Historical sketch of machine translation in Eastern and Central Europe[1]

John Hutchins

Abstract

This historical survey[2] describes machine translation (MT) in the former Soviet Union and its satellites in Eastern and Central Europe from 1954 to the present day. The first half concentrates on machine translation in Russia between 1954 and 1990, covering the research groups active in Moscow and Leningrad during the 1950s and 1960s, which after 1974 were brought together with other USSR groups as the All-Union Translation Centre. Before the mid 1970s the focus of most Russian groups had been on theoretical studies – primarily because of the lack of computer facilities – but after 1974 the focus shifted to the practical delivery of translations whatever the quality of the MT systems. Outside Russia, during the dominance of the USSR, there was also significant MT activity in Czechoslovakia, the German Democratic Republic, and Bulgaria. After 1990, there emerged two commercial systems from the Leningrad research: the STYLUS (PROMT) and PARS systems; and research has grown rapidly in the Czech Republic, Hungary, Poland, Bulgaria, Romania, and in the Baltic states. Much of this research activity since 1990 has centred on statistical machine translation, but the older rule-based 'tradition' has continued (particularly for translation between closely related languages).

1. The beginnings

Within months of Stalin's death in March 1953, Soviet researchers were investigating and developing Western developments in the new foreign sciences of computers, information theory, cybernetics – all previously condemned and forbidden as 'bourgeois' studies. Engineers and linguists learnt about machine translation very soon after the first demonstrations in the USA by the Georgetown and IBM groups (e.g. Berkov and Eršov 1955, Ljapunov and Kulagina 1955). In 1955 the first research groups were set up: two in Moscow under Dmitrij Panov and Aleksej Ljapunov and one in Leningrad under Nikolai Andreev.

Panov led the group at ITMVT (Institut Tochnoj Mekhaniki i Vyčislitel'noj Tekhniki); it is reported that he had been to the US and seen a demonstration; he was an officer of the KGB. By late 1954, ITMVT was able to run an English-Russian system, based closely on the Georgetown-IBM system, programmed for the new Russian BESM computer. In early 1956, Panov published a book devoted primarily to a description of the ITMVT system (Panov 1956a).

Ljapunov came across a description of the Georgetown experiment in the October 1954 issue of *Referativnyi Žurnal* and decided to start his own MT group at the Steklov Mathematical Institute. He recruited Olga Kulagina and Igor Mel'čuk to work on translation from French into Russian. Mel'čuk was actually employed at the Institute of Linguistics, but Ljapunov arranged for him to work at his own institute. The first publication of the group was a review of the Georgetown experiment (Ljapunov and Kulagina 1955.) The next year, Kulagina and Mel'čuk (1956) reported their research on French to Russian translation in the principal Russian linguistics journal (*Voprosy Jazykoznanija*). The paper was preceded by a general article on MT by Kuznecov, Ljapunov and

---

[1] Revision and expansion of an unpublished presentation at the conference on *Crosslingual Language Technology in service of an integrated multilingual Europe*, 4-5 May 2012, Hamburg, Germany.

[2] Russian characters are transliterated as follows: a=а, b=б, c=ц, č=ч, d=д, e=е, f=ф, g=г, i=и, j=й, ja=я, ju=ю, k=к, kh=х, l=л, m=м, n=н, o=о, p=п, r=р, s=с, š=ш, šč=щ, t=т, u=у, v=в, y=ы, z=з, '=ь

Reformatskij (1956) outlining the basic linguistic problems to be solved in MT, and tracing back ideas on mathematical linguistics, and consequently MT as well, to the eminent 19th century Russian linguist F.F.Fortunatov (1848-1914).

2. Trojanskij

The same issue of this journal included an article by Žirkov discussing the possibilities of MT in the light of the research reported by Kulagina and Mel'čuk. In it, Žirkov (1956) described briefly the pre-computer investigations into mechanical translation by Smirnov-Trojanskij. It had been mentioned briefly in Panov's book, and a few years later Trojanskij's patent and various other papers showing how he intended to develop his system were gathered together and edited by members of the ITMVT group (Bel'skaja et al. 1959; see also Hutchins and Lovtskii 2000.)

Trojanskij's device consisted of a sloping table over which a broad belt could be moved. The belt recorded entries of a large dictionary, words in up to six languages in parallel columns. The operator located a word of the source language and moved the belt to display in the aperture the corresponding word of the target language. The operator would then type in a code indicating the grammatical category or role of the word in question – codes that Trojanskij referred to as 'signs for logical parsing' – and the combination of target word and code were then photographed onto a tape. Then the next source word would be located and 'translated' in the same way. From a tape of the target language words in sequence, a typist would then produce a 'coherent text' for a reviser to substitute the correct morphological forms for each word based on the assigned codes. As a final stage a 'literary editor' would produce the final target text. The 'logical codes' were in effect interlingual (or language-independent) codes, derived primarily from suffixes in Esperanto.

Although Trojanskij's method of mechanising a dictionary for translation could not be implemented on electronic computers, the basic principles were sound enough. In effect he proposed a mechanizable translation process in three stages, and one based to some extent on 'universal' linguistic elements – in so far as Esperanto can be regarded as 'universal' – and he recognized some of the basic problems of translation (homonyms, synonyms and idioms, the problems of analysis into abstract 'symbols', the need for post-editing), and stated clearly some of the major advantages of mechanization (in particular with multilingual output).

In 1939, Trojanskij approached the Academy of Sciences and asked for experts to evaluate its linguistic aspects. There were discussions over the next few years until in 1944 a plenary session was held with experts from mechanics, electronics and linguistics, and Trojanskij's proposal was rejected.

It may be noted that Trojanskij was not the first to outline ideas for a mechanized dictionary: in 1924 in Estonia, a certain A.Bacher (or A.Vakher) proposed to develop a 'typewriter-translator' (*pišuščaja mašina-perevodnik*). It was reported in an Estonian newspaper *Waba Maa* in February 1924 (Waba Maa 1924).

3. Institute of Precision Mechanics and Computational Techniques (ITMVT)

The ITMVT system of 1955 was an improvement on the Georgetown design insofar as it included some morphological analysis (Panov 1956a, Panov 1956b, Panov 1956c, Zelenkevič et al.1956), otherwise its rules were as ad hoc and text-specific as its model. As at Georgetown the subsequent systems incorporated more layers of analysis (noun phrases, verbs, homonyms, etc.), rules of word reordering, separation of dictionary and program, and independent target-language synthesis. The principal researcher at ITMVT was Izabella Bel'skaja, appointed in January 1955. The English-Russian system she developed was similar in many respects to the Georgetown system of the 1950s and 1960s. The first experiments, both run on machine (the BESM computer) and 'manual' simulations, were performed in December 1955. (The achievements up to 1957 were

reported by Bel'skaja (Bel'skaja 1957; see also Nelyubin 1997). The system was also tested 'manually' on other source languages: Japanese, Chinese and German (Panov 1956b, 1956c).

