

Machine Translation in the Former Soviet Union and in the New Russia

XENIA PIOTROWSKA, RAJMUND PIOTROWSKI, YURI ROMANOV,
NATALIA ZAITSEVA, MICHAEL BLEKHMAN

More than forty years separate us from the advent of the first American and Soviet MT experiments and publications devoted to the analysis of the Natural Language Processing (NLP) problems. It is hardly possible to find another scientific field that has faced such dramatic upheavals. It is enough to mention the powerful reverberation of the MT ideas at the end of the 1950s - a wave that turned into a deep disappointment in the mid-sixties, then again an increase in the number of MT projects in the 1970s and the emergence of new ideas in the eighties followed by the financial crisis of the nineties.

1. ROMANTIC ERA

Soviet linguists and mathematicians were among the first in the world to produce examples of MT. That was in 1954, about only a year after the famous Georgetown University experiment.

This first phase of MT developments was a period of enthusiasm and romantic hopes. Indeed, on the crest of Khrushchev's thaw, with vast administrative and financial support of the Communist party, of the USSR Committee for Science and Technology, as well as of the Academy of Sciences, many research teams in Moscow, Leningrad, Gorky, Kharkov, Kiev, Tbilisi, Yerevan advanced on a number of NLP fronts.

The most significant ideas and tempting projects within the field of MT were those of N.Andreev at Leningrad State University and of O.Kulagina-I.Melchuk at the Steklov Mathematical Institute and the Institute of Linguistics (Academy of Sciences of the USSR, Moscow).

Andreev's research team was developing the MT problem in two directions. First, dozens of algorithms for binary MT from English, German, Norwegian, French, Spanish, Rumanian, Serbo-Croatian, Czech, Hindi, Turkish, Arabic, Japanese, Indonesian, Chinese, Vietnamese, Burmese, etc. into Russian were designed.

Second, the idea of translation via an intermediary language (interlingua) put forward (Andreev 1967:3-27). The interlingua MT required that a metalanguage be designed to summarize the descriptions of the major natural languages (NL). The necessity for such a description stemmed from the fact that the majority of semantic-syntactical relations and many morphological and semantic categories are expressed implicitly in NL, while in the interlingua all of them must be given explicitly. Andreev stressed the point that the interlingua must be found in the language of symbolic logic, taking into account combinatorial and probabilistic limitations for every lexical, morphological or syntactical feature.

This concept, except for the idea of a statistical-combinatorial approach, was accepted also by the Kulagina-Melchuk group. The question of fully automatic high quality MT should be resolved, according to the statement of the Moscow scientists, through the following steps:

- creation of a formal logical description for each NL,
- formal modeling human knowledge,
- formalization of mental processes (Kulagina and Mel'čuk 1967,48).

For some Soviet and, later, Russian researchers the postulate of highly accurate, fully automated translation together with the logical interlingua project came to be the focal theoretical point of NLP development. These ideas were based on Hjelmslev's and Chomsky's assumptions that the natural language, like the artificial languages of mathematics or logic, is a kind of calculus. At the same time, the Anti-Wittgensteinian thesis by Y. Bar-Hillel (1964) was neglected. Meantime, clouds began to gather on the horizon of the Soviet MT. In the early 1960s, Andreev's group sought to test a ready-to-function Indonesian-Russian MT system. That was the first attempt

to reach, by a computer, a coherent Russian-language translation of a randomly selected foreign newspaper text. The attempt failed. As for the other groups, none of the Moscow or provincial research teams were in a position to demonstrate any practical MT product. But the main blow was delivered by the ALPAC Report. It is important to remind that, for the Soviet bureaucracy, an American opinion on scientific, economic or military problems was of a paramount authority. Therefore, the statement of the American experts that high quality fully automated translation in a broad domain was not immediately feasible had fatal consequences for many Soviet MT groups. The Academy of Sciences and a number of ministries stopped the financial support for the development of MT systems. Traditionally-trained linguists with their anti-machine prejudice rose in applause to such a decision. As a result, the majority of pioneers in MT, and among them prof. N.D.Andreev, abandoned language engineering. The romantic age of MT based on the theoretical-logical approach ended, giving way to the realistic, pragmatic engineering one. It was at about this point of time that the Speech Statistics Group came into active operation.

