphological MT will be described by means of the functions

$$\begin{aligned} F(T_2) &= \{d \wedge L \wedge L' \wedge k\} \quad \text{(cf. Appendix A and Fig.4)}, \\ F(T_2') &= \{d \wedge l \wedge q \wedge L' \wedge k\} \quad \text{(cf. Fig.5)}, \end{aligned}$$

SS of topic–comment MT of headlines will be defined by the function

$$F(T_4) = \{d \wedge c \wedge L \wedge s_1 \wedge s_2 \wedge s_3 \wedge L' \wedge s'_1 \wedge s'_2 \wedge k\} \quad \text{(cf. Fig.3 and 4)},$$

SS for abstraction and reviewing document will be defined by the function

$$F(A) = \{d \wedge L \wedge L' \wedge k\} \quad \text{(cf. Appendix B)}$$

etc.

At present the practical implementation of the upper stratum is being carried out through incessant man–computer interaction.

All system and service facilities of the LA are held in strata 1 and 2.

The numbers of functional SS of course may be enlarged. Thus, experimental modules of text generation (including poetic texts) are being worked out according to a given content "motive". Attempts to use the LA have been made teaching foreign languages. By means of providing the automaton with linguo–didactic data (CALL–ware) it may be turned into didactic linguistic automaton (DLA).

**2.2.3.2. Decision Taking System of the LA.** Any NLP–system when faced with uncertainty performs operations of identification, similar to the operations performed by humans in their speech/thought activity. The uncertainty is presented as options to choose form, registered in the AD and in the grammar.

The LA, endowed with certain Artificial Intellect, has to select the right alternative. Therefore we must describe the cconstruction of the LA not only structurally and functionally, but also with a view of its decision making. Similar to other control and management systems, the decision making body of the LA can be described as a hierarchic organisation of the following strata:

a) Self organisation;

b) Adaptation of the LA to the texts processed by it;

c) Choice of the optimal solution for a certain task.

On the first self organisation stratum, a strategy is adopted for the resolution of the task and it is being decided what subsystems and modules will be needed to solve it. This is done usually in man–computer communication mode.

In order to understand better the role of the second stratum we should bear in mind that in the course of resolving the linguistic task the LA usually finds itself in a situation of uncertainty. This uncertainty, or ambiguity, arises as a result of the polysemy of the lexical units listed in Dictionary, the polysemy of the morphemes and of the syntactic structures that form the text, and also, as a result of the insufficient linguistic and encyclopedic knowledge contained in the database. Therefore, the decision making body of the LA must possess means to resolve this type of uncertainty, to begin with, such as filtering algorithms that will partially resolve the ambiguity ([36], pp.91-145). Also, ways be found to adapt the LA to the texts processed by it. Most important of these is the updating the AD by registering in it all new geographical names, personal names and terms that characterise the sublanguage in question, and also, the creation of new algorithmes and revision of the ones already in operation. All this reeducation of the LA is carried out both in interactive regime and in the autonomous, software mode, in which the computer chooses the most frequently used alternative solutions met in the text ([32], pp.296-298).

The most important problem in the development of the concept of a LA is how to organise and make function the mechanisms of the third stratum. We'll look into this matter more closely.

**2.2.3.3. The choice of the best way to solve the task.** The Sp.St. group has worked out several methods aimed at minimizing faults due to engineering and linguistic limitations affecting LA. A part of them has been already turned into software. Two methods are of special interest.

The first of them explains how to organise, hierarchically, the performance of all subsystems and modules, taking into account the fact that they can function independently of each other. The hierarchic organisationn is done in the following order:

- the man–computer decision making body is the highest control stratum (see above);
- the subsystems and the modules belonging to the upper strata determine the work of those on the lowel strata;
- if the subsystems and the modules of a lowel strata cannot reach a decision while working in autonomous mode, or when they are faced with several options, then the text processing results are passed on to the upper stratum for a final decision.

Let's discuss these procedures, using as an example the system for topic–comment MT (see Fig.3) of German article titles from the SF "Sewage Waters" ([9]). The lexico–grammatical processing of the title strats in block 2, wich is a lower stratum of the LA. The computer uses a German, Russian AD and electronic morphology stored in the LIDB. The LIDB contains also inf. about the morphological border markers between syntactic structures, semantic inf. to be used for decision taking at a higher level, and translation patterns.

Then the title of the article is being split into semantico–syntactical segments, in block 3, on the basis of the morphological indicators obtained in block 2 and the syntactical border markers extracted from the LIDB. After that, the segments are passed on to blocks 3, 4 and 5, where they are being screened by a series of filters till one of the alternative solutions concerning the communicative (topic or comment) nature of the title is being reached. The first filter, in block 3, relies entirely on probability and syntactical inf., because the statistical data have shown that 90 p.c. of the 1–st segments of German article headings coincide with the topic (or are part of it). 70 p.c. of the terminal segments of the headings represent the topic. The 2–nd and 3–d segment of the title belong either to the topic or to the comment (statistically it is difficult to say to which). All this means that the position of the segment in the title does not allow for a single solution. Therefore, the results obtained in 2–nd block nd in block 3 must be sent to a higher level block 4, for further analysis of the meaning of the text and also to check if the segmentation of the title, carried out in block 3 is correct. There is a filter in block 4 designed to check if the lexical units constituting the title coincide whith the lexical units in the LIDB listed as indicators of topic or comment. The results obtained by this filter are compared with the results from the lower level (block 3). This comparison leads to one of the following alternatives.

