[From: *International Journal of Translation*, vol.13, no.1-2, Jan-Dec 2001]

First Steps of Language Engineering in the USSR: The 50s through 70s

BORIS PEVZNER

Language engineering was one of the most interesting fields of science and technology in the former Soviet Union. However, few people (if any) know details of its development. The authors of the present paper have been and are active in machine translation, computational lexicography, information retrieval, automatic abstracting and indexing. Boris Pevzner began practical work in language engineering in the 60s, while Michael Blekhman, Dr. Pevzner's pupil, became a researcher in mid 70s. Thus we can provide analysis “from inside”.

In 1954, a year after Joseph Stalin's death, the Soviet Union “rehabilitated” cybernetics, which had been considered an “imperialist” science before that. The beginning of the new era in science was marked with Acad. Solodovnikov's public lecture on cybernetics given in the Polytechnic Museum in Moscow. Several more years passed since then before those fundamental ideas laid in the foundation of cybernetics were understood and appreciated by the young generation of scientists.

One of the results of this understanding was the appearance of Russian translations of classic books on cybernetics and applied mathematics. The translations were made by the leading Soviet researchers: V.Cherniavsky, D.Lakhuti, A. Yesenin-Volpin, and some others. Papers on theoretical issues of machine translation were published regularly in the issues of “Masshinnyi perevod” (‘Machine Translation’). They presented translations of papers by the most prominent linguists and cyberniticians, Noam Chomsky among them. Centers of research and development in this field were the All-Union Institute for Scientific Information - VINITI (headed by

A.Mikhailov) and Laboratory for Electric Modeling (headed by A. Vasilyev).

Original Russian publications appeared at the same time, the most widely readable edition being “Problemy kibernetiki” (‘Problems of Cybernetics’). It was there that various aspects of automatic text processing were discussed for the first time in the Soviet Union. In particular, papers by O.Kulagina and N.Moloshnaya covered some practical issues of machine translation.

In the 50s and 60s, I.Belskaya suggested a detailed algorithm of automatic English-Russian translation. The algorithm was published as a large two-volume book.

All those break-through ideas were not tested on representative text corpora, however. An illusion existed that they were absolutely logical and non-contradictory and would lead to quick and efficient problem solution. As one of the linguists pointed out, having large dictionaries and detailed traditional grammars, one only needed a powerful computer to develop real-life text processing systems.

In the late 50s, L.Gutenmacher published one of the first Russian monographs on cybernetics – “Information Retrieval Systems”. He discussed various aspects of information retrieval, such as software, hardware, and communication through telephone channels. Gutenmacher foresaw the time when one would be able to access remote libraries from home via telecommunication channels (a prototype of Internet).

G.Lesskis, a Moscow-based linguist and literature researcher, carried out a deep statistical analysis of a large fiction text corpora and found out an extremely interesting phenomenon: the lengths of sentences vary depending on the narration phase: introduction, culmination, and epilogue.

In the early 60s, two large scientific conferences on automatic text processing were held in Moscow: ‘Problems of Semiotics’ at the Institute of Foreign Languages, and ‘Problems of Automatic Information Retrieval’ at VINITI. Many researchers took part in both. One of the sessions at the former was presided by Acad. A.Markov, the prominent scientist who created an algorithmic model, descriptive enough for linguistic analysis. The algorithm was formulated in terms of word conversion (the Markov algorithm). A broad spectrum of problems was approached dealing with algorithmic text processing:

- semiotic analysis of fiction, formalization of the ties between the characters;
- formalisms describing sense relations in texts;
- morphological, syntactic, and semantic analysis;
- formal languages, both well-known and "exotic" ones, such as the language of the cinema (S.Genkin) and many other problems.

The conference on automatic retrieval approached both theoretical and purely practical issues. Quite a number of contributions were made by young researchers who worked in the laboratories headed by G.Vlaeduts and V.Cherniavsky.

One of the most important events was developing the world's first personal computers of the *Mir* series at the Institute of Cybernetics of the Ukrainian Academy of Sciences in Kiev. Some time later, Acad. V.Glushkov, director of that institute, published his breakthrough monograph named "Paperless Informatics" suggesting electronic data processing and exchange (a prototype of modern information technologies).

In the mid-60s, two exhibitions on information technology were held in Moscow. The contributions, made by the members of the Council for Economic Cooperation (socialist countries), embraced operational models rather than commercial products: 2nd generation computers *Minsk* and *Ural*, photocopying machines, office stationery, information retrieval systems based on those machines. Some stands displayed Western technology, although no Western representatives attended the exhibitions.

One of the most remarkable events was Norbert Wiener's visit to Moscow and his lectures at Moscow State University. The great scientist was met with understanding and enthusiasm from the young Russian students and experienced researchers. He appreciated the high professional level of the audience.