The system described by Bel'skaja (1959) covered applied mathematics and was also tested on literary works; in 1959, the English dictionary contained 2000 words. In a trial of 3000 sentences, few "unknown" words were encountered, and so it was decided to branch into other subjects. There were three basic processes: vocabulary analysis, grammatical analysis, and grammatical synthesis. In the first process, each word of the source texts is assigned a Russian translation and a set of grammatical and semantic 'characteristics'. It included a sequence of tests for homographs and polysemes – similar to the tests in the first ITMVT system. Dictionary analysis ('look-up') was followed by stages of grammatical analysis: verb analysis – in five sections – punctuation analysis, syntax analysis (i.e. identification of word-classes), noun analysis, and adjective analysis. Then came 'change-of-word-order' and a program for target-language synthesis – mainly to derive morphological endings, and intended to be independent of the source language, since the ultimate aim was a multilingual system translating into Russian – and for Russian to be an 'inter-language' (cf. also Panov 1956b, 1956c). The paper gives a detailed example of the analysis of three sentences. There is no doubt about the ad hoc nature of the rules or procedures. Like many MT systems of the time (in the US as well as USSR) there was no theoretical foundation to the system. The system was designed specifically to translate mathematical texts; however, as mentioned above, Bel'skaja believed that the foundations had been laid for more ambitious tasks, in particular the translation of literary texts. Her argument (Bel'skaja 1959: 201) was that:

> the applicability of MT depends on whether it is possible to identify the implicit set of rules governing this or that particular sphere of language applications, be it as narrow a sphere as say, Wordsworth's poetry, and further, on whether these rules can be formulated into a formal set.

Examples of translations from Galsworthy, Poe, and Aldridge were included in her 1959 paper at the IFIP conference (Bel'skaja 1959). Like many researchers at the time exaggerated claims were made – in essence, claiming that the fundamental problems of MT in general had been solved. The ITMVT group "were publicly exposed, and their results recognized as false" (Mološnaja 2000: 228). Shortly afterwards both Panov and Bel'skaja died (the latter in 1964), and the focus of the MT research at the KGB went to Jurij Motorin (Marčuk 2000). The ITMVT moved to other computational areas under a new director, the linguist V.V. Ivanov, who had no interest in MT (Marčuk 2000, Mel'čuk 2000).

4. AMPAR

The work of Jurij Motorin and his colleagues remained largely unknown outside Russia until the English-Russian system which they developed, TYPHOON, was handed over to the All-Union Translation Centre in 1974 (see below, sect.10) and renamed AMPAR. Motorin's laboratory for MT was established in 1956. By mid-1957 a large corpus of social-political articles had been processed comprising almost 5 million words (Nelyubin 1997). These constituted large frequency dictionaries arranged according to word lengths. In 1959 Motorin's team claimed their programme could translate any social-political article and a range of scientific and technical texts. "This algorithm of MT was the first economically efficient one" due in part to "the substantial resources and technical means with which the laboratory was provided" (Nelyubin 1997: 135).

Jurij Marčuk joined the group at about this time. He was a translator in the KGB, and had become interested in MT after seeing reports of the Georgetown experiment. AMPAR (Avtomatizirovannyj Mašinnyj Pervod s Anglijskogo jazyka na Russkij) was explicitly characterised as a one-directional bilingual 'direct translation' system (Marčuk et al.1982). The developers stressed the integration of linguistic data and algorithmic processing and the interlocking of analysis and synthesis procedures – in fact there are no explicit synthesis stages: the rearrangement of source sentences is determined by local lexical and grammar rules. Its 'direct translation' features (similar in many ways to the ITMVT system) are demonstrated by the

seventeen stages: dictionary search and morphological analysis, segmentation (by punctuation marks), idiom dictionary search, treatment of English *–ing* forms, homograph resolution, 'grammatical analysis' to determine information required for synthesis, translation of unambiguous words, translation of polysemes (using a 'contextological dictionary' and morphological and syntactic information from the source text), translation of residual polysemes (by a 'best match' table), generation of Russian morpho-syntactic information (sometimes involving further analysis of the source text, e.g. pronouns, gender of nouns), and morphological synthesis of Russian word forms. There is clearly no phrase structure analysis; no distinctions between levels of analysis (morphological and syntactic information are mixed at many stages). The designers rejected syntactic analysis in favour of lexical context; the system was basically 'dictionary-driven' – in 1975 there were 25,000 English and 35,000 Russian entries in the dictionaries, including a special dictionary for computer science and programming. The contextological dictionary, containing the information for resolving homographs, was compiled by students at the Moscow State Pedagogical Institute for Foreign Languages.

5. Steklov Mathematical Institute

The other Moscow research group founded in 1954 was headed by Aleksej A. Ljapunov, a prominent mathematician, whose enthusiasm for the new science of cybernetics led to the foundation of the Russian journal *Problemy Kibernetiki*, where many of the most important Russian articles on MT were published. For Ljapunov, MT was a subdivision of cybernetics and to be approached as part of an effort to find out what areas of thought processes might be subject to automation. He put a great deal of effort into overcoming the distrust and scepticism of linguists about the use of computers in their research; in the end he succeeded with prominent figures such as P.S.Kuznecov, A.A.Reformatskij, V.Ju.Rozencveijg and I.I.Revzin. For his MT research he recruited researchers from his own organization, the Steklov Mathematical Institute (Olga Kulagina), and by a private arrangement from the Institute of Linguistics (Igor Mel'čuk, who worked for the group without pay; see Kulagina 2000).

Their first effort starting in December 1954 and ready for testing on the STRELA computer in early 1956 was a system for French-to-Russian translation, FR-I (Kulagina and Mel'čuk 1956). In some respects the system was similar to the contemporary Georgetown University system: every French sentence was scanned several times for morphological and syntactic information, each step resulting in the analysis of words belonging to the same grammatical class (verbs, prepositions, nouns, pronouns, etc.), and on the basis of this information the selection of Russian equivalents. FR-I was the basis for more ambitious experiments, including, for example, useful methods for dealing with collocations (multi-word expressions) either as wholes or as separable parts. These included specialised mathematical terminology and 'idioms' such as *le long de*, *parce que* and *mettre en doute*. In the dictionary most nouns and adjectives were entered in their stem forms, some verbs were listed under several stems (*faire*, *fais*, *fasse*, *fe*) in order to avoid listing all forms. Russian equivalents were also entered under stem forms, often several. Morphological synthesis followed grammatical analysis of English.

Kulagina's work on English analysis and dictionary lookup procedures resulted in numerous independent operations. Further study revealed that simplification could be achieved by reducing operations to 17 elementary 'operators' such as checking for morphological information, checking for grammatical information, and marking and deleting procedures. One theoretical result of Kulagina's research was the development of her 'set theory' model of language which formulated definitions of relationships between fundamental concepts: word, phrase, syntagma, sentence, etc. Forming the basis of Revin's mathematical model of language (Revzin 1961), it attracted much attention outside Russia.

The next project at the Steklov Institute was a system for English-to-Russian translation (from 1955), with Tat'jana Mološnaja as chief linguist. When Mološnaja joined the MT group she was at first assigned to help with the program for Russian synthesis (from French). For the English-Russian project the idea was to analyze not by separate words but by word combinations (as with collocations in the French-to-Russian system). As with previous systems the focus was on mathematical texts (Mološnaja 2000). Mološnaja found that English analysis demanded a different approach to Kulagina's for French analysis. Less information could be derived from morphology and more weight had to be given to syntactic analysis. For this, Mološnaja based her approach on the work of Charles Fries (1952) and Otto Jespersen (1937). Mološnaja's programs for analysis were adopted and widely admired by other MT groups in the Soviet Union; however, she herself grew disappointed with the poor results achieved by MT systems and turned increasingly to linguistics, now at the Institute of Slavonic Studies.

In 1967, Olga Kulagina began a new French-to-Russian system (FR-II) based on dependency tree structures within a 'transfer' framework (i.e. a tripartite system of analysis, transfer, and synthesis) of the kind being adopted widely at the time in Western Europe (cf. Hutchins 1986). Nodes of dependency trees could represent words, groups, punctuation marks, or collocations. About 60 syntactic links were identified. Analysis produced all plausible structures, which were then filtered for the most likely ones, each of which were then passed to the Russian synthesis program (Kulagina 2000). The system was tested on some 1500 sentences from mathematical texts and the output evaluated in 1974-75 by experts for comprehensibility and adequacy; around two thirds were considered "good".