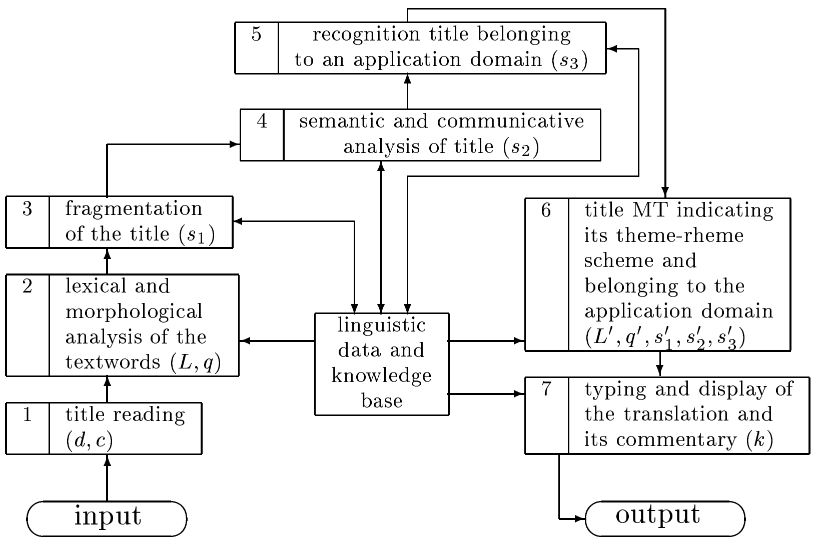


Figure 3: The theme–rheme (topic–comment) MT of German titles

1. The lexical units, identified as indicators of the topic, constitute the 1–st segments, while the lexical units, identified as indicators of the comment, constitute the terminal segments.

   In other words, the results obtained on each of the two levels coincide, and the LA takes the following decision: the terminal segment represents the topic. The attributive phrases adjoining the topic or the comment segments can also determine the topic or the comment.

2. The results obtained on level 3 contradict the results obtained on

level 4. In this case, priority is given to the results obtained on the higher (the 4–th) level.

3. If block 4 does not offer a single solution for the topic/comment segmentation of the title, the results obtained in block 3 and 4 are refered to block 5. Here the title of the article is attributed to one (or more) of the domain of the Subject Field (SF) in question. This was made possible, since each terminological wordform, listed in the AD of input terms has a tag (an index), showing the domain(s) of the SF (in our case "Sewage Waters") in which it is used, or its use in other related SF.

Our statistical data show, that the terms belonging to the SF "Sewage Waters" are more often used in the topic of the title, while the terms, belonging to other SF are usually used in the comment. This means that the presence, in a segment, of a wordform, attributed to a particular SF may provide additional inf. that will help the LA decide if the segment belongs to the topic or to the comment.

Let's illustrate this using the topic/comment analysis of a German title from the journal "Wasserwirtschaft Wassertechnik" (1985, No.1, S.4). A computer printout is shown on Fig.4.

It was not possible to identifly the topic–comment of the title in blocks 2 to 4. Then, all inf. concerning the task is passed on to the upper (the 5–th) block. Here, the terms, contained in the title, are being attributed to the respective domain or SF, which, in turn, leads also to the attribution of the title to a particular domain. It is important to note, that the terms, attributed to the "Sewage Water", are regarded as weak indicators of the topic, while all other wordforms and phrases, attributed to a domain or SF other than "Sewage Waters", are believed to be indicators of the comment. On the basis of this rule, the segment "eines Auswerterrechners" is included in the comment of the title. On the other hand, the subject field index of the segment "Trinkwasseraufbereitung" confirms its attribution to the topic of the title. The nature of the segment "bei Verfahrensuntersuchungen" is still unclear and computer program asks the user for help.

Title: Einsatz // eines Auswerterrechners // bei Verfahrensuntersuchungen // in der Trinkwasseraufbereitung.
Einsatz S, N/D/Ac, Sg, R (941)
eines Art, G, Sg, Gs (0)
Auswerterrechners Cp, S, G, Sg, M (105)
. . . . . . . . .
Trinkwasseraufbereitung Cp, S, Cc, Sg, T,W (4001)
Einsatz eines Auswerterrechners – comment
in der Trinkwasseraufbereitung – topic
bei Verfahrensuntersuchungen – no solution (nature unclear)
Auswerterrechners – SF "Machines and Equipment"
Trinkwasseraufbereitung – domain "Management of Water Industry"
Translation (in Russian): Primenenije vyčislitelnoj mašiny pri issledovanii metodov podgotovki pitevoj vody 'Application of computer in the search of methods to process drinking water'.

Indexes: // – borders of the segment; Ac – Accusative; Art – Article; Cc – Common Case; Cp – Compound word; D – Dative; G – Genetive; Gs – border signal; M – SF "Machines and Equipment"; N – Nominative; R – comment; S - Noun; Sg - Singular; TW - domain "Water Industry Management".

The figures in brackets show the places of the Russian equivalens in the memory.