2. PROSAIC TIMES

The Speech Statistics Group (SSG), originally named “All-Union Speech Research Group”, comprised linguists and language learners, mathematicians and programmers, psychologists and psychiatrists working in Minsk, Kishinev, Leningrad (St. Petersburg), Riga, Tartu, Kharkov, Baku, in the republics of Central Asia, and later in the USA, France, the Netherlands, Switzerland, Israel and Afghanistan. For the last thirty years the SSG has been busy with NLP problems including practical MT, as well as with linguistic and linguo-didactic problems of artificial intelligence (AI). Starting with a fundamental investigation of the statistical and informational nature of text and at the same time designing and developing small practical NLP systems, overcoming the lack of access to suitable equipment, the group, in the late 1970s, was coming into its own. Prof. F.E.Knowles from the University of Aston, who visited some Soviet MT centers in 1978, wrote:

One of the largest groups on the engineering linguistics “scene” is “All-Union Speech Research Group”, which is active in over thirty centers and which has published to date well over a thousand articles, the members of which have defended more than fifty DSc/DLitt theses and have published over a score of substantial volumes of works devoted to particular themes as well as a dozen or so university text books relating to this general field of activity. The field ranges very widely in fact from “pencil and paper”, but nonetheless algorithmic lexicography, via statistical linguistics, to the design and development of complex ISR and MT systems (Knowles 1979:70; cf. Tambovtsev 1985: 50-53).

In the early 1980s the SSG designed and developed the first practical NLP-system, in the USSR. That was a Chinese-Russian MT program processing wire communication texts from continental China (cf. Andrezen et al. 1992). The program served as a component of a great system tracking the political and economical situation in the People’s Republic of China. To create this program the Leningrad SSG research team was awarded a State premium in 1982.

The decay of the Soviet empire caused the SSG to change its line of activity. Losing the state’s financial support, the SSG research teams in new Russia, the Ukraine, Moldova and Byelarus’ began to develop and put on the European market commercial NLP programs, such as MULTIS-SILOD, STYLUS, SARMA, PARS etc. systems (SILOD 1986; Blekhman 2000; Piotrowski 1996:85-92) which were oriented towards the processing of business texts and were quite competitive with the international MT products.

Another line of today’s SSG activity is the elaboration of a methodology and developmental technology for a speech understanding system based on synergetics and semantic-pragmatics ideas (Kosarev, Piotrowski 1997:113-118). Certain work has been done towards the creation of a working model performing, automatically and in both directions, the following chain of tasks: perception and “comprehension” of oral speech by a computer-creation of a corresponding text file-translation of this file into a foreign language and creation of the corresponding foreign language text file-acoustic play-back of this foreign language text file in the corresponding

foreign language. Certainly, in a long-term prospect, any languages could be on either side of this chain.

Lastly, it should be particularly emphasized that the SSG has at its disposal a large stock of language resources for several languages, such as a few dozens of frequency dictionaries of text words, as well as of word triads and tetrads, micro-thematic glossaries, corpus-based dictionaries, robust parsers etc. (cf. TS 1969-1970; Alekseev 1984:109-118; *Statistica linguistica* 1973/1968; LE 1971-1985; SSAAT 1971-1980). These resources have been collected over the period of the last 30 years.

What was it that helped the SSG to overcome the adversities of the last decades and to carry out some practical MT systems at a time when a dozen of other Soviet teams failed to put into practice their widely publicized theories and promises to develop a highly accurate, fully automated MT and other NLP systems?

On the one hand, it was a direct link the SSG had with the actual users of the NLP product being developed. The users financed and controlled the research work, and this stimulated the development of a really practical MT system.

On the other hand, the totality of the MT pioneers' experience, both successful and unsuccessful, has been studied extensively. At the same time, comparative theoretical-experimental investigations of natural language and of the linguistic possibilities of the computer were pursued.

The SSG is the only research body on the NIS territory whose teams develop the real practical MT systems today. That is why we shall describe its activity in more detail here.

2.1. Theoretical Investigations in the SSG

The theoretical research has provided insight into the nature of the existing discrepancies between the human verbal/mental activity and the speech/intellectual capabilities of the computer. As it turned out, there are some "genetic" antinomies that form an invisible barrier between the natural language (NL) and the artificial language of the computer (Piotrowski 1984:41 -56). In the development of a real human NLP application there are at least three major obstacles.

