

PERESTROIKA AND MACHINE TRANSLATION IN THE USSR

Last month, some politicians changed a law, a businessman made a speech, and Electric Word got a fax from Russia. All three events illustrate the gaps that exist between reality and preconception, the talk and the action in the West's reaction to post-perestroika Eastern Europe.

In the same week that President Bush consulted politicians about what to do about "high tech" trade with the Eastern Bloc, in a little-reported speech, Olivetti's managing director Vittorio Cassoni expressed his belief that "both local-area and wide-area networking will provide the underlying foundation for the unification of Europe."

Contrast this logic with that of the US government's decision to relax controls on the export of computers and other high technology equipment to Eastern Europe – excluding the Soviet Union.

US State Department spokesperson Margaret Tutwiler explained why the USSR was excluded from the COCOM regulation relaxation. "The situation has not evolved to the same extent in the Soviet Union – which retains a formidable strategic capability that still poses a substantial threat to the West," she said.

What is the strategic capability of a country whose infrastructure has perhaps months to survive? How much less substantial a threat would it be if it were stable?

Cassoni the businessman pointed out that the USSR differs from the rest of Eastern Europe in having vast natural resources it can trade for hard currency with which to buy the computers and phone lines to create the LANs and WANs that could give it the stability.

In other words, if politicians got off the computer industry's back now instead of in two years time, the USSR might have a fighting chance of rectifying the fact that it has about 500,000 PCs (or less than one per cent of the 60 million in the US), is about ten years behind the West in hardware technology, has 12 phone lines for every 100 people (compared to the USA's 51), and has dodgy switching and transmission equipment.

The lucky few with access to computers are reputed to have a very good knowledge of assembly and machine code. The first most people heard about Russian programming skills was last year when Alexey Pajitnov, a researcher in AI and sound/voice recognition at the Academy of Sciences Computer Department, developed Tetris – one of the most popular (and non-violent) computer games in recent years.

Which brings us to the EW fax. As their space program has shown, when the tools and materials are made available to Soviet scientists, Great (and very rapid) Leaps Forward can occur. Such a leap could come in machine translation.

TRANSLATION BUREAU

"Soviet machine translation" may conjure up images of gaunt chess-playing scientists sweating over steam-powered mainframes. Indeed, the Institute for Oriental Studies, according to the Japanese-to-Russian MT specialist there, Zoia Shalyapina, "has no money and only managed to get its own computers (Olivetti PCs and a Soviet mainframe) 18 months ago." Before then, work had been "purely theoretical."

But the USSR Center for the Translation of Scientific and Technical Literature and Documentation, aka *Interperevod* (VCP), has got computers – and the use of phone, telex, fax, e-mail, and DHL courier as means of communication. (Having found it impossible to keep a phone connection up for longer than a minute, Electric Word opted for the fax.)

Given the tools, VCP has knitted a range of service, research, and commercially available MT software into an impressive translation bureau and its own range of commercially available MT software – albeit ten years behind the West.

Created in 1956 as the translation bureau of the USSR Academy of Sciences, VCP is now the biggest translation center in the Soviet Union. Last year it translated

over two million pages of scientific and technical documentation. It also provided interpreting for conferences, negotiations, and seminars in 30 foreign languages into Russian and vice versa.

VCP has already successfully developed and put on the market a series of its own MT products, based around specialized technical subjects such as computing, electronics and engineering.

One such is ANRAP, which comes in two language pairs: English-Russian (ANRAP-A) and German-Russian (ANRAP-N). Both run on Soviet-made EC-type mainframes. ANRAP-A is offered with integrated dictionary-updating facilities, which can be operated by users unfamiliar with the peculiarities of MT.

VCP director Oleg Doubovov says the system can grind out 90 pages per hour, including input and printout time, of a quality "that makes it possible to use it for indicative translation." Since 1986, ANRAP has been used on a commercial basis and translated 100,000 pages of text.

VCP's dictionary research has also produced a multilingual dictionary with language-independent software. This has mainly been used for the partial automation of lexicographic research, allowing the continuous acquisition of new terms and easier update of translation dictionaries. It contains 250,000 words in English, German, Russian, and French, has several subject fields, and offers online updating and reversibility – that is, queries in any language.

SPAS

In partnership with SO DZU in Bulgaria, VCP also markets interactive computer-assisted translation software for AT/compatibles, called SPAS, which VCP claims doubles the production of a human translator.

Harddisk-resident SPAS (a Russian acronym for "Translation System with the Help of the Automatic Dictionary") is based around a core 20,000-word automatic dictionary, specialized in scientific domains, though VCP says dictionaries can be customized.

Text is input by the user, and unknown words marked. Following a lookup wait of three seconds, a translation or choice is presented, the user's choice marked and sent for output to the printer or communications network. A screen editor allows the user to correct spelling and grammar mistakes, text structure, and format. There is also a module which allows queries in the form in which they occur in the translated text.

SPAS comes as a PC standalone product (for use in a Soviet-built EC 1832, based around Intel's 80C88 16-bit microprocessor), or a standalone harddisk for

connection to an AT, or as a set of diskettes for a PC or network.

So far, SPAS handles English and Bulgarian to Russian, but this month, VCP plans to release a German-Russian/Russian-German SPAS translator's workstation with automatic dictionaries to run on XT/compatibles. And similar systems are coming for Spanish (September 1990) and French (end of the year) with subject fields for computing, commerce, and common usage. Also at the end of the year, SPAS 2.1 – a souped-up English-Russian model with a dictionary lookup rate of ten words per second – is expected for the computing, technical terms, and common usage domains.

LINTRAN

Another VCP MT system is Lintran English-Russian. Targeted for market at the end of the year, Lintran has an English wordstock of 30,000 words and 25,000 phrases, plus enough Russian words (35,000) and phrases to cover the source dictionary. Lintran uses its own programming language for coding linguistic algorithms. It also performs context analysis and translation of polysemic words algorithmically. This system will also run on the EC-type mainframes (under the OC EC operating system). VCP expect a translation rate of 150,000 characters per hour when run on senior ECs.

Doubovi says the USSR Translation

Center is looking to cooperate with foreign firms in the following areas:

- exchanging lexical bases for different language pairs, including source dictionaries of stems and/or word forms, and bilingual dictionaries on machine-readable or conventional media;
- exchanging linguistic algorithms for text processing, and information on MT and automatic dictionaries, including reciprocal invitations to conferences on automatic text processing.

Interperevod Firm, 14 korp. 1, Krzhizhanskogo ul, 117218 Moscow, USSR. Tel: 124 70 24. Fax: (095) 310 70 24. Telex: 411618 VCP SU.