Figure 4:Topic–comment MT of a German title

The second method for optimal solution relies on the decomposition capability is the ability of the LA to modify the task $(P)$ finding a simple way to present it, in cases when no other solution is found, or when much time and memory is need to resolve it, and the LA cannot afford it at the moment. The task $(P)$ is decomposed and presented as a series of specific tasks $P_1, P_2, \ldots, P_k$. We'll consider an example from the experimental Russian–Turkish MT. The incompatibility of the Turkish nominal and verbal paradigms with the corresponding Russian one is really great. Therefore, the lexico–morphological modules $L$ and $L'$ prove to be incapable, without modules $s/s'$, of forming Russian text–words and phrases morphologically coinciding with Turkish input–items. As these modules have not yet elaborated we'll present the lexico–morphological task $L/L'$ as a sequence of following procedures, functioning independently of one another:

$P_1$ — analysis of the Turkish wordform and its dissection into constituent morphemes (root plus affixes), – see module $q$;

$P_2$ — to determine the grammatical role of each affix (module $q'$);

$P_3$ — translating of the root (the word basis) (modules $l/l'$).

Then the user translates himself the input Turkish sentence, using the results of each procedure (see Fig.5).

A typical example showing the necessary to substitute the task $P$ with a modified and simplified task $\hat{P}$ is a that when the LA has not got sufficient morphological and semantico–syntactical resources to construct a surface and a deep structure of the input sentence and has to switch it self to word/phrase MT. The decomposition and simplification of $P$ increases considerably the flexibility and the viability of the LA since it helps to escape from deadlocks that occur when the automaton fails to follow a prespecified text processing format.

## 3 Conclusion

The concept for the construction of multistructural and polifunctional LA, able to model the speech/thought activity of man, came as result of decades of research into the linguistic aspects of the Artificial Intelligence. This research was carried out by over two hundred scientests from different specialities, organised in a team know as Speech Statistics Group. For the past twenty five years several experimental and industrial LA were constructed on the basis of the above concept.

In the rest of the world, the idea to create multifunctional NLP–systems and to develop them experimentally, appeared in the few years ([23]; [22]).

In goes without saying, that the existing multifunctional NLP–systems are still long way off from becoming full fledged LA. The trouble is that much of the feedback and module interaction is left to man. There is much more human interference in the control of the upper structural/functional and decision making strata, than in the lower, primitive blocks of the LA. It is to be expected that in the very near future more efforts will be concentrated to broaden the decision making

edinilen bilgiye göre iş bankasi, öncelikle federal almanyada dört ya da beş
şube açmayi planlamiştir

edin polucat' '(to) receave'
il PASSIVE VOICE
en PARTICIPLE
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
bilgi svedenie 'information'
ye DATIVE
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
gore soglasno 'according'
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
is trud 'labour'
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
banka bank 'bank'
si ATTRIBUTIVE POSTFIX
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
öncelikle v pervuju očered' 'primarily'
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
federal federativnyj 'federal'
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
almanya Germania 'Germany'
da LOCATIVE
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
dört četyre 'four'
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
ya da ili 'or'
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
beş pjat' 'five'
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
şube filial 'branch'
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
açma otkrytie 'opening'
yi ACCUSATIVE
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
planla planirovat' '(to) plan'
miştir PERFECT

Post–editing of the target (Russian) text:

Soglasno polučennym svedenijam, Trudovoj Bank zaplaniroval otkrytie v pervuju ocered' v FRG četyrjoh ili pjati filialov.

English translation:

According to the information recieved the Labour Bank plans to open primarily in the Federal Republic of Germany four or five branches.

Figure 5: Turkish–Russian lexico–morphological MT with postediting

capability of the communicative pragmatic operator that controls the speech/thought activity of man.

Abbreviations used in the references below:

**ACT** — Applied Computer Translation (journal, UK);

**ADML** — Automatic Documentation and Mathematical Linguistics (jounal, USA);

**ALLC** — Association for Literatury and Linguistic Computing (UK);

**LLC** — Literary and Linguistics Computing (journal, UK);

**RRL** — Revue Roummaine de Lunguistique (journal, Roumania);

**SL** — Statistica Linguistica (ed. R.Patron, 1971, Italy);

**StRAAT** — Statistica Reči i Avtomatičeskij Analiz Teksta (in Russian, collected works, issues 1971, 1973, 1974, 1978, 1980, Moscow–Leningrad).

# Appendix A. SpSt – group unrevised English-Russian translation

### English:

geneva. the u.n. commission on human rights has called on soviet authorities "to ensure human rights and fundamental freedoms in latvia and lithuania".

in his statement on february 26, the commission chairman enrique bernales ballesteros noted "the positive developments in the soviet union in respect to human rights and fundamental freedoms", but he expressed "grave anxiety over the recent tragic acts of violence including violations of human rights" in the baltic republics.

the statement will be included in the final report of the commission this year, and no further action is expected by the commission on the question of human rights in the baltic republics.

the commission noted the readliness of the government of the soviet union to give without delay to the chairman of the of the commission the results of this investigation, urged the authorities concerned to ensure that human rights and fundamental freedoms are fulfilled in latvia and lithuania. this statement will be included in report of the commission.

### Russian:

Женева. Комиссия ООН по правам человека призвала советские власти "гарантировать права человека и фундаментальные свободы в Латвии и Литве".

В его заявлении 26 февраля, председатель комиссии enrique bernales ballesteros отметил "положительные обстоятельства в Советском Союэе в уважении (отношении) прав человека и фундаментальных свобод", но он выразил "серьезное беспокойство по недавним трагическим действиям насилия включающего нарушения прав человека" в балтийских республиках.