Also in the 60s, VINITI began publishing what became the most authoritative and respectable Soviet journals in the field of information processing and language engineering, the monthly “Nauchno-technicheskaya informatsiya” (‘Scientific and Technical Information’). Since the first issue, its readers and contributors have been both practical researchers and theoreticians from all over the USSR.

At the same time, a theoretical department “Semiotics” was created at VINITI. Its main fields of research were formal grammars, multiple-valued logics, information retrieval, and formalized methods of text processing. A scientific council and a postgraduate center for scientific and technical information were established at VINITI at the same time.

In the same period of time, a fundamental monograph was published by A.Mikhailov, A.Chernyi, and A.Gilyarevskyi: “Foundations of Informatics”. It became a textbook for thousands of students in the Soviet Union.

In the mid-60s, information technologies gained a great momentum in the USSR. All-Union, Republican, departmental, and local information centers were set up. Their main task was creating, processing, and distributing textual and factual databases as well data processing technologies. The databases comprised millions of entries. Search mechanisms were based on keywords and topic categories. The All-Union system of scientific and technical information was unprecedented in the world. It had a layered multidivisional structure and served thousands of institutions all over the USSR. Data exchange was carried out using magnetic tapes. Telecommunications were also introduced.

MACHINE TRANSLATION IN THE USSR

This topic deserves special discussion. Due to the lack of space, however, we will only shed light on a few points of thousands.

The Moscow State University group headed by O.Kulagina and N.Moloshnaya developed algorithms for automatic morphological and syntactic analysis.

At VINITI, a group led by Yu.Shreider developed algorithms for recognizing proper names in real-life texts. The algorithms were based on calculating distances between words in the texts. The group focused attention on theoretical aspects of machine translation.

Researchers led by I.Mel'chuk at the Institute of Foreign Languages pioneered in using a semantic component in machine translation. That component was introduced as a explanatory-combinatorial dictionary. Each word-entry in the dictionary was made according to a universal scheme using the so-called lexical and logical functions. Such an approach was supposed to provide complete description of the word-entry including phrases and idioms.

Research and developments in the machine translation area were also carried out at the All-Union Patent Institute, Leningrad State University, and some other centers.

The most outstanding person to mention is, to our opinion, Raimund Piotrowski, professor at the Leningrad Pedagogical Institute, a man whose role in Soviet language engineering has been really great. He is both a brilliant linguist and a very energetic organizer. In the early 1970s, he founded the All-Union linguistic group, which he called “Statistica Rechi” (‘Speech Statistics’). It united language engineers from all over the USSR: Leningrad, Moscow, Ukraine, Kazakhstan, Moldavia, Uzbekistan, Azerbaijan, etc.

The first operational Soviet MT system was developed in 1976 at Chimkent Teachers Training College in Kazakhstan, by the Kazakhstan subgroup Speech Statistics headed by Prof. K.Bektayev and Prof. P.Sadchikova. The system ran on IBM-compatible mainframes and performed word-for-word and phrase-for-phrase English-Russian translation of patent chemical texts. The system was used at the Institute of Chemistry, Kazakhstan Academy of Sciences.

Piotrowski's Moscow colleague, Prof. Yuri Marchuk, Director of the All-Union Center for Translations, headed a project covering 3 MT systems: English-Russian (AMPAR), (German-Russian (NERPA), and French-Russian (FRAP). The AMPAR system was launched in 1977. It was used for generating raw translations of technical texts both at the Center and at some departmental research institutes. Marchuk published a 2-volume English-Russian contextual dictionary that can be used (and I plan to use it!) for disambiguation purposes. Yevgeni Lovtski developed a special language for representing linguistic rules in AMPAR. Doctors Boris Tikhomirov, Zoya Shaliapina, and Nina Leontieva investigated into various aspects of semantic-based and transfer-based MT. I believe that Zoya was the best expert in Japanese-based MT in the USSR.

Boris Pevzner published in the early 70s a series of papers on example-based text processing.

The 70s were a period of scientific confrontation of two conceptions: the practical ("engineering") approach to machine translation, most vividly expressed by Raimund Piotrowski, and the theoretical approach, backed by such outstanding linguists as Igor Melchuk and Yuri Apresian. They opposed the idea of automatic translation to Piotrowski's machine translation, and argued that the linguist's task is to offer an in-depth description of the language as the foundation of an AT algorithm instead of gradual improving an imperfect MT system. Apresian's group developed the ETAP family of pilot MT systems translating from French and English into Russian. It's interesting that the word-for-word English-Russian translation module was used for translating patent titles in the INPADOC patent information retrieval system.

However, it was in the 1990s, with the advent of personal computers, that machine translation was made accessible to hundreds of thousands of end users. Would it be possible without the first steps made by the pioneers?