6. Institute of Linguistics

After his collaboration with Olga Kulagina on French-to-Russian translation, Igor Mel'čuk turned to the challenging task of Hungarian to Russian. He made extensive detailed analyses of Hungarian and outlined routines for translation (Mel'čuk 1960), following the general procedure in the French-to-Russian scheme. The system was not implemented, since Mel'čuk came to the conclusion that a much deeper level of analysis and an 'interlingual' representation was needed. From these beginnings, Mel'čuk developed over the next decade his 'meaning-text-model' of language in collaboration with Aleksandr Žolkovskij (e.g. Žolkovskij and Mel'čuk 1967). Mel'čuk's conception of an interlingua was quite different from those of others at the time, in particular from the interlingua of Nikolaj Andreev in Leningrad (see below), who conceived an interlingua as a 'real language' with its own vocabulary and grammar. Mel'čuk's conception was an interlingua comprising sets of language-independent 'lexical functions' (such as synonymy, antonymy, etc.) and semantic relations (such as verb and agentive nouns (*write* and *writer*), verb and its causative form (*lie* and *lay*), nouns and their inceptive verbs (*conference* and *open*, *war* and *break out*), nouns and causative verbs (*foundations* and *lay*, *camp* and *pitch*). These relationships within a source-language text form the framework for equivalent relationships in the target language. Such relationships are not a 'language' but a means of representing texts (in any language) at multiple levels: phonology, morphology, surface syntax, semantics, and discourse (Kulagina and Mel'čuk 1967; Mel'čuk 1970; Papp 1966: 119-121). He was criticised for constructing a theory which was quite impractical for MT, and indeed no operational MT system was created on this model. He admits that his first interest had always been the theoretical aspects of linguistics and MT. "For me, MT was a discipline separate from practical implementations on particular machines… I was never really interested in practical MT as such" (Mel'čuk 2000: 216). It was perhaps this attitude, together with his political support of dissidents such as Sakharov, that contributed to his forced emigration in 1977. His linguistics was, however, very influential both in the Soviet Union and in the West (e.g. the CETA group under Vauquois in Grenoble, France), and

he has continued to work on the 'meaning-text-model since leaving Russia; for a comprehensive bibliography see Mel'čuk 2000.

Mel'čuk and colleagues such as Apresjan explored in depth the lexicographic aspects of the model, with the elaboration of an 'explanatory-combinatorial' dictionary designed for the generation of texts from a semantic representation (Apresjan et al. 1969). It was combinatorial in the sense that it showed the potential combinability of lexical items, and explanatory in the sense that it provided semantic interpretations of combinations.

7. Institute of Foreign Languages

Jurij Apresjan worked at the Laboratory of Machine Translation at the Institute of Foreign Languages. The laboratory had been founded by Viktor Rozencveijg, a fluent speaker of many languages, for whom MT was an unofficial interest and who founded the irregular journal *Mašinnyi Perevod i Prikladnaja Lingvistika*, which published many of the most significant papers on MT in this period (Mel'čuk 2000). The other outlets were the *Problemy Kibernetiki* (edited by Ljapunov) and the 'official' linguistics journal *Voprosy Jazykoznanija*. Other MT workers at the Institute of Foreign Languages included Žolkovskij, Leont'eva and Martem'janov.

In 1968 Apresjan refused to endorse the Soviet invasion of Czechoslovakia; he was dismissed and sent to the Institute of Heavy Electrical Machinery (later Institute for Information Transmission Problems), which had its own information service, Informélektro. There, Apresjan (later with Leonid Iomdin and Igor Boguslavskij) developed the ETAP system for translation between Russian and English based on Mel'čuk's 'meaning↔text' model (e.g. Apresjan et al. 1992, Boguslavskij 1995). In this stratificational approach, texts pass through multiple stages: morphological analysis, syntactic analysis, normalization (through the source combinatorial dictionary), transfer, expansion (through the target combinatorial dictionary), syntactic synthesis and morphological synthesis. In recent years, the group has participated in the interlingua foundations of the UNL (Universal Networking Language) project (Boguslavskij et al. 2000).

8. Leningrad State University

At the Leningrad State University, Nikolaj Andreev founded an MT group in 1955, and its associated Experimental Laboratory for Machine Translation (ELAMP) in 1958. The group enthusiastically designed numerous algorithms for bi-directional translation of Russian and languages such as Romanian, German, Norwegian, Serbocroat, Czech, Hindi, Turkish, Arabic, Chinese, Vietnamese, Japanese, Burmese, and Indonesian. Indonesian was Andreev's own speciality, having lectured in it at the university after serving as an officer in the NKVD and KGB from 1941 to 1945 (Piotrovskij 2000).

The distinctive feature of the Leningrad group was its development of the concept of an interlingua for MT. Andreev had been inspired by his studies of Indo-European languages and Esperanto. The interlingua was intended to provide representations for all semantic and syntactic relations in all major languages, whether expressed explicitly or implicitly. The interlingua was to be founded on symbolic logic and take account of combinatorial and probabilistic limitations for all lexical, morphological and syntactic features (Andreev 1967; Papp 1966: 121-123). Lexical items from 'major' languages (Chinese, Hindi, English, Russian) were weighted greater than those from 'minor' languages (Arabic, Swahili, etc.). Likewise syntactic features were weighted (adjective before noun rather than noun-adjective, subject-predicate rather than predicate-subject, predicate-object rather than object-predicate, etc.) It was, therefore, to be a 'language' which brought together all potentially expressible concepts. It was a conception quite different from that of Mel'čuk. There was an attempt in the early 1960s by Andreev and his group to test an Indonesian-to-Russian

translation system. But the attempt failed, as indeed had other Soviet systems at the time (Piotrovskij 2000.)

9. Speech Statistics Group

By the mid-1960s it was becoming clear that the complex linguistics-based approaches to MT were not going to succeed. The ALPAC report (1966), although intended as a survey only of American funding, put an end also to many Soviet projects – the bureaucracy in Russia took the view that if the Americans could not envisage working MT systems in the future, then there was no point in the USSR supporting research in this field. Andreev abandoned the attempt and the ELAMP laboratory became part of the Speech Statistics Group (SSG), founded in 1964 by Raimund Piotrovskij (1980).

The SSG had been established in Minsk by linguists, mathematicians, engineers, psychologist and programmers. There were also branches in Kharkov, Kishinev, Riga, Tartu, Baku and in republics of Central Asia. The focus of the group has been the development of practical MT systems (and various linguistic aspects of Artificial Intelligence). It started with investigations of the statistical foundations of texts. The leading researcher of the group, Raimund Piotrovskij, identified three major discrepancies between human language behaviour and the 'intellectual' capabilities of computers were identified: the divergences between (i) the 'fuzzy' continuity of human language and the discrete (non-fuzzy) nature of computer language, (ii) the open and dynamic character of natural language and the static (deterministic) nature of computer processes, and (iii) the rigidity of computer 'understanding' and the polysemantic (pragmatic) understanding of humans (Piotrovskij 1980, Piotrovskij 2000, Piotrowska et al.2001). The SSG produced simple but practical translation systems for organizations throughout the Soviet Union – such as in Kazakhstan (Chimkent Teachers Training College), Kiev, Kishinev (Moldova), and Samarkand (Marčuk 2000).

10. All-Union Translation Centre

In 1974, much of the MT activity in the Soviet Union was centralised at the All-Union Translation Centre (Vsesojuznyj tsentr perevodov naučno-tekhničeskoj literatury i dokumentacij) under the management of Jurij Marčuk (Kotov et al. 1983). There were two English-to-Russian systems; one was AMPAR (described above) which became operative in 1979 after revision of linguistic rules by Evgeni Lovtskij and reprogramming by Boris Tikhomirov for the new IBM-type mainframe computer. The other was a system developed by Martem'janov under Rozencveijg at the Moscow Institute for Foreign Languages (Martem'janov 1977). A German-to-Russian system (NERPA) and a French-to-Russian system (FRAP) became operational in 1981. All systems were described as 'multi-functional', integrating MT, lexicographical work and an abstracting service.