Заявление будет включено в заключительное сообщение комиссии в зтом году, и никакое дальнейшее действие не ожидается комиссией по вопросу прав человека в балтийских республиках.

Комиссия отметила что готовность правительства Советского Союза чтобы давать без задержки председателю комиссии результаты зтого исследования, стимулировала власти затронутые (озабоченные) гарантировать что права человека и фундаментальные свободы выполнены в Латвии и Литве. Это заявление будет включено в сообщение комиссии.

**English:**

moscow, march 1 (xinhua). soviet defense minister said here thursday that iraq's crushing defeat in the gulf war shed light on the drawbacks of the soviet national defense on which the iraqi army largely depends.

tass quoted the minister who said that iraq calls for a review of the soviet union's entire air defense system. he said that the iraqi air defense system, which mostly consisted of soviet missiles, " failed in most cases".

earlier, however, the newspaper "izvestia" quoted a senior soviet military expert who said that the gulf war did not demonstrate the advantages of u.s. weapon because the iraqi army had little of the latest soviet technology.

the expert said: "iraq's weapon including soviet made, was produced mainly in the 1960s and 1970s".

**Russian:**

Москва, 1 марта (xinhua). Советский министр обороны сообщил здесь в четверг что сокрушительное поражение Ирака в войне залива проливало (роняло) свет на недостатках (препятствиях) советской национальной обороны на которых иракская армия в значительной степени зависит.

ТАСС процитировало министра кто сообщил что Ирак призывает к пересмотру (обзору) целой системы противовоздушной обороны Советского Союэа. Он сообщил что иракская система противовоздушной обороны, которая главным образом состояла иэ советских ракет (ракетных установок), "потерпела неудачу в наибольшем количестве случаев".

Ранее, однако, газета "Известия" процитировала старшего советского военного эксперта кто сообщил что война залива не демонстрировала преимущества американского оружия потому что иракская армия имела немного из последнего советской технологии.

Эксперт сообщил: "оружие Ирака включающее советское, было произведено главным образом в шестидесятых годах и семидесятых годах".

# Appendix B. Automatic processing of French patent

FRENCH PATENT

[15] BREVET D'INVENTION
[11] 2.073.829
[19] RE2PUBLIQUE FRANCAISE
[21] 70.45321
[22] 16 DE2CEMBRE 1970 14 H. 7 MN.
[46] B.O.P.I. "LISTES" N 39 1-10-1971
[51] F 16 D 13/00 B 60 K 17/00
[54] ACCOUPLEMENT A8 DISQUES DE FRICTION, EN PARTICULIER EMBRAYAGE
POUR BOITE DE CHANGEMENT DE VITESSES DE VE2HICULES.
[71] SOCIE2TE2 DITE: DR. ING.H.C.F. RFA
-*-
REVENDICATIONS 2.073.829 .1. ACCOUPLEMENT A8 DISQUES DE FRICTION,
EN PARTICULIER EMBRAYAGE POUR BOITE DE VITESSES DE VE2HICULES,
COMPORTANT 2 DISQUES DE FRICTION RELIE2S A8 L' ARBRE D' ENTRE2E
DE LA BOITE DE VITESSES, CHACUN D' EUX, E2TANT ASSUJETTI A8 UNE
PLAQUE D' APPUI QUI EST SOUMISE A8 L' ACTION D' UN RESSORT
BELLEYILLE PRENANT APPUI SUR LE COUVERCLE DE L' ACCOUPLEMENT ET
CENTRE2 DANS UN PALIER D' EMBRAYAGE COULISSANT AXIALEMENT,
CARACTE2RISE2 EN CE QUE LES DISQUES DE FRICTION (16,16',17) SONT
MONTE2S SUR UN TOC D' ENTRAINEMENT COMMUN (20), QUI EST MUNI
D' E2LE2MENTS AMORTISSEURS (19) AGISSANT PAR LE TOC D' ENTRAINEMENT (20)
SUR LES DEUX DISQUES DE FRICTION (16,16',17).
2. ACCOUPLEMENT A8 DISQUES DE FRICTION SUIVANT LA REVENDICATION 1
CARACTE2RISE2 EN CE QUE LE TOC D' ENTRAINEMENT (20) EST MONTE2
COULISSANT AXIALEMENT SUR L' ARBRE D' ENTRE2E (14) DE LA BOITE
DE VITESSES, ET EN CE QUE L' UN DES DISQUES DE FRICTION (16,16',17)
PEUT COULISSER AXIALEMENT, INDE2PENDAMMENT DU TOC D' ENTRAINEMENT (20).
3. ACCOUPLEMENT A8 DISQUES DE FRICTION
SUIVANT LA REVENDICATION 2
CARACTE2RISE2 EN DE QUE
LE DISQUE DE FRICTION (16)
POUVANT COULISSER AXIALEMENT
INDE2PENDAMMENT DU TOC D' ENTRAINEMENT (20) EST
RELIE2 A8 CE DERNIER PAR PLUSIEURS
DOUILLES DE GUIDAGE (25) RE2GULIE8REMENT RE2PARTIES
A8 LA PE2RIPHE2RIE DU TOC D' ENTRAINEMENT (20).
4. ACCOUPLEMENT A8 DISQUES DE
FRICTION SUIVANT LA
REVENDICATION 3 CARACTE2RISE2 EN CE QUE LES
DOUILLES DE GUIDAGE (25), RE2ALISE2ES EN ACIER OU
EN MATIE8RE SYNTHE2TIQUE, SONT MONTE2ES SUR LES
BOULONS DE FIXATION (24) DU TOC D' ENTRAINEMENT (20).
5. ACCOUPLEMENT A8 DISQUES DE FRICTION SUIVANT
LES REVENDICATIONS 1 ET 2 CARACTE2RISE2 EN CE QUE
L' UN DES DISQUES DE FRICTION (16') EST
GUIDE2 DANS LE SENS AXIAL PAR LE TOC
D' ENTRAINEMENT (20), PAR L' INTERME2DIAIRE
D' UNE DENTURE (26).

