[*Language Industry Monitor* 21, May-June 1994]

The Finnish Formula

With Nokia Telecommunication's decision to implement the Kielikone MT system in its document production process, the world is a commercial high-end MT system richer.

Helsinki, Finland – One of Finland's largest companies recently licensed its Unix-based Finnish-English machine translation system, it has sold twelve thousand hand-held dictionaries in the past several years, it licenses its Finnish morphology and spellchecker to major OEM suppliers, and its electronic dictionary packages are enjoying growing popularity in Finland these days. With close ties to both the translation industry and the research world, Kielikone has achieved a unique position in this country of five million people. What is the Finnish secret?

After several years of evaluation and co-development, Nokia Telecommunications officially accepted the Kielikone machine translation system last fall and is currently implementing it within its documentation department. For the linguistic engineering community, this is important news, because it is (still) not every day that a new, high-end MT system is brought on-line in an industrial context. In Finland, news of the Kielikone system, which was formally announced in February, made a bit of a splash in the public eye as well: mainstream media coverage included appearances on TV (including one in which the "Key Deed of the Month" award was bestowed upon the company) and a feature article in the science and technology section of a major Finnish daily (written, fortunately, by Kielikone's managing director, Harri Arnola).

As befits a modern MT system, the Kielikone system is designed along the lines of a "translator's workstation." In interactive mode, this Unix-based program displays both source and target texts synchronized on-screen, and highlights the sentence being translated. In addition to basic editing functions, the system offers some special post-editing functions, such as an optional pop-up window listing translation equivalents for a given target language term, and function keys for quickly modifying English articles (a/an/the – Finnish, like Japanese, has no articles and this poses problems for MT systems), and capitalization (upper-, lower-, and proper-case). Unknown words are transferred as is and embedded in the text between double asterisks. The system's general Finnish-English dictionary tallies nearly fifty-thousand entries.

The Kielikone MT system is the result of an unusually close collaboration between Kielikone and Nokia. As Leo Kulikov, managing director of the MT group, explains, Kielikone built the general lexicon while Nokia compiled the domainspecific lexicon (for "information technology"). Strictly-speaking, the Kielikone system is domain-independent and is not "tuned" for the Nokia domain, but the system has been exhaustively tested on Nokia texts. Nokia took responsibility for the dictionary development tools because it had its own extensive termbases on mainframe systems which it wanted to port to the MT system. However, Kulikov adds that Kielikone has its own dictionary interface tool under development. Currently, users can only allow add nominal lexical entries (no verbs), but later this year Kielikone will start looking for ways of allowing users to teach and tune the system without actually touching the grammar rules. For development purposes, Kielikone has a carefully compiled five-thousand sentence bilingual corpora which it carefully examines for changes when new grammar rules are added. The group grades the translation of each sentence as good, ugly, or bad. Says Kulikov, “grammar rules can be dangerous.”

Nokia Telecom is Kielikone’s first genuine user, but the system has also been installed for test purposes at Trantex, a translation company and Rautaruukki Oy, a steel company. Both companies are in the process of evaluating the possibility of using the system for production, and are partly supporting the development of the system. Additional support for development comes from TEKES, a research agency administered by the Finnish Ministry of Trade and Industry. At the moment, the MT group is formally a part of Kielikone, although it is a separate cost-center within the company. However, there are plans to spin the MT group off as a separate company, possibly as early as this summer. According to Harri Arnola, Kielikone is also investigating the possibility of offering MT services in conjunction with a partner. He wonders whether such a service might not be able to address a “latent demand” for translations; texts which companies might consider translating if it could be done quickly and cheaply. “A rough but ‘true’ translation of a patent, for example, might be sufficient, since such a translation would be carefully checked by translators and lawyers anyway,” Arnola points out.

While delivering a high-end commercial MT system is an impressive achievement, Kielikone has been surprisingly successful in more modest arenas as well, namely with consumer and OEM products. M.O.T. is Kielikone’s line of bilingual electronic dictionaries available for the language pairs Finnish-English, Finnish-German, Finnish-Swedish, Finnish- French. The company supplies both “general” as well as “business” and “technical” dictionaries; a Finnish thesaurus is an additional option. M.O.T., which was developed entirely in-house, is available in DOS, Mac, Windows, and Unix versions.