The NERPA system was based on AMPAR, developed by Tikhomirov and Vlasov, and shared a unified software environment. Its main difference was the considerably greater role devoted to morphological analysis, in order to deal with German compound nouns (Marčuk 1984). The FRAP system was a development of the FR-II system built by Kulagina (see above), i.e. a 'transfer-based' system, expanded to cover electronics, computer science, and aircraft construction as well as its original target of mathematics texts. (The head for the FRAP project was now Nina N. Leont'eva.) One development for all the Centre's systems was the use of 'operands' to facilitate programming by linguists, e.g. checking word forms for specific grammatical categories or morphological features, checking the class of subsequent or preceding word forms, etc.

The main focus of the Centre (Oubine and Tikhomirov 1988) has been the strictly practical one of producing what are admitted to be 'low-quality' translations adequate for information purposes. The centre merged AMPAR and NERPA into a combined modular system ANRAP with

more advanced software. In 1988 the centre was developing a system for Japanese-to-Russian based on the AMPAR model (with the cooperation of the Institute for Oriental Studies). It took over the operations of the SILOD system from the Leningrad group (English, French and Spanish into Russian, a simple (almost word-for-word) system for technical texts requiring extensive post-editing. It assumed operations of the ETAP-2 system for French-Russian translation of a superior quality. Finally, it has conducted basic research on the LINTRAN framework system for automatic generation of MT systems.

The centre has developed translation workstations, and subject-specific translation aids from English to Russian (titles of warehouse patents, word-for-word systems for petrochemistry and for polymer chemistry, and for the Chimkent Pedagogical Institute). In 1989 it launched the first PC system in Russia, MULTIS, a multilingual system based on the MARS automatic dictionary for English-Russian and French-Russian. The system was developed by Larisa Beliayeva, with Svetlana Sokolova and Alexander Serebriakov as programmers (Blekhman 2001.)

11. Operational and commercial systems from the former USSR

Since 1964, an English-to-Russian system has been in use at the USSR Central Scientific-Research Institute of Patent Information, in order to translate patent texts from the US Patent Bureau *Official Gazette* The program comprised 16 sub-routines (including dictionary search, processing of idioms and unknown words, homograph resolution, text segmentation; syntactic analysis of segments (antecedents of pronouns, morphological analysis, noun combinations); and Russian synthesis. The research group was led by Leonid G. Kravec (1967). Roberts and Zarechnak (1974: 2835) reported that considerable post-editing was necessary, which reduced the economic effectiveness of the system.

In 1976, the SSG group in Kazakhstan under K.Bektaev and P.Sadchikova at Chimkent Teachers Training College, developed a system running on IBM-compatible mainframes which performed word-for-word and phrase-for-phrase translations of patent chemistry texts from English into Russian (Pevzner 2001). It operated at the Institute of Chemistry of the Kazakhstan Academy of Sciences.

The commercial MT system STYLUS translation grew out of the MT activities of the SSG researchers in Leningrad. Its first systems (in 1991) developed by Svetlana Sokolova (1997) were for English-to-Russian, Russian-to-English and German-to-Russian business correspondence. Soon further language pairs were added and the range of special dictionaries considerably expanded. The system was transfer-based with multiple levels and using various computational approaches (ATN, finite-state automata, etc.) A major feature of the system was the availability of the bilingual dictionaries to users (purchasers) for updating and revising. By 1998 (Sokolova 1998) the system was available for translation from and to Russian for English, French, German, and Italian, and also German-French and English-French. The PROMT system started from a strictly rule-based approach, but recently it is taking on the form of a hybrid system with increasing use of statistical information. It was and is still the most popular system for Russian-to-English and English-to-Russian translation.

There was only one other commercial system arising from activities of the SSG group; this was the PARS system developed by Mikhail Blekhman in Kharkov, initially for translating medical abstracts from English to Russian. Blekhman followed what Piotrovskij called the ‘engineering approach to language modelling’, the cyclical sequence of: modelling of a language segment, operational test, analysis of results, modification of the model, and so forth (Blekhman 1999a). The first version of PARS was operational in 1989 at the Georgian Medical Information Centre (Blekhman 2001). In 1990 appeared the world’s first bidirectional Russian-to-Ukrainian system (later called PARS/RU), and in 1992 a bidirectional English-Ukrainian system (PARS/U). Over the following years PARS systems for German appeared: German-to-Russian (1994), German-to-

Ukrainian (1998), and German-to-English (1999). Crucial features of the PARS systems are the facilities for user updating and modification of dictionaries, and the large selection of specialised dictionaries. Blekhman (1999a) characterises the PARS systems as being word-for-word with subsequent intensive modification and reordering. By the end of the 1990s, PARS systems were in widespread use by companies in the former USSR states. In the late 1990s the company introduced a series of bi-directional dictionaries as computer aids for translators for the same language pairs, PG-PARS (Blekhman 1999b).

12. MT research outside Russia in the Soviet era

During the 1950s, the Soviet Union was probably more active than the United States. In the summer of 1958, for example, the first All-Union conference on MT was held in Moscow with scientists from Moscow, Leningrad, Gorky, Kharkov, Kiev, Yerevan, Tbilisi, Petrozavodsk, Kazakhstan, etc. There were 340 representatives from 79 institutions. The range of languages studied for translation into Russian was impressive: Arabic, Armenian, Burmese, Chinese, Hindi, Indonesian, Japanese, Norwegian, Vietnamese (Papp 1966: 106). Outside Moscow and Leningrad, however, these seem to have been studies by individuals and did not result in working systems. Harper (1963) considers that 1959 marks a change of emphasis from earlier ad hoc accumulations of rules to a much greater emphasis on close studies of specific language phenomena and the theoretical linguistic foundations of translation – the latter undoubtedly influenced by the inadequacies of equipment. It is in this period that researchers speculated on the formation of an interlingua for MT, indeed a 'language' for facilitating indexing and abstracting as well as translating: an 'information language' (Andreev 1961, 1967)

Before 1990 and the breakup of the Soviet Union and its satellites there was relatively little research activity outside Russia. There were, as mentioned, some attempts to build MT systems in Georgia, Armenia, Kazakhstan and Ukraine. It is reported, for example, that Mel'čuk assisted in a program for the analysis of Georgian (in 1959) and Mološnaja helped colleagues in the Ukraine in a program for Ukrainian (Mološnaja 2000). It is clear that the original enthusiasm did not disappear entirely, although information about it is difficult to find.

13. MT research outside the USSR before 1990

The situation in countries of Central Europe outside the USSR indicates that there was quite substantial activity, although much hampered by the dearth of adequate computer facilities. An algorithm for MT from English-to-Romanian was developed by Erica Domonkos in Romania from 1959 until 1962 to be run on the Romanian MECIPT-1 computer, which had a very limited memory. The program followed a familiar pattern of the time: input of English sentence from punched tape, dictionary search, stemming of not-found words for base forms (e.g. deletion of endings: *-s*, *-ed*, *-ing*, *-ly*, etc.), recognized words assigned to seven categories (noun, adjective, verb, invariable words, conjunctions, homonyms), replacement by numeric code pointing to Romanian translation equivalent, reordering of English words to normal Romanian sequences, replacement of English words by Romanian words, generation of inflected forms, and print-out. The limitations of the computer memory allowed for a dictionary of only 80 words. The experiment in 1962 demonstrated translations of 4 sentences, and a further demonstration in 1965 translated 20 more sentences, with quite good results. An account of the system was published in English (Domonkos 1962, 1966).