**

## AUTOMATIC INDEXING OF FRENCH PATENT, EXTRACTING AND TRANSLATION OF KEY WORDS

[15] ВИД ПУБЛИКАЦИИ:
ПАТЕНТ
[11] НОМЕР ПАТЕНТА:
2.073.829
[19] СТРАНА ПАТЕНТОВАНИЯ:
ФРАНЦИЯ
[21] РЕГИСТРАЦИОННЫЙ НОМЕР ЗАЯВКИ:
70.45321
[22] ДАТА ПОДАЧИ ЗАЯВКИ:
16 ДЕКАБРЬ 1970 14 ЧАС. 7 МИН.
[46] ДАТА ОПУБЛИКОВАНИЯ ФОРМУЛЫ ИЗОБРЕТЕНИЯ И N БЮЛЛЕТЕНЯ:
B.O.P.I. "LISTES" N 39 1-10-1971
[51] МЕЖДУНАРОДНАЯ КЛАССИФИКАЦИЯ ИЗОБРЕТЕНИЙ:
F 16 D 13/00 B 60 k 17/00
[54] НАЗВАНИЕ ИЗОБРЕТЕНИЯ:
СЦЕПЛЕНИ (СОЕДИНЕНИ) С ДИСК ТРЕНИ, В ОСОБЕННОСТИ
СЦЕПЛЕНИ ДЛЯ КОРОБК ПЕРЕКЛЮЧЕНИЯ ПЕРЕДА
[71] ЗАЯВИТЕЛЬ:
SOCIE2TE2 DITE: DR. ING.H.C.F. RFA

ФОРМУЛА ПАТЕНТА СЦЕПЛЕНИЕ (СОЕДИНЕНИ), ДИСКИ, ТРЕНИЕ, СЦЕПЛЕНИЕ, КОРОБКА ПЕРЕДАЧ, ТРАНСПОРТНЫЕ СРЕДСТВА, ВХОДНОЙ ВАЛ, ОПОРНАЯ ПЛАСТИНА, ДЕЙСТВИЕ, ПРУЖИНА, КРЫШКА, ПОДШИПНИК СЦЕПЛЕНИЯ, ПАЛЬЦА (ХОМУТИК), АМОРТИЗАТОРЫ, ДИСК, НАПРАВЛЯЮЩИЕ ВТУЛКИ, ПЕРИФЕРИЯ, СТАЛЬ, СИНТЕТИЧЕСКИЙ МАТЕРИАЛ, СКРЕПЛЯЮЩИЕ БОЛТЫ, НАПРАВЛЕНИЕ, ЗУБЧАТОЕ СЦЕПЛЕНИЕ.

## AUTOMATIC FRAGMENTATION AND RUSSIAN ABSTRACTING

ФОРМУЛА ПАТЕНТА 2.073.829.