“M.O.T. is doing surprisingly well, even exceeding our expectations,” says Arnola with obvious delight. “It’s currently our major source of revenue.” M.O.T. appeals to Finland’s export-based companies, and many of the orders are for large site licences. What is the appeal of M.O.T.? Arnola: “Well, we offer both general dictionaries and technical dictionaries, and companies like having both. The combination seems to be the key, particularly in the engineering community, where our sales are concentrated.”

Kielikone’s Finnish spellchecker is available both as an end-user product, called Morfo, or as an OEM package. While spellchecking for a language like English is a trivial task, only requiring a modicum of linguistic knowledge for making accurate suggestions, spellchecking for a highly declined, agglutinative language like Finnish is futile without morphological reduction. The Finnish spellchecker originally supplied with a well-known American word processing package was based solely on wordlists. As a result, its performance was unacceptable. “People were very indignant about it,” recalls Arnola. “It was considered scandalous.” Kielikone has now been contracted by the American company’s OEM supplier to replace the original Finnish spellchecker with that of Kielikone. It is worth noting that the usefulness of the familiar search and replace function standard in many wordprocessors is also something that would benefit from morphological reduction in the case of Finnish, but the mainstream (ie, American) software vendors have yet to go that far.

Several years ago, Kielikone also introduced a grammar and style checker for Finnish called VIRKKU, which was awarded software product-of-year in Finland in 1992. According to Arnola, the forthcoming version should be a significant improvement over the previous one. The system’s rule base system is being rewritten in more powerful descriptive language which should make more accurate syntactic checks possible. At the moment, it has not been decided whether the final version will allow users to enter their own rules. The new release will be available in DOS, Windows, and Mac versions. The price for the new version will be considerably less than that of the current package, FIM2,900 (±US$480).

In yet another market, that of hand-held electronic “travel” dictionaries, Kielikone has also enjoyed a measure of commercial success. According to Mika Herpiö, managing director of Käännöskone Oy, a separate company established to market these devices, some twelve thousand bilingual handhelds have been sold over the past two years at a price of FIM795 (±US$160). The Kielikone unit sports 54,000 entries, while its only competitor on the Finnish market, the Canon WordTank, offers just 30,000. Currently, versions are available for Finnish-English, Finnish-Swedish, and Finnish-German. Plans are afoot to merge Käännöskone and Kielikone, with Herpiö becoming managing director of the new organization.

According to Arnola, Kielikone has not embarked on ambitious marketing activities for its consumer products, but it has been fortunate enough to have the support of several large distributors who undertake telemarketing on behalf of the Kielikone line. They appear to have been quite successful.

While Kielikone is now clearly a commercial enterprise, it was originally established as a research project in 1982 to study natural language database interfaces for Finnish; it was financed by SITRA, a public venture capital fund. Early on, the project generated a tangible spin-off in the form of a high-quality morphological analyzer for Finnish, which made it possible to spellcheck the richly inflected Finnish language for the first time.

"When we started," says Arnola, "a natural language interface for Finnish seemed like a suitable goal. But we encountered a lot of very nasty problems. Plus, it appeared that natural language databases were and are still not commercially viable." The group therefore switched to machine translation in 1987 in response to "a large number of requests from potential customers." Kielikone, the research project (Kielikone means "machine language" in Finnish), thereby metamorphosed into Kielikone Oy, the company. Nokia was a pilot customer from the very start. While this undertaking attracted considerable attention, there were also doubts expressed in some quarters, partly because of the lack of a strong linguistic background. But the group defied the sceptics. By 1992, the MT project reached the product development phase, and in the fall of 1993 Nokia announced that it would start using the system for production. Arnola says that it is difficult to estimate the total investment in the MT system. Four to five people worked on it full-time since 1987, but the group had a morphology analyzer before it started and already had some experience with dependency parsing. However you slice it, Kielikone appears to have obtained a lot of mileage from its original research work: all of the Kielikone applications, from the handhelds to the MT system, share the same core lexicon, which the company compiled itself with the assistance of the English department of the University of Helsinki. More recently, Kielikone also compiled Finnish-Swedish and Finnish-German lexicons.

