

MT Systems from Russia: A Hot Find?

With the dismantling of the Soviet military complex, everything from night binoculars to tanks are on the auction block. You can even find an MT system or two.

A&X Automatisering, a small software house in Zwijndrecht, Holland, recently obtained exclusive worldwide marketing rights to a PC-based machine translation system developed in Russia. While the company's documentation studiously avoids specifying the precise origins ("the Russian Universities"), A&X's director Herman van der Meer reveals the system, called SILOD, originated at the Herzen Pedagogical University in St. Petersburg, where it was developed by Prof. Raymund Piotrovskij. Piotrovskij and colleagues have reputedly been working on the system for more than thirty years, and Van der Meer believes development of SILOD was largely funded by the KGB. One intriguing aspect of the system is the surprisingly large number of language pairs which are available for it, albeit in various stages of development. For source languages, these include all of the official EU languages (except Dutch and Danish), and Finnish, Turkish, Hebrew, Arabic, Japanese, and Chinese. In addition, there is also a Russian into English version. It is encouraging that SILOD's developers tackle this mammoth task domain; less heartening is to see how broadly some of domains are defined: they include "informational" (apparently newspapers in KGB,speak), "business," "technical," "medical," and "nuclear power station construction" (we dearly hope this last one has been rigorously compiled). All in all, the surprisingly large number of language pairs and a highly informal impression of the system's linguistic capabilities lead us to the conclusion that SILOD still profoundly reflects the bulk text scanning proclivities of its original sponsors.

Van der Meer, who has ties to Russian (his wife is Russian), has an apartment in St Petersburg, and has plenty of entertaining anecdotes about doing business in Russia, discovered SILOD during a visit to St Petersburg two and a half years ago. During a visit to the University, he met Piotrovskij, who offered to demonstrate the system to him. Van der Meer himself speaks no Russian but he says the interpreter who accompanied him was impressed by the output of the system. One thing led to another and a deal was struck up whereby A&X acquired part ownership and exclusive distribution rights to SILOD.

Now that Van der Meer has SILOD in his pocket and is currently paying the salaries of the Russian programmers at the University (a pittance in Western terms but still a considerable burden on his company), he is looking for ways to market it. Van der Meer has demonstrated the system to several potential customers in Holland, including an AT&T office and the European Space Agency in Noordwijk. The latter was interested in SILOD and A&X submitted a hundred-thousand guilder (\pm US\$50,000) proposal to develop a version of the system for the "space" domain but the project did not go through. Realizing there may not be many takers at that price, Van der Meer is considering offering the basic system for, say NLG20,000 (US\$10,000), and charging NLG15,000 (US\$7000) per "domain".

Like lots of Russian technology, SILOD looks a couple of decades behind the times. As Van der Meer readily acknowledges, the system's developers have scant appreciation for the art of user interface design, although the DOS-based program now sports a rudimentary menu system. SILOD's interactive mode is scarcely interactive: "interactive"

here means simply that you can type text freely in the source text window.

A trial English-to-Russian translation of a text Van der Meer had on hand revealed peculiarities in the Russian output that even a non-Slavicist could detect. Sentence segmentation looked shaky at best, while the abbreviation “no.” at the end of a sentence was blithely translated as *nyet*, leading us to believe the system did not perform much in the way of syntactic parsing. However, in one respect, SILOD is marvelously up-to-date: it is fitted with a dongle, one of Van der Meer’s most tangible contributions to the effort. In the beginning the Russians didn’t take it seriously, he says, but after six months they still have yet to hack it.

SILOD is A&X’s first foray into linguistic software; the six-strong company’s primary line of business is a mainframe-based maintenance application which A&X has lately decided to downsize and market more actively. Van der Meer freely admits he has absolutely no knowledge of the translation market. He feels a market study might be the right step forward but is apprehensive about the cost involved.

Van der Meer is quite proud of and even a bit protective of his Russian find; he believes the team in St Petersburg are great programmers and that he has chanced upon something unique. In Holland, he received positive feedback from translators to whom he has shown SILOD, and these are people whom you would think would be naturally sceptical of such a system. But Van der Meer remains out of touch with reality by relying on such anecdotal evidence, particularly given the relative “distance” of the English-Russian language pair (no Dutch translator would be caught dead praising an Dutch-English system). He has not had the linguistic capabilities and the lexical coverage of the system thoroughly analyzed by a professional computational linguist, something that would probably result in an unpleasant verdict.

As problematic, however, as his lack of linguistic knowledge is Van der Meer’s woeful lack of insight into the state of translation software today, and the economic forces which govern it. In many but not all quarters, file format compatibility is equally if not more important than linguistic capabilities; SILOD can only read ASCII files. No MT user today will accept not being able to update the dictionary – or at least create a user dictionary. Only SILOD’s developers can enter new terms; potential users would have to wait for monthly updates from St Petersburg. And avowed accuracy rates of “eighty- five” percent mean absolutely nothing without measurements of what kind of effort it takes to correct the remaining fifteen percent. Finally, more polished programs of approximately similar linguistic capabilities (albeit not for Russian) cost between a hundred and a thousand dollars in the US. To turn SILOD from what may be an adequate if crude system for scanning content into a polished system suitable for translating technical documentation of publishable quality would require an outlay that would probably make even major software development companies think twice (and many have) before embarking upon such an ambitious endeavor. Common wisdom dictates that the last ten percent of the software development cycle is the most difficult; it takes tremendous motivation, discipline, and effort to complete a program that people will want to use day in and day out. If you add to that equation the fact that the generalities of language are relatively easy to code and the exceptions are the hardest, then, charitably put, SILOD’s authors have probably done the first fifty percent – and that’s just the easy part.

A&X Automatisering, Postbus 330, Zwijndrecht, 3330 AH, The Netherlands; Tel: +31 78 19 12 55, Fax: +3 1 78 1 9 12 66 %>