НАЗВАНИЕ ПРЕДМЕТА ИЗОБРЕТЕНИЯ:
1. СЦЕПЛЕНИЕ (СОЕДИНЕНИ) С ДИСКАМИ ТРЕНИЯ, В ОСОБЕННОСТИ СЦЕПЛЕНИ
НАЗНАЧЕНИЕ ПРЕДМЕТА ИЗОБРЕТЕНИЯ:
ДЛЯ КОРОБКИ ПЕРЕДАЧ ТРАНСПОРТНЫХ СРЕДСТВ.
СОВОКУПНОСТЬ ОГРАНИЧИТЕЛЬНЫХ ХАРАКТЕРИСТИК ПРЕДМЕТА ИЗОБРЕТЕНИЯ:
СОДЕРЖАЩ 2 ДИСК ТРЕНИЯ, СОЕДИНЕНН С ВХОДНЫМ ВАЛОМ КОРОБКИ ПЕРЕДАЧ, КАЖД ИЗ НИХ, БУДУЧИ ПРИКРЕПЛЕНН К ОПОРНОЙ ПЛАСТИНЕ, КОТОР. ЯВЛЯЕТСЯ ПОДВЕРЖЕНН. ДЕЙСТВИЮ ПРУЖИНЫ BELLEYILLE, ОПИРАЯСЬ НА КРЫШКЕ СЦЕПЛЕНИЯ (СОЕДИНЕНИ) И ЦЕНТРИРОВАНН. (СОСРЕДОТОЧЕНН.) В ПОДШИПНИКЕ СЦЕПЛЕНИЯ СКОЛЬЗЯЩ. (СКОЛЬЗЯ) АКСИАЛЬНО.
СОВОКУПНОСТЬ ОТЛИЧИТЕЛЬНЫХ ХАРАКТЕРИСТИК ПРЕДМЕТА ИЗОБРЕТЕНИЯ:
ОТЛИЧАЮЩ ТЕМ, ЧТО ДИСКИ ТРЕНИЯ (16,16',17) МОНТИРОВАНЫ (СОБРАНЫ, УСТАНОВЛЕНЫ) НА ПАЛЬЦАХ (ХОМУТИК) ОБЩИЙ (АЯ,ЕЕ) (20), КОТОР. ЯВЛЯЕТСЯ ОСНАЩЕНН АМОРТИЗАТОРАМИ (19), КОТОР. ДЕЙСТВУЕТ (ДЕЙСТВУЯ) ПРИ ПОМОЩИ (ЧЕРЕЗ) ПАЛЕЦ (ХОМУТИК) (20) НА 2 ДИСКОВ ТРЕНИЯ (16,16',17).
НАЗВАНИЕ ПРЕДМЕТА ИЗОБРЕТЕНИЯ:
2. СЦЕПЛЕНИЕ (СОЕДИНЕНИ) С ДИСКАМИ ТРЕНИЯ
ССЫЛКА НА ПРЕДЫДУЩИЕ ПУНКТЫ ФОРМУЛЫ ИЗОБРЕТЕНИЯ:
ПО ПУНКТУ ФОРМУЛЫ ИЗОБРЕТЕНИЯ 1
СОВОКУПНОСТЬ ОТЛИЧИТЕЛЬНЫХ ХАРАКТЕРИСТИК ПРЕДМЕТА ИЗОБРЕТЕНИЯ:
ОТЛИЧАЮЩ ТЕМ, ЧТО ПАЛЬЦА (ХОМУТИК) (20) МОНТИРОВАН (СОБРАН, УСТАНОВЛЕН), СКОЛЬЗЯЩ. (СКОЛЬЗЯ) АКСИАЛЬНО НА ВХОДНОМ ВАЛЕ (14) КОРОБКИ ПЕРЕДАЧ,
СОВОКУПНОСТЬ ОТЛИЧИТЕЛЬНЫХ ХАРАКТЕРИСТИК ПРЕДМЕТА ИЗОБРЕТЕНИЯ:
И ТЕМ, ЧТО ОДИН (ОДНА, ОДНО) ИЗ ДИСКОВ ТРЕНИЯ (16,16',17) МОЖЕТ СКОЛЬЗИТЬ АКСИАЛЬНО, НЕЗАВИСИМО ОТ ПАЛЕЦ (ХОМУТИК) (20),

НАЗВАНИЕ ПРЕДМЕТА ИЗОБРЕТЕНИЯ:
    3. СЦЕПЛЕНИЕ (СОЕДИНЕНИ) С ДИСЦАМИ ТРЕНИЯ
ССЫЛКА НА ПРЕДЫДУЩИЕ ПУНКТЫ ФОРМУЛЫ ПРЕДМЕТА ИЗОБРЕТЕНИЯ:
    ПО ПУНКТУ ФОРМУЛЫ ИЗОБРЕТЕНИЯ 2
СОВОКУПНОСТЬ ОТЛИЧИТЕЛЬНЫХ ХАРАКТЕРИСТИК ПРЕДМЕТА ИЗОБРЕТЕНИЯ:
    ОТЛИЧАЮЩ ТЕМ, ЧТО ДИСК ТРЕНИЯ (16), КОТОР. МОЖЕТ СКОЛЬЗИТЬ
    АКСИАЛЬНО НЕЗАВИСИМО ОТ ПАЛЕЦ (ХОМУТИК) (20) СОЕДИНЕН
    (А,0) С ЗТО ПОСЛЕДН. ПРИ ПОМОЩИ (ЧЕРЕЗ) НЕСКОЛЬКО
    НАПРАВЛЯЮЩИХ ВТУЛОК (25), РЕГУЛЯРНО РАСПРЕДЕЛЕНН
    (РАЗМЕЩЕНН) С ПЕРИФЕРИЕЙ ПАЛЕЦ (ХОМУТИК) (20).
НАЗВАНИЕ ПРЕДМЕТА ИЗОБРЕТЕНИЯ:
    4. СЦЕПЛЕНИЕ (СОЕДИНЕНИ) С ДИСКАМИ ТРЕНИЯ
ССЫЛКА НА ПРЕДЫДУЩИЕ ФОРМУЛЫ ПРЕДМЕТА ИЗОБРЕТЕНИЯ:
    ПО ПУНКТУ ФОРМУЛЫ ИЗОБРЕТЕНИЯ 3
СОВОКУПНОСТЬ ОТЛИЧИТЕЛЬНЫХ ХАРАКТЕРИСТИК ПРЕДМЕТА ИЗОБРЕТЕНИЯ:
    ОТЛИЧАЮЩ ТЕМ, ЧТО НАПРАВЛЯЮЩИЕ ВТУЛКИ (25), СДЕЛАНН ИЗ
    СТАЛИ ИЛИ В СИНТЕТИЧЕСКОМ МАТЕРИАЛЕ, МОНТИРОВАНЫ (СОБРАНЫ,
    УСТАНОВЛЕНЫ) НА СКРЕПЛЯЮЩИХ БОЛТАХ (24) ПАЛЕЦ (ХОМУТИК) (20).
НАЗВАНИЕ ПРЕДМЕТА ИЗОБРЕТЕНИЯ:
    5. СЦЕПЛЕНИЕ (СОЕДИНЕНИ) С ДИСКАМИ ТРЕНИЯ
ССЫЛКА НА ПРЕДЫДУЩИЕ ПУНКТЫ ФОРМУЛЫ ИЗОБРЕТЕНИЯ:
    ПО ПУНКТАМ ФОРМУЛЫ 1 И 2
СОВОКУПНОСТЬ ОТЛИЧИТЕЛЬНЫХ ХАРАКТЕРИСТИК ПРЕДМЕТА ИЗОБРЕТЕНИЯ:
    ОТЛИЧАЮЩ. ТЕМ, ЧТО ОДИН (ОДНА, ОДНО) ИЗ ДИСКОВ ТРЕНИЯ (16')
    ЯВЛЯЕТСЯ УПРАВЛЯЕМ (А,0) В НАПРАВЛЕНИИ АКСИАЛЬН. ПРИ ПОМОЩИ
    (ЧЕРЕЗ) ПАЛЕЦ (ХОМУТИК) (20), ПОСРЕДСТВОМ ЗУБЧАТОГО
    СЦЕПЛЕНИЯ (26).

