First-Perso
Singula

Language International *talks with Michael Blekhman, trans*
and designer of the Pars MT (Machine Translation) Sys


Michael Blekhman

As founder of the machine translation company Lingvistica '93 and creator of the Pars family of MT systems, what are the Russian/Ukrainian roots in MT development?

Michael Blekhman: Machine translation dates back to the 1950s in the ex-Union. I graduated from the Kharkov State University in 1974, where I studied under Raimund Piotrowski. In the early '70s, he founded the All-Union linguistic group which he called "Statistica Rechi" (Speech Statistics).

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

Piotrowski's Moscow colleague, Yuri Marchuk, director of the All-Union Center for Translations, headed a project covering three MT systems: English-Russian, German-Russian, and French-Russian. The English-Russian system was launched in 1977. It was used for generating raw translations of technical texts both at the Center and at some departmental institutes.

In the early 1970s, my old friend and teacher, Boris Pevzner, published a series of papers on example-based text processing, which I consider revolutionary. The Pars "distant phrases" feature has a lot to do with Pevzner's ideas.

The Leningrad subgroup of Speech Statistics, headed by Raimund Piotrowski and Larisa Beliayeva, launched an integrated language-engineering project in the '80s that included:

- Multis, a multilingual-MT system based on what Larisa called Mars—a multi-aspect. Russian automatic dictionary (my Pars system includes a grammatical Russian dictionary that resembles Mars to some extent); the main language pairs were English-Russian and French-Russian.
- A system for automatic topic recognition preceding machine translation of information messages.
- The automatic abstracting of information messages, in which I was involved. This system was used by a large governmental analysis bureau that processed hundreds of such messages every day.

Multis was the first Soviet PC-MT system and became operational in 1988-1989 with Larisa Beliayeva as the designer, and Svetlana Sokolova and Alexander Serebriakov as the programmers. Multis implemented several basic ideas put forward by Piotrowski back in 1971. One of them was to assign a single generalizing translation to each polysemic word, instead of several separate translations (as is done in Pars).

Piotrowski's main idea was what he called the "engineering approach" to language modeling. He argued that developing an MT system is a complicated process consisting of numerous stages. The linguist models the text, implements it in an operational (not hypothetical) program, analyzes the results, modifies the model, and so on, thus "growing" the system up from diapers to a full suit. That's exactly what we have been doing to the Pars systems for almost 10 years now.

As for Pars itself, the first version was launched in 1989 and implemented at the Georgian Medical Information Center for generating raw translations of the Medline database abstracts.

"YOUR PROGRAM TRANSLATES

LIKE A STUDENT."

"A FRESHMAN OR A SENIOR?"

"LIKE A SOPHOMORE."

"YOU FLATTER ME!

A STUDENT IS AT LEAST HUMAN!"

But it was the arrival of the PC in the 1990s that finally made machine-translation systems such as Pars accessible to hundreds of thousands of end-users, professional and otherwise.

You are a practicing translator—you translated Alice's Adventures in Wonderland, *for instance. How has your translator's experience and knowledge influenced the design of Pars (and its ongoing developments), compared with purely "computer-science" type MT designs?*

Michael Blekhman: Being a translator is not necessarily an advantage in developing an MT system. Translating is art, and you can't train anyone, especially a computer, to be an artist. However, when I worked as technical translator at the VNII Telektromash Research Institute, I was able to successfully translate material I didn't really understand because I did act like a computer program: I simply substituted Russian words for the English ones and put them in the proper morphological forms.

Yet being a human translator is a real advantage for a machine-translation design. I knew that my colleagues would be disappointed with the childish mistakes that Pars would make. I even thought that they would prefer extensive electronic dictionaries to a full MT program. So, I am always happy when Pars gets praise from a translator. And paradoxically, it is always the most skilled and experienced translators who find Pars useful in their everyday work.

When I presented Pars at the Antonov Aviation Plant in Kiev and asked a translator his reactions, he replied, "Well, it translates like a student." "A freshman or a senior?" I countered. He thought for a moment and said, smiling, "Like a sophomore." He meant that the translation was quite understandable but rather primitive. "You flatter me!" I replied. "A student is at least human!"

Where my translator's experience helps in designing Pars is in the service options rather than in the translation algorithms. For example, one of the major features of our systems is their target-text, post-editing facilities. The unique "pen" editor currently being developed by a team headed by Alexander Kazakov provides hot keys for the most typical editing operations performed by professional translators.

Another special feature is the dictionary-updating subsystem. I am really happy to hear translators say that Pars is user-friendly in this regard, and that they can create dictionaries of their own reflecting their professional experience.