The first antinomy lies in a divergence between the continuity of NL founded on tolerant fuzzy sets and the discrete nature of the computer language operating with non-fuzzy traditional sets (Zadeh 1973 ;Nalimov 1978).

The second antinomy consists in a discrepancy between the open dynamic (diachronic) character of NL and the static (synchronous) nature of its counterpart, i.e. the language of the computer. This antinomy lies in the fact that language, along with the human verbal/mental activity (VMA), is not a calculus, but an associative communicative fuzzy system (Dreyfus 1972; Melnikov 1988/1978). It follows that human performance in the process of text generation, perception and comprehension, its self-developing and self-enriching qualities are to be modeled in accordance with the strictly determined calculus used by the computer. Hence, NL VMA variability is reduced to a limited set of computer possibilities. In practice, the procedures are determined by a static model of an average expert competence and are specified in linguistic and knowledge databases.

The third obstacle, the so-called antinomy of three senses, resides in a discrepancy between the only possible monosemantic message which the computer should infer and the polyaspectual nature of a human utterance addressed by one person to another, which may have at least three different facets of meaning, two of them depending on the author's and recipient's pragmatics, while the third is independent of the interlocutors' pragmatics, being objective and universal.

Experimental work conducted along these lines has proved that the quality standard of MT depends much on the results that may be achieved in moderating or reducing the influence of these "genetic" antinomies. In this respect, the SSG was started in the early 1960s with the following aims in view:

- to study the informational and statistical structure of the text,
- to create a semiotic model of text generating and "understanding" by the computer.

Informational investigation was carried out for the Slavonic, Germanic, Romance, as well as for the Turkish and the Finno-Ugric

languages. Certain informational text characteristics important for development of the NLP systems were revealed (Piotrowski 1999: 113-121).

All the investigated languages, no matter to what type they belong, have redundancy in the range of 65-96 percent. The deviations within this interval depend on the type of text. Thus, business and patent documents display the highest redundancy level and are most explicit and adaptable to the computer language (85-96 p.c.); then come scientific/technical and political/sociological texts (about 80 p.c.). Fiction, as well as spontaneous irregular speech, displays the low redundancy level. At the same time, it has become clear that, to understand the general content of a text, it is sufficient to infer about 70 p.c. of information gathered from the lexicon of the text.

Informational investigations have proved also that between 18 p.c. (in the agglutinative and synthetic languages) and to 35 p.c. (in the languages with predominant isolating features, such as English) of the syntactic and semantic information is concentrated in the contextual links and relations. The rest of the information (from 82 p.c. for the synthetic and to 65 p.c. for the isolating ones) is concentrated in the lexical units (word-forms and phrases).

As a result of the experimental work, a hypothesis was formulated that the lexicon alone (that is, words and phrases), without going into the analysis of the syntactic structure of the sentence, may sometimes give information enough for the understanding of its general content. This holds true especially for the synthetic languages, in which word-forms themselves provide up to 20-30 p.c. more information than in the languages with predominant isolating features, such as English.

From the above line of reasoning, it is clear that the focus of SSG processing, for both the input and the output text, is at the level of the lexicon, i.e., the word, word-forms and phrases, rather than at the level of syntax. Thus, an automatic dictionary (AD) is the nucleus of any NLP system and more attention is focused on its performance, primarily, on the choice of the lexical units.

The frequencies of widely used full words and phrases, as well as those of syntactic words, are of greater stability in texts,

independently of their style and sublanguage, whereas the frequency of terms is basically determined by the time when the text was written and by its sublanguage. Hence, when developing an AD, syntactic and widely used full lexical units (LU) may be selected from explanatory, phrase and frequency dictionaries. Terminology, as it becomes outdated with time, should be periodically renewed and therefore it must be selected from sublanguage corpus-based dictionaries.

Without denying nor doubting the importance of informational and statistical as well as traditional linguistic investigations, it must be admitted that they are not sufficient to serve as a foundation to create a computer architecture that is analogous to the synergetics of human verbal/mental activity. Such theoretical foundations should be looked for in semiologic hypotheses derived from man-man and man-computer communication.