In the German Democratic Republic (GDR), an MT group involving Erhard Agricola, Jürgen Kunze, Stefan Nündel, Ingeborg Brand, Gerda Klimonow and Ingrid Starke was active during from 1963 to 1977 at the Deutsche Akademie der Wissenschaften (DAW) (Agricola et al. 1969). The researchers concentrated on English-to-German and Russian-to-German, with a few

small experiments in other languages (e.g. Georgian-to-German, Fähnrich 1970). In-depth studies were undertaken on German prepositional phrases, Russian adverbs, semantic relations, syntactic disambiguation; and an outline program of Russian-to-German was developed (1969) but not fully implemented. The system had various stages of analysis and synthesis (similar in some respects to the AMPAR system), and the designers claimed that German synthesis operated quite independently of the Russian input. From the mid-1980s, the goal of a working MT system was effectively abandoned (mainly through inadequacies of computer hardware), so the researchers turned to more theoretical investigations, e.g. an experimental syntactic transfer-based system for translating German verbs or verbal groups into Russian, focussing particular on problems of tense and aspect (Buschbeck et al. 1991, Kunze et al. 1991) After 1990, there were plans for the MT group to be attached to Siemens' METAL project and a company (Gesellschaft für Multilinguale Systeme mbH) was set up in Grünheide for the further development of the Russian-German system for METAL. – later transferred to SAIL Labs GmbH in Berlin (Höser and Klimonow 1994, Geldbach 1999). The system was so good that it attracted the attention of the German intelligence service (Bundesnachrichtendienst) where development continued. There were plans also for a Russian-to-English system.

One researcher from outside the GDR was also involved in the DAW project during 1976. This was Alexander Ljudskanov, a researcher well known throughout Eastern and Central Europe. As a Bulgarian, Ljudskanov's knowledge of MT came from Russian translations – his mother was Russian, his father a Bulgarian diplomat. He began as a teacher, but then studied translation from a linguistic viewpoint. In 1963 he attended the Fifth International Congress of Slavicists held in Bulgaria. At this conference, Ljudskanov presented papers on MT parsing, and met major MT figures such as Igor Mel'čuk; Bulgarian mathematicians encountered MT for the first time, and as a result an MT group was formed in 1964 in the Mathematical Institute of the Bulgarian Academy of Sciences, comprising a small team of never more than four assistants (Paskaleva 2000). An experimental Russian-to-Bulgarian was began but remained unfinished – most of the work being done in Rozencveijg's MT group in Moscow – using features of the 'meaning-text' model. His most important contribution was to the theoretical aspects of MT formulated within a stratificational framework. His principal work (*Prevežđat čovek't i mašinata*, "Man and machine as translators", 1965) was translated into French and German (Ljudskanov 1969, 1972), but not into English. He maintained close contact with researchers in the GDR (Jürgen Kunze), in Czechoslovakia (Petr Sgall), and in Hungary (Ferenc Papp). Almost his last act was the organization in Varna in 1975 of an international conference, *Application of Mathematical Models and Computers in Linguistics*.

14. Czechoslovakia

Despite the oppressive regime, researchers in Czechoslovakia achieved impressive results before the end of the Soviet era (Kirschner 2000.) In the late 1950s two MT groups were formed at the Charles University in Prague: one at the Center of Numerical Mathematics, the other at the Faculty of Philosophy. The leader and pioneer for MT and linguistic theory in Czechoslovakia was Petr Sgall, whose chief collaborator was Eva Hajičová. Building upon the strong pre-war tradition of Czech formal linguistics, the distinctive functional-generative framework for the explicit description of languages – based on dependency and stratificational relations – was sufficiently developed by 1958 to be applied to MT experiments (e.g. Palek et al. 1964, Konečna et al. 1966, Sgall and Hajičová 1966). The aim was English-to-Czech translation. There were links to Rozencveijg's group and the Leningrad group, and the algorithms for English analysis were based on those of Mološnaja and Kulagina, and later on the predictive analysis of Kuno and Oettinger (1962). The intermediary between source and target was to be the 'deepest' techtogrammatical level of the Czech stratificational model; the aim was a multilingual MT system.

With the arrival in 1960 of a (relatively) powerful new computer SAPO, the first experiments in MT took place, although initially with not very satisfactory results. The ALPAC report in 1966 had no repercussions in Czechoslovakia, partly because of distracting political events in 1968, and partly because the MT researchers were still confident of progress. Nevertheless, the authorities dismissed theory in favour of practical systems; and the Prague group obliged by developing systems of information retrieval (TIBAQ, MOSAIC, ASIMUTH, KODAS) – see Kirschner (2000: 355).

The group were able to return to MT with support from the Canadian TAUM team in Montreal. The Czech group had access to the Q-system software (Panevova and Sgall 1980), and the English analysis program (dependency-based) at TAUM was used for the Prague MT system APAC – Automatic Translation from English to Czech – developed by Zdeněk Kirschner (1982, 1987.) The Czech generation program developed for APAC was later used for RUSLAN, a system for Russian-to-Czech translation (Oliva 1989).

In November 1989 a new era began, with a new approach for MT. The linguistics-based systems of the past were replaced (gradually at first) by corpus-based MT systems (primarily statistical machine translation (SMT)). The researchers in Prague have taken a leading position in this new research paradigm, with Jan Hajič, Ondřej Bojar, Petr Homola, and others making major contributions, particularly to methods of dependency-based SMT models and to systems for closely related languages (from Czech to Latvian, to Lithuanian, to Polish, to Russian, to Slovak, etc., see Hajič et al. 2003, Homola and Kuboň 2004). Finally, the *Prague Bulletin of Mathematical Linguistics* is now taking a leading world-wide role in the publication of SMT papers.

15. Hungary

From Mel'čuk's paper on Hungarian in 1960 until 1990, the contribution from Hungarian researchers was meagre. Otherwise Ferenc Papp (1966) and Ferenc Kiefer (1964) at the Hungarian Academy of Sciences followed closely developments in MT and maintained theoretical studies in mathematical and computational linguistics as contributions to MT modelling, e.g. Papp (1964) at the University of Debrecen where Mel'čuk had held seminars. But no systems were constructed.

After 1990, however, there has been a resurgence with contributions to MT lexicography, translation memory systems, evaluation of SMT, and word sense disambiguation, mainly but not exclusively at MorphoLogic Ltd. and Pázmány Péter Catholic University. The main researchers have been Gábor Prószéky, Márton Mihăltz, László Tihanyi and Attila Novák on the MetaMorpho systems (Prószéky and Tihanyi 2002), which combine example-based and rule-based approaches to MT from English-to-Hungarian but without an abstract 'transfer' representation.

16. Bulgaria and Romania

After the death of Ljudskanov in 1976 there was no significant MT research until Walther von Hahn instigated in 1992 a research project with Galia Angelova (Bulgarian Academy of Sciences, Sofia) on a knowledge-based translator workstation, DB-MAT (Deutsch-Bulgarisch Machine Aided Translation (Hahn 1995; Angelova and Bontcheva 1997). This was continued in 1996 with German and Romanian (DBR-MAT) with Florentina Hristea (Bucharest University) and including students from Prague, St.Petersburg and Cluj-Napoca. Otherwise, more recently, there have been contributions by members of the Bulgarian Academy of Sciences to the development of corpora for Swedish-to-Bulgarian SMT (Iliev and Genov 2012) and to Bulgarian-to-English SMT (Wang et al. 2012).