ФРАГМЕНТИРОВАНИЕ ТЕКСТА ЗАКОНЧЕНО.

END

## References

[1] Alekseev P.M. Statistishe Lexikographie. Zur Typologie, Erstellung und Anwendung von Frequenzwörterbüchern. Bochum, Brockmeyer, 1984, 157 p.

[2] Alekseev P.M., German-Prozorova L.P., Piotrowski R.G., Ščepetova O.P. Foundations of the Statistic Optimalisation of Foreign Language Teaching // StRAAT. 1974, Leningrad, Nauka, 1974, pp.195-234. (Russian)

[3] Altmann G., Hammerl D. Diskrete Wahrscheinlichkeitsverteilungen I. Bochum: N.Brockmeyer, 1989, 242 p.

[4] Altmann G., Zöring P. Diskrete Wahrscheinlichkeitsverteilungen II. Bochum: N.Brockmeyer, 1992, 264 p.

[5] Beöthy C., Altmann G. // Piotrowski — Gesetz und der Lehnwortschatz // Zeitschrift fur Sprachwissenschaft, Bd 2, 1982, pp.171-178.

[6] Botnaru R.V. Automatic recognition of the meaning of scientific and technical texts on the basis of reference frames. Ph.D. Dissertation, Leningrad: Leningrad Univ.Press, 1985, 15 p. (Russian)

[7] Chafe W.L. The Recall and Verbalization of Past Experience // Cole R.W. (ed). Current Issues in Linguistic Theory, Bloomington and London, 1977, pp.215-246.

[8] Chomsky N. Lectures on Government and Binding. Dodrecht: Foris, 1981, 210 p.

[9] Chizhacovsky V.A., Popescu A.N. Automatic selection of documental information. Proccedings of the International Conference on Microelectronics and Computer Science. ICMCS-92. Kishinev, Moldova, October 21-23, 1992, Vol.3. Computer systems and image processing. Kishinev: Politechnical Institute, 1992, pp.11-16.

[10] Danejko M.V., Maškina L.C., Nehaj O.A., Šaranda A.N. Studio statistico della distribuzione lessicale delle forme. SL, 1971, pp.255-266.

[11] Fillmor Ch. The Case for Case // Bach E. and Harms R.T. (eds). Universals in Linguistic Theory. New York: Holt, Rinehart and Winston, 1968, pp.1-88.

[12] Gardner H. The Mind's New Science. A History of the Cognitive Revolution. With a new epilogue by the author: Cognitive Science after 1984, NY, Basic Books, Inc., 1985-87, 400 p.

[13] Georgiev H. English Algorithmic Grammar. ACT, Vol.1, No 3, 1991, pp.29-40.

[14] Georgiev H., Piotrowski R.G. A New Method of Measuring Moaning. Language and Speech, Vol.19, Part 1, 1976, pp.41-45.

[15] Greenberg J.H. (ed). Universals of Language. Cambridge, Mass.: MIT Press, 1963, 350 p.

[16] Halliday M.A.K. Notes on Transivity and Theme in English. Journal of Linguistics. Parts 1-2. Vol.3, 1967, pp.37-81, 199-214; Part 3. Vol.4, 1968, pp.179-215.

[17] Ignatieva M.C. Semiotics of verbal forms // Bogodist V.I. (ed). Linguostatistics and Computerlinguistics in Second Language Teaching. Ulyanovsk: Ulyanov Pedagogical State Institute, 1989, 30p. (Russian)

[18] Joshi A.K. Introduction to tree adjoining grammar // Manster–Ramer A. (ed). Mathematics of Language, Amsterdam: John Benjamin, 1987, pp.10-115.

[19] Kaširina M.C. Types of distribution of the lexical units // StRAAT, 1974, pp.336-360. (Russian)

[20] Kaušanskaja M.V., Lukianenkov K.F. Statistical selection of the key and terminological lexical units for attribution and summarizing of French scientific technical texts // Piotriwski R.G. (ed). Avtomatičeskaja Pererabotka Teksta, Kishinev, Štiinca 1972, pp.117-131. (Russian)

[21] Kokočašvili T.G. Distribution of the lexical units in the texts (Georgian and English scientific and technical texts). Ph.D. Dissertation. Leningrad: Leningrad Univ.Press, 1990, 16 p. (Russian)

[22] Kuhns R.J., Little A.D. News Analysis: A Natural Language Application to Text Processing. American Association for Artificial Intelligence Systems, Palo Alto, CA: Stanford Univ.Press, 1990, pp.120-123.