Research in the field of machine semiosis was carried out from the early 1960s till the late 1970s, with the aim of creating text-generation and text-comprehension models that could be used to form adequate computational procedures for text analysis and synthesis. On the one hand, the search was directed at adopting the existing linguistic, psycholinguistic and cognitive models for the needs of man-machine communicative purposes. On the other hand, a thorough search was carried out with the purpose of finding a new approach using the data of psychiatric linguistics, free from any machine metaphors.

As a result of these efforts, an extended Saussurian model of the linguistic sign was built, on the basis of which a semiotic communicative scheme comprising several hypotheses of stratificational generation and comprehension of a message were derived (Berzon, Blekhan, Piotrowski 1984:16-34; Piotrowski 1986:36-38; 1994:16-31).

This scheme is a psycho-linguistic prototype to build NLP "intellectual" systems. It describes message modeling beginning with the sender's denotatum purport (Dn1) reflecting a certain fact of objective reality, through a topic-comment designatum scheme, towards lexical and grammatical verbalization of the message and its linearization and, finally, to graphemic or phonemic encoding.

The procedure is carried out under the control of a communicative and pragmatic operator (CPO), which regulates the transfer of the message from one level to another by providing the choice of necessary information from the sender's thesaurus and his linguistic competence (LC).

As to the perception and decoding of the message, the researchers follow two hypothetical schemes.

According to the first scheme, the receiver compares sound and visual (graphical) signals to phonemic/phonetic and graphemic patterns stored in LC. If the comparison gives a positive result, the surface lexical and grammatical analysis of the message starts, including the analysis of its constituents - word combinations and text words. Then follows a deep topic/comment analysis at the designatum level, based on the semantic and syntactic information from the thesaurus, LC and from the contextual analysis. Finally, at the conclusive denotatum level, a summarized interpretation of information collected at the previous levels takes place. The operation is performed by the receiver on the basis of his personal pragmatics, presupposition, and background knowledge of the situation. All this should result in adequate modeling of the receiver's denotatum (Dn2), i.e. of a generalized simultaneous model of the fact or situation described in the given message. The equality $Dn1=Dn2$ means that the message is comprehended by the receiver in accordance with the purport. If $Dn1$ is not equal to $Dn2$ (cf. the antinomy of three senses), the decoding of the message is inadequate to the purport.

Another search of $Dn2$ goes within the frames of sensorial, lexical, and grammatical decoding of the message. At the initial stage the key marks (separate words, phrases, simple semantic-syntactic schemes) are interpreted. This search is performed by the receiver on the basis of his pragmatic purpose and anticipation including orientation in the referential field and presupposition. Then, on the ground of the previously received information, hypotheses are formed as to the content of the received message. Further, on the basis of the receiver's pragmatic purpose, his presupposition, using, if necessary, a lexical and grammatical analysis and, finally, by comparing received information with the frames destined for

syntactic and semantic processing, a more suitable solution is chosen. This solution is used to restore the denotative content of the message. All these operations of text comprehension are carried out under the control of the CPO.

2.2. *Linguistic Automaton*

From the late 1950s on, many language engineering groups tried to develop various MT-systems as well as some other independent working NLP-models. However, the closing years of the sixties witnessed the creation of operative systems of multifunctional text processing such as automatic abstracting of a foreign language document with its machine translation, or a computer-assisted language instruction incorporating MT techniques. Some attempts to develop a similar system, named LINGTON (LINGuistic automaTON), were launched in the late 1960s by the Soviet SSG. The idea of the LINGTON became a focal point of the SSG's theoretical and technological activity.

LINGTON is intended as a multi-purpose NLP-system which should model, in a robust way, the verbal/mental behavior of human beings in a particular social role: that of a translator, a text interpreter or referrer, and a language teacher (Czyakowski, Piotrowski 1993 161 -189). It should comply with the following requirements:

- (1) be multifunctional, i.e. to be able to perform such tasks as the initial statistical text processing, language/style/variety (British English vs. American English etc.) recognition, spell-checking, indexing, annotation, abstracting, man-machine dialogue, machine translation;
- (2) possess linking features which will enable its connectivity with the most common optical character recognition and word processing software;
- (3) possess a possibility of developing and use the resident dictionaries,
- (4) allow for further developments and improvements, by adapting LINGTON to the communication informational

evolution of society, for example, to Internet, to speech recognition and “understanding”, and to the changing pragmatic outlook of the actual users of information;

- (5) be adjustable to the processed linguistic material through feedback with operator;
- (6) possess a built-in ability to preserve its most essential properties in case of failure caused by viruses, RAM breakdowns, distortions of words etc.