Walther von Hahn was also instrumental in setting up research projects with Cristina Vertan, firstly on an English-to-Romanian menu-driven translation aid (Vertan and von Hahn 2003) and later on the implications for MT of the Semantic Web. Otherwise, research in Romania has

mainly been in the context of SMT. The chief centre has been the Institute for Artificial Intelligence of the Romanian Academy of Sciences. The focus has been on problems of alignment and of word sense disambiguation, with the main figure being Dan Tufis. Recently, however, there has also been investigation of example-based MT for English to Romanian (Irimia 2009)

17. Poland, Slovenia and the Baltic states

These countries were not contributors to MT research until the 1990s, and even now only on a relatively small scale.

Krzysztof Jassem (2000) has worked over many years on various systems for Polish MT, a bidirectional Polish-English transfer-based system (POLENG), a Russian-to-Polish rule-based system, a bidirectional rule-based Polish-to-German system, and also SMT systems for Polish-to-Italian, French and Spanish. With colleagues Jassem has also worked on machine-aided translation.

In Slovenia, Tomaž Erjavec (Institut Jožef Stefan) has made major contribution to the development of resources for Slovenian and other Central European languages (*JRC-Acquis*, *MULTEXT-East*).

In Latvia, various researchers connected with the Tilde company (e.g. Raivis Skadiņš and Andrejs Vasiļjevs) have made significant contributions to MT evaluation, to the morphology of Baltic languages (Skadiņš et al. 2011), and to the problems of under-resourced languages (Skadiņa et al. 2010), and are developing a cloud-based on-line platform for building MT systems (*LetsMT!*, cf. Vasiljevs et al. 2011). At the Vytautas Magnus University in Lithuania there have been similar efforts on building language resources and on various aspects of SMT. Other systems are Lithuanian-to-English and English-Lithuanian SMT systems developed jointly by the Institute of Lithuanian Language and the Tilde company. In addition, there has been one rule-based system for English-to-Lithuanian based on the PROMT platform (see sect.11).

Likewise, in Estonia there have been studies on parallel corpora and miscellaneous aspects of SMT systems (e.g. Fishel and Kirik 2010). A general investigation of the problems encountered in the translation of Baltic languages is provided by Khalilov et al. (2010). From experiments in SMT for six language pairs (English to and from Lithuanian, Latvian and Estonian) it is concluded is that translation from and into Latvian is the least complex while translation from and into Estonian is the most complicated.

18. Conclusion

From this review it is evident that until 1990 nearly all the mentioned MT systems were either purely experimental with no subsequent operational implementation or, if operational, produced only 'low-quality' output which had to be extensively revised (post-edited) or used only for 'information purposes'. It was an outcome which was not anticipated in the first two decades of MT activity in the Soviet era. For example, the excellent surveys by Papp (1966: 100-126) and by Roberts and Zarechnak (1974: 2839-2851) took largely optimistic views of the then current research in the USSR. During the earlier Soviet period (up to ca.1975) little was known in the West about Russian MT activity, although a few Americans managed to make visits and reported on what they could find or were shown (Harper 1960, Oettinger 1958).

Until 1990 MT research and MT system developments in the Soviet orb were dominated by large institutes (most under the aegis of state academies) within Russia itself. Since then, however, research has been much less concentrated: Russian researchers have declined in numbers while researchers from countries of Central Europe have increased. Partly this is attributable to the change from linguistics-based systems requiring large teams of researchers (computer staff and linguists) to corpus-based approaches (particularly SMT) involving small groups of researchers, often from many different countries. It is also partly attributable to the availability of open-source computer

aids (such as Moses) and easier international communication and cooperation, all facilitating research in small teams and individuals. Research and development is now less bound to particular institutions and countries. Whereas in the past researchers would generally remain many years in one particular organization (usually in their native country), now we see researchers going from one institution to another (often from country to country), seeking and following in most cases the prospects of better funding. It is seen already in Central Europe and the Baltic states and it may be expected to happen elsewhere in Russia and other parts of Eastern Europe.

19. Acknowledgements

I am most grateful to the following for information about systems in Bulgaria, Poland, Russia, Germany, Romania, and Lithuania: Walther von Hahn, Krzysztof Jassem, Evgenii Lovtskij, Gregor Thurmair, Dan Tufis, and Andrius Utka.

20. References

Agricola, E. (ed.) *Automatische Sprachübersetzung*. 4 vols. Berlin: Akademie Verlag, 1969-1977.

ALPAC (1966): *Language and machines: computers in translation and linguistics*. A report by the Automatic Language Processing Advisory Committee, Division of Behavioral Sciences, National Research Council. (Washington, D.C.: National Academy of Sciences.)

Andreev, N.D. (1961): The universal code of science and machine languages. In: Allen Kent (ed.) *Information retrieval and machine translation* (New York: Interscience), 1061-1073.

Andreev, N.D. (1967): The intermediary language as the focal point of machine translation. In: A.D.Booth (ed.) *Machine translation* (Amsterdam: North-Holland), 3-27.

Angelova, G. and K.Bontcheva (1997): Refining domain ontologies for flexible explanation generation. *Recent Advances in Natural Language Processing*, 11-13 September 1997, Tzigov Chark, Bulgaria, 258-263

Apresjan, Ju.D., I.A.Mel'čuk, and A.K. Žolkovskij (1969): Semantics and lexicography: towards a new type of unilingual dictionary. In: F.Kiefer (ed.) *Studies in syntax and semantics* (Dordrecht: Reidel), 1-33.

Apresjan, Yu., I.M.Boguslavskij, L.L.Iomdin, A.V.Lazurskij, V.Z.Sannikov, and L.L.Tsinman (1992): ETAP-2: the linguistics of a machine translation system. *Meta* 37(1), 97-112.

Bel'skaya, I.K. (1957): Machine translation of languages. *Research* 10 (10), 383-389.

Bel'skaya, I. K., L. N. Korolev and D. Yu. Panov, eds. (1959): *Perevodnaja mašina P.P.Trojanskogo: sbornik materialov o pervodnoj mašine dlja perevoda s odnogo jazyka na drugie, predložennoj P.P.Trojanskim v 1933 g*. [The translating machine of P. P. Troyanskii: collection of materials on a translating machine for translation from one language into others, proposed by P. P. Troyanskii in 1933]. Izd. Akad.Nauk SSSR, Moskva.

Berkov, V.P. and B.A.Eršov (1955): O popytkakh mašinnogo perevoda. [On attempts in machine translation] *Voprosy Jazykoznanija* 1955, 6, 145-148.

Blekhman, M.S, O.Bezhanova, A.Kurshin, and A.Rakova (1999a): The PARS family of MT systems – a 15-year love story. *Machine Translation Summit VII,* 13th-17th September 1999, Kent Ridge Digital Labs, Singapore. Proceedings of MT Summit VII "MT in the Great Translation Era", 340-345.

Blekhman, M.S., M.Byezhanova, A.Kurshin, I.Backlan, A.Kazakov and V.Petrov (1999b): Computer-aided translation tools for Russian, Ukrainian and English. *Machine Translation Summit VII,* 13th-17th September 1999, Kent Ridge Digital Labs, Singapore. Proceedings of MT Summit VII "MT in the Great Translation Era", 598-600.

Blekhman, M.S. (2001): General remarks on machine translation. *International Journal of Translation* 13 (1-2), December 2001; pp.65-80.

Boguslavskij, I. (1995): A bi-directional Russian-to-English machine translation system (ETAP-3). *Machine Translation Summit V*: proceedings, Luxembourg, July 10-13, 1995; 2pp.

Boguslavskij, I., N.Frid, L.Iomdin, L.Kreidlin, I.Sagalova, and V.Sizov (2000): Creating a Universal Networking Language module within an advanced NLP system. *Coling in Europe: proceedings of the 18th International Conference on Computational Linguistics*, Universität des Saarlandes, Saarbrücken, Germany, 83-89

Buschbeck, B., R.Henschel, I.Höser, G.Klimonow, A.Küstner, & I.Starke: Limits of a sentence based procedural approach for aspect choice in German-Russian MT. *Fifth conference of the European Chapter of the Association for Computational Linguistics*: Proceedings of the conference, 9-11 April 1991, Congress Hall, Alexanderplatz, Berlin, Germany; pp. 269-274.