Since its inception as a research project more than ten years ago, Kielikone has been guided by Harri Amola (formerly Jäppinen), currently Kielikone's managing director. While Arnola appears to have quite successfully piloted the group technologically and organizationally into the commercial sector, his primary passion remains the development of the Kielikone parser. He currently divides his time between managerial activities at the Kielikone offices and further refinement of the parser at home, safely ensconced away from ringing telephones. "I must admit," says Arnola, "I'm very proud of the parser ," which he describes as a "deterministic dependency parser." While parsers for English and other Western European languages tend to be based on constituent models (ie, noun phrase, verb phrase, etc.), a dependency model uses the verb as a point of departure and builds up a structure around it, (ie, subject, object, indirect object, etc). Arnola says it is well suited for inflectional languages like Finnish (and, incidently, Japanese). He adds that once you have Finnish morphology licked, parsing the language comes easier, since so much grammatical information is encoded in the inflections.

From an implementation point of view, an important feature of Amola's deterministic parser is that it generates only one parse tree for a given sentence. "Because we don't generate all possible trees," says Arnola, "parsing times in relation to sentence length increase in a linear rather than a logarithmic fashion, as is the case with non-deterministic parsers." This means, quite simply, Arnola's parser is very efficient: it is fast and economical with memory. Arnola has parsed and checked more than five thousand Finnish sentences from various sources over the past two years. A recent evaluation with a small corpus of forty-four newspaper editorials (twenty-five to thirty-five sentences) achieved an accuracy rate of over ninety percent, lexical gaps not included, a high score for such “general” texts, which are carefully written yet complex.

While many of his colleagues will argue that semantic information is required to improve the accuracy of parsers, in particular to resolve ambiguity, Arnola remains highly sceptical of that approach. “Semantics is a swamp,” says Arnola. “It brings with it a raft of difficulties: data representation problems and great complexity problems.” The Kielikone system does not attempt any semantic processing, although Amola points out that there is a “flavor” of semantics in the dependency model. For MT, Arnola feels pure syntactic processing can produce translations of sufficient quality.

While linguistic software by definition is never truly finished, Arnola hopes to reach a plateau with his parser towards the end of the summer, whereupon he will “freeze” it (not difficult in this chilly land). It will then be, you could say, *Finnished*. And after that? A half year sabbatical – a welcome break after more than ten years of uninterrupted effort. However, Amola has no plans to turn his back on either computational linguistics or commerce for good. He has plenty of ideas for other commercial applications for the Kielikone technology, not to mention possible entrepreneurial excursions beyond the Finnish border. For a start, Kielikone recently signed a contract with a Swedish publisher to market electronic dictionaries in Sweden, and it is now actively looking for partners in other countries. A companion English-Finnish system is not in the pipeline, although Arnola hints that that could change if he is able to acquire a good English parser.

Owing to the idiosyncrasies of the Finnish language, the small market, and probably a few other factors, Kielikone has achieved a position within Finland which is difficult to find a parallel to in any other country. Does Amola have any insights to explain the company’s success? “You need a well thought-out theoretical basis for applied computational linguistics, such as spellchecking, morphology, or MT ,” he says. “But you also have to keep implementation in mind early on. Certain theoretical decisions can have dire consequences for efficiency.” Because he was originally trained as an engineer, Amola says he has an ingrained awareness of limited resources. “An engineer only has a limited amount of concrete with which to build a bridge,” he says. The physical world has natural constraints; in the digital world such constraints are less visible but they nonetheless remain. Previous MT projects, he points out, started out with strong theoretical foundations, but they did not keep efficiency in mind, and hence this became a major obstacle to building a working system.

Finland is set to join the European Union in January, 1995. Does Arnola foresee any useful opportunities in terms of participating in EU-funded research programs? Arnola waxes a bit hesitant. “I don’t know – all that paperwork,” he sighs. But surely there is paperwork attached to public funding in Finland? “Yes, but it is nowhere near as bad.” A final thought: maybe Kielikone’s secret weapon is that not having had to rely on Euro-funding, its researchers did not have to spend months writing proposals, they did not have to travel to meetings in various parts of Europe, they did not have to show their faces in Luxembourg once in a while, and they were not required to work with(geo-)politically correct partners. Instead, they could get some work done.

Prices: M.A.T., FIM1700 (±US$280) a single language pair; additional two,way lexicons are FIM850 (±US$140); technical and business lexicons, FIM1450 (±US$240)

Kielikone Oy, Vattuniemenkuja 4, PL 126, Helsinki, SF-00211, Finland; Tel: +358 0 6820 211, Fax: +35806820 167, Email: kkoy@kielikone.fi