2.2.1. *Linguistic Strategy for the Development of LINGTON*

There existed two main points in the LINGTON’s design and development.

The first one required deciding whether the lexical or the grammatical element should have the priority in designing the general algorithm of LINGTON. To this end, the SSG had to take into account the following considerations. On the one hand, the lexicon conveys the bulk of information the text contains. On the other hand, the automatic analysis and synthesis of separate lexical units is less subjected to the effect of the “genetic” paradoxes of NLP than syntactic processing. Therefore, the focus of processing of both input and output texts is at the level of the lexicon rather than at that of the grammar.

The second solution required to choose between Hjelmslev’s strictly deductive tradition of a linguistic calculus and a functional grammar relying upon the probabilities of text units. Here the SSG faced two possible ways:

- (1) to follow Chomsky’s scheme according to which a text is internally and inherently organized into canonical patterns, and, hence, such concepts as structure, planning, and selective mechanisms are to be involved into the procedure.

or

- (2) to assume the Markovian approach that suggests the process of text production as a unit-by-unit sequencing, which involves the concepts of probabilistic functional grammar,

where a regular structure does not necessarily imply a regular predetermined relation.

It is well known that our speech is, as are the majority of printed texts, full of syntactic errors, stylistic and semantic mistakes, it is disrupted by repetitions and hesitations. That is why the informational and statistic analysis proved that text production was a complex Markovian process, and fairly rigid planning and exact preconditioning of the structure worked only for short text segments, whereas distant linguistic units displayed quickly weakening stochastic bonds. Hence, it was only within the grammatical system that it was possible to use the model schemes built on the traditional basis of the set theory, mathematical logic, and relational algebra. In modeling the input text parsing and in generating the target text inside LINGTON, it was indispensable to use the functional text linguistics, relying on the valence models of frames, on the probability models, to be able to resolve ambiguity, and, in the end, on the formal recognition of semantic patterns.

Thus, the SSG strategy differed on the whole from the MT approaches used by the most Soviet research teams. The difference lay in the priority given to the probabilistic lexical studies, adapted to the demands of the users, as well as to the ideas of the functional text linguistics.

2.2.2. LINGTON *Architecture*

There were two possible approaches to the development of LA. The first one was a deductive development according to a rigid top-down scheme, i.e. from semantic and pragmatic levels to lexical-grammatical and encoding blocks. The second approach implied iterative development of the automaton, bottom up, i.e. from elementary lexicon blocks towards the more complicated semantic, syntactic and pragmatic levels.

With all its tempting simplicity the first approach had two serious drawbacks: it excluded the simultaneous realization of all the tasks arising in development of a polyfunctional practical NLP system and did not allow to make use of those scientific achievements,

which appeared while working on the system and after its fulfillment, without changing its architecture. As a result, the system built according to a rigid deductive scheme was inadequate to cope with the effect of the above-mentioned man-machine paradox and diachrony-synchrony antinomy.

The iterative approach proved to be more efficient in minimizing these antinomies and the effect of rejection it involved. This approach provided for an open (module level) stratification, allowing, on the one hand, to remove, from LINGTON, some of the modules and replace them with other ones, and, on the other hand, to relate each module to a particular level of generation/perception of the message.

LINGTON is a complex system, therefore a multiform representation is required to describe it. Two representation schemes are of primary importance. These are a structural-functional scheme and a management decision scheme.

2.2.3. *The structural-functional description*

Without considering the physical substratum of LINGTON, this description has the following four strata.

The lower stratum is taken by the linguistic and encyclopedic database (LEDB), which is analogous to the linguistic competence contained in the verbal/mental apparatus of man. The LEDB includes input and output dictionaries of the most frequently used word-forms, stems, phrases, as well as lists of grammatical affixes, toponyms, antroponyms, abbreviations, etc. For a detailed description of the database used in the SSG, see: Beliaeva, Piotrowski 1989/90:26-35.