My elder daughter uses our MT systems together with the Polyglossum scientific and technical dictionaries to do professional translations. She claims that a professional can't do without MT and machine-assisted translation if they want to be competitive. She post-edits 30 pages of technical Russian-to-English MT output per day. If the subject area is covered by Pars specialist dictionaries, editing the raw translation is vastly easier than translating the same text manually.

Deep in my heart, I suspect that machine translation as a scientific task is a mathematical problem. But my practical experience tells me that an operational MT system can only be designed by a working linguist. I don't think I would be able to head an MT team if I were a mathematician rather than a linguist and translator.

DEEP IN MY HEART, I SUSPECT THAT MACHINE TRANSLATION AS A SCIENTIFIC TASK IS A MATHEMATICAL PROBLEM. BUT MY PRACTICAL EXPERIENCE TELLS ME THAT AN OPERATIONAL MT SYSTEM CAN ONLY BE DESIGNED BY A WORKING LINGUIST.

What are the key priorities in developing an MT system?

Michael Blekhman: The dictionary-updating and compiling tool is one of the most important features of an operational MT system. Boris Pevzner taught me that I should only develop technologically efficient systems, i.e., systems that could be easily tuned to as broad a user-base as possible.

A flexible dictionary-updating program is sometimes an even more important condition for professional translators than translation quality itself. An MT system is made by linguists and programmers for people who have nothing to do with linguistics and programming (translation and linguistics are two different things), and we have to develop a procedure suggested by the program that is natural and understandable.

Let me compare this with translating *Alice's Adventures*. When I came across a pun or some specific English expression, I did my best to find a translation, but I never included it in the final variant until I offered it to Olga, who was seven then—Alice's age. I never told her I was translating something—she didn't have to know about my technical problems. I simply offered her my Russian joke, and the only criterion was whether she smiled or not. My rule is if a user smiles when entering new words into the dictionary, then your design is OK.

Together with ETS Publishers, we have designed a "conveyer-belt" technology for inputting new and existing dictionaries. We are currently converting the world's largest specialist English-Russian and German-Russian bidirectional dictionaries into Pars format.

Having a single start-up client is another *sine qua non*. One of my principles is to have a definite user in mind when developing an MT system. This principle dates back to 1980, when I began developing an information-retrieval system. It was hard to convince my experienced, yet conservative, bosses to finance the work, but I gained the welcome support of Vladimir Terletsky, head of the powder-metallurgy laboratory, in trying to make the system technologically efficient. So my principle is: develop your system for someone you know very well, and there is every chance it'll be accepted by many others.

I believe MT design should be about setting both quality and quantity targets. But since great pilot systems do not always

(continued on page 46)

Blekhman (continued from page 39)

scale up to industrially efficient systems, I prefer compromises. A simpler, yet larger system is better for the end-user than a smaller, more complicated MT system. You can use and improve a simpler system. But no one will ever use a complicated system, however good it appears to be in a demo.

Finally, you have to think about computer platforms. As soon as we come up with a new version of a translation system, "those Americans" invent something else, so we have to update our products. Since 1994, we have moved successively from DOS systems to Windows 95 and NT. All Parses are linked to MS Word 6.0, Word 7.0, and Office 97. And people from all over the world still keep asking me: "Can't you make your systems compatible with Unix or Macintosh or Sun?"

At present, Windows 95 and NT, with Word 7.0 and Office 97 are very widely used around the world, including Ukraine and Russia, which makes our life easier. There are no problems with fonts, and Microsoft Tool Kits simplify compatibility.

How do you read future prospects for the market for MT systems in Russia, Ukraine, and elsewhere?

Michael Blekhman: The market in Ukraine and Russia will depend largely on the governments. If they do not make a serious effort at prosecuting computer pirates, then most MT companies will simply disappear. These days, you can find all kinds of pirated CD-ROMs in almost any Ukrainian or Russian computer shop. They are even available on the Web. One of the latest is "Flint's Treasures" which has a picture of an ugly pirate taking his treasures out of a sack—hundreds of thousands of copies of MT systems such as Stylus and Pars!

But more important to the future of MT in general is training qualified language engineers. They are simply not trained at Ukrainian universities. Language engineering as a scientific discipline is not listed among the specialties of the Ukrainian Highest Qualifications Commission. This makes it extremely difficult to defend dissertations on language engineering in this country.

I hope that there will be some support for my initiative to organize a language-engineering department at Kharkov Slavonic University. I also plan to organize an international student language-engineering group, whose practical goal would be to develop a new generation of Parses (plus new language pairs), to be used by governmental organizations in Europe and the Americas. I keep dreaming of an organization with financial backing from some of those European foundations.

Contact

www.polyglossum.com, www.jourist.de