Domonkos, E. (1962): Report on the algorithm for automatic translation from English to Romanian (Bucharest: Mathematical Linguistics Commission of the Academy of Peoples Republic of Romania, 1962.)

Domonkos, E.N. (1966): Algorithm for automatic translation from English to Romanian. Bucuresti: Didactic and Pedagogical Publishing House, 1966; 302pp.

Fähnrich, H. (1970): Zur Gestaltung der maschinellen Übersetzung der georgischen Sprache ins Deutsche. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 23 (5), 475-495.

Fishel, M. and Kirik, H. (2010): Linguistically motivated unsupervised segmentation for machine translation. *LREC 2010: proceedings of the seventh international conference on Language Resources and Evaluation*, 17-23 May 2010, Valletta, Malta; pp.1741-1745.

Fries, C. (1952): *The structure of English*. New York: Harcourt, Brace.

Geldbach, S. (1999): Anaphora and translation discrepancies in Russian-German MT. *Machine Translation* 14(3-4), 217-230.

Hahn, W.von (1995): Knowledge-based machine aided translation: the project D-B MAT. *Contemporary Topics in Computational Linguistics, International Summer School*, Tzigov Chark, Bulgaria, 9-12 September 1995; 14pp.

Hajič, J., P.Homola, and V.Kuboň (2003): A simple multilingual machine translation system. *MT Summit IX*, New Orleans, USA, 23-27 September 2003, 157-164

Harper, K.E. (1960): Soviet research in machine translation. *Proceedings of the National Symposium on Machine Translation*, held at the University of California, Los Angeles, February 2-5, 1960; ed. H.P.Edmundson. (Englewood Cliffs, NJ: Prentice-Hall, 1961); pp.2-12.

Harper, K.E. (1963): Machine translation. *Current Trends in Linguistics*, *vol.1: Soviet and East European Linguistics*. (The Hague: Mouton), 133-142.

Homola, P. and V.Kuboň (2004): A translation model for languages of accessing countries. *9th EAMT Workshop "Broadening horizons of machine translation and its applications*, 26-27 April 2004, Malta, 90-97.

Höser, I. and G.Klimonow (1994): Russisch-Deutsch: ein neues METAL-Sprachpaar. *Sprache und Datenverarbeitung* 18 (1), 53-64.

Hutchins, J. and E.Lovtskij (2000): Petr Petrovich Troyanskii (1894-1950): a forgotten pioneer of mechanical translation. *Machine Translation* 15(3), 187-221.

Iliev, G. and A.Genov (2012): Expanding parallel resources for medium-density languages for free. *LREC 2012: Eighth International Conference on Language Resources and Evaluation*, 21-27 May 2012, Istanbul, Turkey, 3937-3943.

Irimia, E. (2009): EBMT experiments for the English-Romanian language pair. In: M.A.Klopotk, et al. *Recent Advances in Intelligent Information Systems*, Kraków, Poland, June 15-18, 2009, 91-102.

Jassem, K., F.Graliński, and G.Krynicki (2000): POLENG – adjusting a rule-based Polish-English machine translation system by means of corpus analysis. *Fifth EAMT Workshop "Harvesting existing resources",* May 11 - 12, 2000, Ljubljana, Slovenia, pp.47-50.

Jespersen, O. (1937): *Analytic syntax*. Copenhagen: Levin and Munksgaard.

Khalilov M., Pretkalniņa, K., Kuvaldina, N. and Pereseina, V. (2010): SMT of Latvian, Lithuanian and Estonian Languages: a comparative study. *Human Language Technologies – the Baltic Perspective. Proceedings of the Fourth International Conference Baltic HLT 2010*, pp. 117–124.

Kiefer, F. (1964): Mathematical linguistics and machine translation in the Computing Centre of the Hungarian Academy of Sciences. *Computational Linguistics* 2, 203-211

Kirschner, Z. (1982): A dependency-based analysis of English for the purpose of machine translation. *Explizite Beschreibung der Sprache und Automatische Textbearbeitung* 9. Prague: Faculty of Mathematics and Physics of the Charles University.

Kirschner, Z. (1987): APAC-3: an English-to-Czech machine translation system. *Explizite Beschreibung der Sprache und Automatische Textbearbeitung* 13. Prague: Faculty of Mathematics and Physics of the Charles University.

Kirschner, Z. (2000): Pioneer work in machine translation in Czechoslovakia. In: W.J.Hutchins (ed.) *Early years in machine translation: memoirs and biographies of pioneers* (Amsterdam: John Benjamins), 349-360.
Konečna, D., P.Novák, and P.Sgall (1966): Machine translation in Prague. *Prague Studies in Mathematical Linguistics* 1, 185-194.
Kotov, R.G., Ju.N.Marčuk, and L.L.Neljubin (1983): Mašinnyj perevod v načale 80-kh godov. *Voprosy Jazykoznanija* 32(1), 31-38.
Kravec, L.G., A.L. Vasilevskij, and A.M.Dubitskaja (1967): Eksperimental'naja sistema avtomatičeskogo perevoda publikacij iz Amerikanskogo patentnogo eženedel'nika "Official Gazette." *Naučno-Tekhničeskaja Informacija* Ser 2, No.1, 35-40.
Kulagina, O.S. and I.A.Mel'čuk (1956): Mašinnyj perevod s francuzskogo jazyka na russkij [Machine translation from the French language into Russian]. *Voprosy Jazykoznanija* 5(5), 111-121.
Kulagina, O.S. and I.A.Mel'čuk (1967): Automatic translation: some theoretical aspects and the design of a translation system. In: A.D.Booth (ed.) *Machine translation* (Amsterdam: North-Holland), 137-171.
Kulagina, O.S. (2000): Pioneering MT in the Soviet Union. In: W.J.Hutchins (ed.) *Early years in machine translation: memoirs and biographies of pioneers* (Amsterdam: John Benjamins), 197-204.
Kunze, J. et al. (1991): [collection of articles, no title.] *Sprache und Datenverarbeitung* 15 (1-2), 3-117.
Kuno, S. and A.G.Oettinger (1962): Multiple-path syntactic analyzer. *Information Processing 1962: proceedings of IFIP Congress 62,* Munich 27 August – 1 September 1962 (Amsterdam: North-Holland, 1963), pp.306-311.
Kuznecov, P.S., A.A.Ljapunov, and A.A.Reformatskij (1956): Osnovye problemy mašinnogo perevoda [Basic problems of machine translation] *Voprosy Jazykoznanija* 5(5), 107-11
Ljapunov, A.A. and .O.S.Kulagina (1955): Ispol'zovanie vyčislitel'nykh mašin dlja perevoda s odnogo jazyka na drugoj [Use of calculating machines for translation from one language into another]. *Priroda* 1955, 8, 83-85.
Ljudskanov, A. (1969): *Traduction humaine et traduction automatique*. Paris: Dunod.
Ljudskanov, A. (1972): *Mensch und Maschine als Übersetzer*. (Leipzig: Niemeyer.)
Marčuk, Ju.N., B.D.Tikhomirov, and V.I.Ščerbinin (1982): Ein System zur maschinellen Übersetzung aus dem Englischen ins Russische. In: H.E.Bruderer (ed.) *Automatische Sprachübersetzung* (Darmstadt: Wissenschaftliche Buchgesellschaft), 319-336.
Marčuk, Ju.N. (1984): Machine translation in the U.S.S.R. *Computers and the Humanities* 18 (1), 39-46.
Martem'janov, Y.S. (1977): Eliminating formal constraints of language. In: *Overcoming the language barrier*: third European Congress on Information Systems and networks, Luxembourg, May 1977 (München: Vlg. Dokumentation), 625-627.
Mel'čuk, I.A. (1960): Machine translation from Hungarian to Russian. *Problems in Cybernetics* 1960, 1, 251-293. [Translated from Russian, dated April 1957.]
Mel'čuk, I.A. (1970): Towards a linguistic meaning↔text model. In: F.Kiefer (ed.) *Trends in Soviet theoretical linguistics* (Dordrecht: Reidel), 33-57.
Mel'čuk, I.A. (2000): Machine translation and formal linguistics in the USSR. In: W.J.Hutchins (ed.) *Early years in machine translation: memoirs and biographies of pioneers* (Amsterdam: John Benjamins), 205-226.
Mološnaja, T.N. (2000): My memoirs of MT in the Soviet Union. In: W.J.Hutchins (ed.) *Early years in machine translation: memoirs and biographies of pioneers* (Amsterdam: John Benjamins), 227-231.
Nelyubin, L.L. (1997): A survey of machine translation in the former USSR. *Perspectives: Studies in Translatology* 5(1), 125-140.
Oettinger, A.G. (1958): A survey of Soviet work on automatic translation. *Mechanical Translation* 5 (3), December 1958; pp. 101-110.
Oliva, K. (1989): A parser for Czech implemented in Systems-Q. *Explizite Beschreibung der Sprache und Automatische Textbearbeitung* 16. Prague: Faculty of Mathematics and Physics of the Charles University.
Oubine, I.I. and B.D.Tikhomirov (1988): The state of the art in machine translation in the U.S.S.R. In: D. Maxwell, et al. (eds.) *New directions in machine translation*: conference proceedings, Budapest 18-19 August 1988 (Dordrecht: Foris), 75-84.
Palek, B., P.Pit'ha, and P.Sgall (1964): Mathematical linguistics in Czechoslovakia. *Prague Bulletin of Mathematical Linguistics* 1, 9-23.