The middle stratum is represented by a set of functional modules each of which performs a specific linguistic task, modeling a certain function of the human verbal/mental activity. This stratum incorporates the following modules:

- modules for graphemic/phonetic decoding or coding of the text,
- spellcheckers,
- modules for analysis/generation of lexical units (LU),
- modules of LU morphological analysis/generation,

- module carrying out the analysis/generation of the sentence surface structure,
- module for analysis/synthesis of the sentence deep topic-comment) structure,
- module carrying out the semantic-pragmatic (denotative) analysis correcting synthesis of the output text,
- abstracting module,
- module performing statistical analysis of the input text,
- tutoring module.

The upper stratum may be presented as a set of functions, each of which is to generate a concrete NLP system or subsystem defined on a set of functional modules. At present the practical implementation of the upper stratum is being carried out through incessant man-machine interaction.

2.2.4. *Decision scheme of LA*

Any NLP system is realized under indeterminacy conditions. This indeterminacy is represented in the LEDB and the algorithm blocks by a set of alternatives. Out of these the LINGTON is to select the correct decision. Therefore its architecture should be described not only from a structural and functional, but also from a decision viewpoint. Similar to other control and management systems, the decision making body of the LINGTON can be described as a hierarchy of the following strata:

- self organization,
- adaptation of LINGTON to the text it processes,
- choice of the optimal decision for a concrete task.

On the first self-organization stratum a strategy to solve the task is worked out and it is being decided what modules will be needed to solve it. This is done usually in the regime of man-machine interaction.

In order to understand better the role of the second level, we should remember that in course of text processing the LINGTON usually finds itself in a situation of uncertainty, caused by the polysemy

of LU, of morphological and syntactic ambiguity, as well as by the shortage of linguistic and encyclopedic knowledge stored in the database. Therefore, the decision making body of LINGTON must possess means to resolve this type of uncertainty, such as filtering algorithms which will partially eliminate the ambiguity. Also, ways should be found to adapt LINGTON to the text processed by it. The most important of these are updating the automatic dictionary by registering in it all the new toponyms, personal names, as well as terminological word-forms and phrases that characterize the sublanguage in question, and also the creation of new algorithms and revision of the ones already in operation. All this reeducation of LINGTON is carried out both in the interactive and the autonomous regimes.

The most important stratum for the development of the LINGTON concept is the third one. Therefore, the problem of its organization and functioning should be considered in detail. The SSG has worked out several methods aimed at minimizing faults due to engineering and linguistic limitations affecting LINGTON. Several of them have been already turned into lingware. Two of the methods are of special interest.

The first of them explains how to organize, hierarchically, the performance of all the modules, taking into account the fact that they can function independently of each other. The hierarchic organization is the following:

- the man-computer decision making body is the highest control level,
- modules belonging to the upper level determine the work of those of the lower levels,
- if the modules of a lower level cannot reach a decision or when they are faced with several options, then the text processing results are passed on to the upper level for the final decision.

EPILOGUE

It goes without saying that the existing multifunctional NLP-systems are still long way off from becoming fully-fledged linguistic automata. 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, more primitive blocks of LINGTON. It is to be expected that in a very near future more efforts will be made to model the decision making capability of the communicative pragmatic operator that controls the synergetics of verbal/mental activity of man.

REFERENCES

- Alekseev, P.M. 1984. *Statistische Lexicographie. Zur Typologie, Erstellung und Anwendung von Frequenzwörterbüchern*, Bochum: Studienverlag N. Brockmeyer.
- Andreev, N. 1967. The intermediary language as the focal point of machine translation. In: Booth, A.D. (ed.), *Machine Translation*, Amsterdam: North-Holland Publishing Company.
- Andrezen, V., Kogan, L., Kwiatkowski, W., Minvaleev, R., Piotrowski, R., Shumovsky, V., Tioun', E., Tovmach, Yu. 1992. *Automatic dictionary organization in NLP systems for Oriental languages*, Actes de COLING-92, Nantes, 23-28 août 1992. Nantes: Geta (IMAG) Assoc. Champollion.
- Bar-Hillel, Y. 1964. *Language and information: Selected Essays*. Reading, (Mass.): Addison-Wesley.
- Beliaeva, L.N., Piotrowski, R., Kondratieva, Ju., and Sokolova, S. 1989-1990. *Abstracts from the Leningrad MT Projects*, [SCCAC- Society for Conceptual and Content Analysis by Computer, Newsletter N50], Bowling Green, Ohio: State University.
- Berzon, V., Blekhman, M., Piotrowski, R. 1984. Connections, Units and Unities on the Intersentence Level, *Tartu Riikliku Ülikoole Toimetised*. Vihik 689. Tartu: Riikliku Ülikoole, (in Russian).
- Blekhman, M., Bezhanova, O., Kursin, A., Rakova, A. 2000. Translation Software by Lingvistica '98 Inc.: The PARS Family of Machine Translation Systems. *International Journal of Translation*. Jan-Dec 12:1-2., New Delhi: Bahri.

