

Sietec's Metal (Also) Does Russian

Klaus Schubert, LANGUAGE INDUSTRY MONITOR's correspondent in northern Germany, filed this dispatch.

Berlin, German – Like most things in life, market forces seem to dictate the choice of language pairs developed for machine translation systems. If that is the case, we have come a full circle, for today there appears to be a demand again for Russian MT, just like in the 1940s and 50s, when MT was first conceived. On the 25th of April, a Russian-German prototype of the Metal machine translation system was presented in Berlin by a three-member consortium. The event took place in an early capitalist palace in a part of the city appropriately called Siemensstadt. The prototype was funded by two German government agencies, the Amt für Auslandsfragen (AFA), which gathers foreign information, and the Bundesministerium für Forschung und Technik (BMFT).

The Russian prototype was developed by a group of eight computational linguists and computer scientists from the former East German Academy of Sciences. The group, led by Gerda Klimonow, took the plunge into the market economy as the Gesellschaft für Multilingual Systems (GMS), a company set up in Germany for this purpose by a Munich-based software house. Working under the auspices of Sietec, the team was able to exploit Sietec's development environment and the existing Metal modules. The AFA had selected Sietec to develop the system after a thorough, two-year evaluation of existing MT systems, with particular attention paid to their extensibility. In turn, Sietec discovered the Academy team at the Coling '90 conference in Helsinki demonstrating its PC-based German-Russian verb translation system.

The prototype has a good coverage of the grammatical complexity of the language pair. With some 7,000 entries, it boasted a set of lexicons considerably larger than is usual for a prototype. It was demonstrated with texts from two domains, nuclear power and aeronautics. Naturally, the prototype was fine-tuned to these texts, and the system produced ready-to-use texts of a quality unknown to ordinary Metal users, but in no way did Klimonow pretend the system was perfect. She listed no less than eleven major grammatical obstacles that were as yet only partially resolved. Some of these are specific to Russian-German transfer, such as the interpretation of the Russian aspect category, which does not exist in German, and the generation of definite and indefinite articles in German, which are likewise unknown in Russian.

The achievement shown in Berlin is due to a combination of expertise and thoroughness not often found in the MT world. Gerda Klimonow's team has a long-standing record of work both in German and Russian grammar as well as in computational linguistics. The team has maintained fruitful cooperation with well-respected sites in Russia and the Czech Republic, where – for want of hardware – they have for many years been forced to think, design, and plan before embarking upon implementation. In their strategy, they also stick to the often neglected distinction of linguistics, computational linguistics, and implementation, which in my experience is extremely expedient to all work in language technology.

The speakers in the Berlin event had much to say about future plans. Peer van Driesten, head of GMS, argued in favor of reversing the language pair, tackling other Slavic languages, and taking other steps toward remote cultures. Gregor Thurmair of Sietec mentioned language pairs like Russian-English and Russian-French along with Polish and Ukrainian and dwelt for quite some time on the possibility of using Metal modules for non-MT purposes, such as

information skimming and routing or index term generation in another languages. Stephan Boden (AFA) advised Sietec to pay more attention to the integration of the system with the customers' work-processes, emphasizing the importance of the workbench concept.

Despite the noble sentiments espoused about the future, the Russian-German Metal project is at the end of its funding, and it is uncertain whether the prototype will be developed into a marketable product. By acquiring a complete team from the Academy, Sietec has acquired expertise far above the value of the three years or so now spent in this project. It would be a shame not to pursue the project. Sietec may not reach the break-even point, however, unless it develops the German-Russian system too. Only the latter truly addresses the needs of German industry rather than the information-gathering priorities of authorities who ordered this prototype. Technological know-how and developmental impetus cannot be put into the deep-freeze. Sietec should not drop the ball – or its competitive edge will be lost.

COPYRIGHT © 1994 BY LANGUAGE INDUSTRY MONITOR