Panevova, J. and P.Sgall (1980): Automatic syntactical analysis of the Czech language. *International Forum on Information and Documentation* 5 (2), 1980; 24-26.
Panov, D.Ju. (1956a): *Avtomaticheskij perevod* [Automatic translation]. Izd. AN SSSR, Moskva. [2nd edition (1958) translated into English, 1960]
Panov, D. (1956b): On the problem of mechanical translation. *Mechanical Translation* 3(2), 42-43.
Panov, D. (1956c): *Concerning the problem of machine translation of languages*. [Moscow]: Academy of Sciences, U.S.S.R. 35pp.
Papp, F. (1964): Mathematical and applied linguistics at the University of Debrecen. *Computational Linguistics* [Hungary], vol.2, 213-218.
Papp, F. (1966): *Mathematical linguistics in the Soviet Union*. The Hague: Mouton.
Paskaleva, E. (2000): Alexander Ljudskanov. In: W.J.Hutchins (ed.) *Early years in machine translation: memoirs and biographies of pioneers* (Amsterdam: John Benjamins), 363-376.
Pevzner, B. (2001): First steps of language engineering in the USSR: the 50s through 70s. International Journal of Translation 13 (1-2); pp.81-86.
Piotrovskij, R.G. (1980): Machine translation in the research group on Speech Statistics: theory, results and outlook. *International Forum on Information and Documentation* 5 (2), 1980; pp. 9-11.
Piotrovskij, R.G. (2000): MT in the former USSR and the newly independent states (NIS): prehistory, romantic era, prosaic time. In: W.J.Hutchins (ed.) *Early years in machine translation: memoirs and biographies of pioneers* (Amsterdam: John Benjamins), 233-242.
Piotrowska, X., R. Piotrowski, Y.Romanov, N.Zaitseva, and M.Blekhman (2001): Machine translation in the former Soviet Union and in the New Russia. *International Journal of Translation* 13 (1-20, Jan-Dec 2001; pp.87-104.
Prószéky, G. and L.Tihanyi (2002): MetaPorpho: a pattern-based machine translation system. *Translating and the Computer 24*: proceedings from the Aslib conference held on 21-22 November 2002 (London: Aslib). 7pp.
Revzin, I.I. (1961): *Modeli yazyka*. Moskva [Translated as *Models of language* by N.F.C.Owen and A.S.C. Ross (London: Methuen, 1966)]
Roberts, A.H. and M.Zarechnak (1974): Mechanical translation. *Current Trends in Linguistics*, vol.12 (The Hague: Mouton), 2825-2853.
Sgall, P. and E.Hajičová (1966): Computational linguistics in Czechoslovakia. Beiträge zur Linguistik und Informationsverarbeitung 8, 1966, 65-74.
Skadiņa, I., A.Vasiļjevs, R.Skadiņš, R.Gaizauskas, D.Tufiş, and T.Gornostay (2010): Analysis and evaluation of comparable corpora for under resourced areas of machine translation. LREC 2010: *Proceedings of the 3rd Workshop on Building and Using Comparable Corpora,* Malta, 22 May 2010; pp.6-14.
Skadiņš, R., K.Goba, and V.Šics (2011): Improving SMT with morphology knowledge for Baltic languages. *Machine Translation and Morphologically- rich Languages*: Research Workshop of the Israel Science Foundation, University of Haifa, Israel, 26 January, 2011; 3pp.
Sokolova, S. (1997): STYLUS – outline of functions and features. *MT Summit VI. Machine Translation: Past, Present, Future.* Proceedings, 29 October – 1 November 1997, San Diego, California, USA, 257-258
Sokolova, S. (1998): ProMT – the current state. *Translating and the Computer 20*…12&13 November 1998 (London: Aslib); 7pp.
Vasiļjevs, A., R.Skadiņš, J.Tiedemann: LetsMT!: cloud-based platform for building user tailored machine translation engines. *MT Summit XIII: the Thirteenth Machine Translation Summit* , 19-23 September 2011, Xiamen, China; 507-511.
Vertan, C. and W.von Hahn (2003): Menu choice translation: a flexible menu-based controlled language system. *Controlled Language Translation, EAMT-CLAW-03*, Dublin City University, 15-17 May 2003, 194-199.
Waba Maa (1924): Perevodnaja mašina v Talline v 1924 goda. In: *Soobščenija po Mašinomu Perevodu* (Tallin: Eestinsu Teaduste Akadeemia: Keele ja Kirjanduse Instituut, 1962; 1, 95-96) [Translated from *Waba maa*, February 1924.]
Wang, R., P.Osenova and K.Simov (2012): Linguistically-augmented Bulgarian-to-English statistical machine translation model. *EACL Joint Workshop on Exploiting Synergies between Information Retrieval and Machine Translation (ESIRMT) and Hybrid Approaches to Machine Translation (HyTra)*, 23-24 April 2012, Avignon, France, 119-128.

Zelenkevič, G.P., L.N.Korolev, and S.N.Razumovskij (1956): Opyty avtomatičeskogo perevoda na élektronnoj vyčislitel'noj mašine BESM. [Experiments in automatic translation on the electronic calculating machine BESM] *Priroda* 1956, 8, 81-85.

Žirkov, L. I. (1956): Granicy primenimosti mašinnogo perevoda [The limits of applicability of machine translation]. *Voprosy Jazykoznanija* 5(5), 121–124.

Žolkovskij, A.K. and I.A. Mel'čuk (1967): O semantičeskogo sinteza. *Problemy Kibernetiki* 19, 177-238. [Translation: Sur la synthèse sémantique. *T.A.Informations* 1970, no.2, 1-85.]