- Czyakowski, W., Piotrowski, R. 1993. Über den gegenwärtigen Stand der automatischen Textverarbeitung in der Forschungsgruppe 'Sprachstatistik' (Zum Problem des linguistischen Automaten), *Glottometrika* 14. Trier: WVT.
- Dreyfus, H.L. 1972. *What computers can't do. A critique of artificial reason*, New York: Harper and Row Publishers.
- Knowles, F.E. 1979. Recent Soviet work on computer techniques for representing natural language meaning, *Informatics* 5, London: The Queen College - Oxford.
- Kosarev, Yu., Piotrowski, R. 1997. Synergetics and 'insight' strategy for speech processing, *Literary and Linguistic Computing* vol. 12, no 2.
- Kulagina, O.S. and Mel'cuk, LA. 1967. Automatic translation: some theoretical aspects and the design of a translation system. In: Booth, A.D. (ed.), *Machine Translation*, Amsterdam: North-Holland Publishing Company.
- LE-Language Engineering - Collections of articles*. 1971; 1973; 1976; 1979; 1980; 1981; 1985. Leningrad: LGPI, (in Russian).
- Melnikov, G.P. 1988. *Systemology and linguistic aspects of cybernetics*, New York, Gordon and Beach [Russian. Original, Moscow: Sov. Radio, 1978].
- Nalimov, V.V. 1978. *Continuity against discretion in the language and in the thinking*, Tbilisi: Tbilisi University Press (in Russian).
- Piotrowski, R. 1984. *Text-Computer - Mensch*, Bochum, N.Brockmeyer [Rus. original, Leningrad: Nauka, 1975].
- 1986. Text processing in the Leningrad research group 'Speech Statistics' - Theory, Results, Outlook, *Literary and Linguistic Computing*. 1:1.
- 1994. Psycholinguistic basis of the linguistic automaton, *International Journal of Psycholinguistics. An Interdisciplinary Journal of Human Communication*, 10.1:27.
- , 1996. Machine translation in new Russia. In: Tufis, D. (ed), *Limba și Tehnologie*, București: Editura Academiei Române.
- . 1999. *Linguistic Automaton and its Verbal-mental Foundation*, Minsk: State Language University (in Russian)
- SILOD. *A Russian-English Translation Support* 1986. New Delhi: Computronics India.
- SSAAT - *Speech Statistics and Automatic Analysis of Text. Collections of articles*. 1971; 1973; 1974; 1978; 1980, Leningrad/Moscow: Nauka, (In Russian).

- Statistica linguistica* (con l'aggiunta di due appendici). 1971. Bologna: Casa editrice Prof. R.Patron [Rus. original, *Statistica rechi*. Leningrad: Nauka, 1968; Germ. translation: *Sprachstatistik*. Hgg. von P.M.Alexejev, V.M.Kalinin, R.G.Piotrowski. Muenchen/ Salzburg: Wilhelm Fink Verlag, 1973].
- Tambovtsev, Yu. A. 1985. Bibliography of the work of the statistics of Speech Group in the USSR, *ALLC Bulletin*. 13.2.
- TS - *Text Statistics. Collection of Articles*. 1969-1970. Minsk, Pedagogical Institute of Foreign Languages, Vol. I: *Linguo-statistical Studies* (1969); Vol. II: *Text Processing* (1970) (in Russian).
- Zadeh, L. A. 1973. *The Concept of a linguistic variable and its application approximate reasoning*, New York: American Elsevier Publishing Company.