[23] Loatman R.D., Post S.D. A Natural Language Processing System for Intelligence Message Analysis. Signal, Vol.42, No1, 1988, pp.41-45.

[24] Luria A.R. Language and Mind, Moscow: Moscow Univ.Press, 1979, 319 p. (Russian)

[25] Miklin G.Z., Piotrowski R.G. and Frumkin V.A. Word code contraction in automatic text processing. ADML, Vol.8, No 3, 1974, pp.73-77.

[26] Miller G.A. and Johnson-Laird P.N. Language and Perception, London: Cambr. Univ.Press, 1976, 150 p.

[27] Nikitina L.S. Nominal three word long phrases in Russian publicistic texts. StRAAT, 1971, pp.262-285. (Russian)

[28] Novac L., Piotrowski R.G. Esperimento di predizione ed entropia della lingua rumena. Sl, 1971, pp.325-363.

[29] Paszkowski W.E., Piotrowskaja W.R., Piotrowski R.G. Ideas and Methods ofn Psychiatric Linguistics // Beliaeva L.N. (ed). Statistica Reci i formaqlizovannyj analiz tekstovyh edinic. LGPI "A.I.Herzen", 1990, pp.109-115. (Russian)

[30] Petrova N., Piotrowski R.G., Guiraud R. L'entropie du francais ecrit. Bulletin de la societe de linguistique de Paris. T.57, fasc.1, 1964, pp.130-152.

[31] Piotrowski R.G. The Linguistics of a Text and Machine Translation. American Journal of Computational Linguistics. Vol.12, No 3, 1975, pp.95-96.

[32] Piotrowski R.G. Text — Computer — Mensch. Bochum: N.Brockmeyer, 1984, 422 p.

[33] Piotrowski R.G. Text Processing in the Leningrad Research Group Speech Statistics: Theory, Results, Outlook. LLC, Vol.1, No.1, 1986, pp.36-40.

[34] Piotrowski R.G. Linguistic Automaton and Computer Assisted Language Learning: a Perspertive, Language Lerning via the Microcomputer. Proceedings of CALL'89, Intern. Conference on

Computer Assisted Language Learning, Institute of Applied Linguistics, Rostock: Univ. of Rostock, 1990, pp.106-110.

[35] Piotrowski R.G., Bektaev K., Piotrowskaja A. Mathematische Linguistik, Bochum: N.Brockmeyer, 1985, 514 p.

[36] Piotrowski R.G., Lesohin M., Lukianenkov K. Introduction of Elements of Mathematics to Linguistics, Bochum: N.Brockmeyer, 1990, 260 p.

[37] Piotrowski R.G., Popeskul A.N., Chažinskaja M., Rachubo N. Automatische Wortschatzanalyse, Bochum: N.Brockmeyer, 1985, 107 p.

[38] Piotrowski R.G., Beliaeva L.N., Popescul A.N., Šingariova E.A. Bilingual Attribution and Summarizing of Text. Developments in Science and Technology, ser. "Informatica" t.7, Automatization of Indexing and of Summerizing of Documents, Moscow, 1983, pp.165-244. (Russian)

[39] Popeskul A.N. Production Network Approach to Modeling the Meaning of Scientific and Technical Texts. Ph.D. Dissertation. Vinnica: Vinnica Politechnic Institute, 1991, 25 p. (Russian)

[40] Ristad E.S. Computational Structure of GPSG Models. Linguistic and Philosophy, Vol.13, No 5, 1990, pp.521-587.

[41] Sampson G. Probabilistic Models of Analysis // Garside R., Leech G., Sampson G. (eds). The Computational Analysis of English, Harlow and London: Longman, 1987, pp.16-27.

[42] Sampson G. How fully does a machine usable dictionary cover English text 7. LLC, Vol.4, No1, 1989, pp.29-35.

[43] Sinclair J. Basic Computer Processing of Long Texts // Leech G., Candlin C.N. (eds). Computers in English Language Teaching and Research, Harlow and London: Longman, 1986, pp.104-203.

[44] Suppes P. Probabilistic Grammars for Natural Languages. Synthese, Vol.22, No1, 1970, pp.95-116.

[45] Tambovtsev Ju.A. Bibliography of the works of the Statistics of Speech Group in the USSR. ALLC Bulletin, Vol.13, No2, 1985, pp.50-53.

[46] Tambovtsev Ju.A. Further bibliography of the works of the Statistics of Speech Group in the USSR (1973-1985). LLC, Vol.1, No4, 1986.

[47] Zelinsky-Wibbelt C. From Cognitive Grammar to the Generation of Semantic Interpretation in Machine Translation // Steiner E.H., Schmidt P. and Zelinsky-Wibbelt C. (eds). From Syntax to Semantics. Insights from Machine Translation, London: Pinter, 1988, pp.38-52.

Ignatieva Maria, Chizhakovsky Valentine,
Popescu Anatol and Piotrowski Rajmund
Technical University of Moldova,
Str. Stefan cel Mare, 168
Kishinev, 277012, Moldova
phone: (373–2) 23–45–28

Received August 8, 1993