

An Enthusiastic MT User

LogoVista E to J gets a workout at one small Japanese translation shop.

By David J. Littleboy

Tetranet is a tiny company, essentially a one man operation, whose main product is English to Japanese technical translation and the associated DTP. It was founded in 1989 by Tetsuro Hayakawa, who previously worked for IBM for nearly twenty years. Many of the documents Tetranet translates are manuals for shrink-wrapped software products by small companies for niche markets, but the range of documents is quite wide, extending from end-user manuals to technical reference works and even scientific papers. Tetranet started out using dedicated Japanese word processors but soon switched to the Macintosh to be able to offer a wider range of DTP services. There was a bit of a boom in Macintosh-based DTP companies at that time, and the Macintosh, despite being incredibly overpriced in Japan, became the only foreign PC to succeed in the Japanese market until quite recently. Unlike the majority of these DTP startups, Tetranet survived the 1990 to 1994 recession that followed the overheated “bubble economy” of the late 1980s, which speaks well for Hayakawa’s abilities as a translator and businessman. Tetranet now handles a volume of about 500 pages per month.

Tetranet’s original mode of operation was to farm out parts of a translation to multiple freelance translators, collate the results and edit the text for errors and stylistic uniformity. Although Hayakawa sees himself as an entrepreneur and not a translator, his role was closest to what would normally be called a “senior translator” in a larger organization. In particular, he takes full technical and editorial responsibility for all the translations he delivers.

Hayakawa was dissatisfied with this system from the very start. Technical translation in Japan always operates on extremely short deadlines, so a single job must almost always be handled by multiple translators. Furthermore, Hayakawa has multiple clients with differing stylistic requirements. Since he is translating into Japanese, all these clients feel they are experts and are quick to criticize even minor points. Thus Hayakawa found himself essentially rewriting everything that came back from the translators. This would not have been such an intolerable situation had the quality of the translators been higher, but Hayakawa was unable to find translators with technical expertise. For Hayakawa then, the idea of MT was extremely attractive, and he has been investigating it actively since he started Tetranet.

Until recently, MT was simply not feasible for a small operation. The PCs available around 1990 and 1991 were simply not powerful enough, and the software itself was problematic. However, by 1993, the situation had changed: desktop computers became powerful enough and commercial MT products, among them LogoVista E to J (developed by Language Engineering Corporation, Belmont MA, USA), became available at a reasonable price.

Hayakawa adopted LogoVista E to J in January 1993, and is adamant that he will never go back to using freelance translators. Although the output from LogoVista requires post-editing, rewriting LogoVista output is far easier than rewriting low quality human translations. First of all, much of the rewriting can be done mechanically. Since Japanese almost never uses the pronoun “you” in actual writing, a simple search and delete operation suffices to resolve this problem. Plural markers can be similarly dealt with. Verb groups can be all rewritten to an unmarked form. (LogoVista goes to all sorts of lengths to render overly complex English verbs in similar Japanese.) There are no missing sections (a major and unavoidable irritation when dealing with human translation) and, of course, LogoVista never makes a typing error. Finally, although LogoVista makes mistakes, its mistakes are easily recognizable, since it has the

consistency of any mechanical system.

All these points, however, would be true of any MT system. What makes LogoVista far superior to other systems Hayakawa has tried is that LogoVista has enough English syntactic expertise to handle input text without pre-editing. Most E to J and J to E systems prior to LogoVista required that the input be broken up into phrases by the user before translation. The prohibitive labor costs of this phase has largely prevented the widespread adoption of MT in Japan.

Logo Vista has several other attractive features as well. Most of Hayakawa's customers are large enough corporations that they have accumulated their own in-house dictionaries in an on-line format. LogoVista allows these dictionaries to be converted quite easily. This is extremely important for Hayakawa because each customer has a multitude of incredibly picky vocabulary requirements. For example, "computer" in Japanese can be written *conpuuta*, *conpuutaa* or *keisanki*, and each customer insists that their favorite be used. As Hayakawa explains, "Japanese has far more 'flexibility' at the lexical level than any other language." Stylistic issues can also be partly dealt with at the dictionary level. Chinese loan words, like their Latin and Greek counterparts in English, are used in Japanese to sound technical and native Japanese words to sound more humanistic. Thus by switching the dictionary definition of "to use" to be either *shiyousurn* (Chinese) or *tsukau* (Japanese) some of the stylistic requirements of the job can be finessed. This is particularly useful now that there is a trend towards technically useless but more "user-friendly" end-user manuals in Japan as well as the US. Although LogoVista also provides stylistic options for the formality level of the output, Hayakawa finds that setting this to the most informal level (the *dachou* level) produces the best results. Human translations become quite idiosyncratic when the formality level increases (to the point that the gender of the translator is immediately obvious), so there has been a trend towards the use of the most informal level in technical manuals, and, Hayakawa points out, the quality of LogoVista output goes down rapidly when it tries to put in formality.

Cost is another major point in MT's favor. E to J human translation costs about ¥15 per input English word, so a 400-page document might cost US\$12,000 for the rough draft. Since this is far more than LogoVista itself plus a high powered Mac to run it on, Logo Vista can pay for itself from the savings on its first job. However, there are complications. Getting a dictionary set up for a particular client and a particular type of job is time-consuming and cannot be billed to the client, so the reality is that MT is only cost-effective if multiple jobs can be run through a given setup.

It should be emphasized here that once LogoVista has produced a rough draft there is still a lot of work to do. Hayakawa handles this as a two stage process, which he refers to as secondary and tertiary translation. Secondary translation consists of the mechanical editing described above and a careful sentence by sentence check of the output against the original. This check is absolutely necessary. Hayakawa stresses that MT good enough that the editor would only have to check the original in places that are clearly questionable is the stuff of radical science fiction. Tertiary translation is an editorial phase in which the linguistic quality of the result is polished. This phase assumes that the result of the secondary translation is correct, and Hayakawa considers only the Japanese (not the original English) during this phase. Although this two-stage approach could be taken with human translation, the consistency of the MT output makes this whole operation much less painful. A final point is that what is going on here is not machine translation but machine-aided translation. While LogoVista is handling the dictionaries and producing rough drafts, Hayakawa reads every sentence in the original and reviews every sentence that is actually delivered to the client. Hayakawa considers it inconceivable that an MT system could ever be trusted to the extent that it would be safe to let even one sentence in the original go unread, and unlikely that an MT system will ever produce a sentence that does not require editing. On the other hand, in my discussion with Hayakawa, I came away with the strong impression that he has never even considered doing translation himself from scratch. Hayakawa admits that given even a clumsy, potentially incorrect translation, it is easier to read the English after having looked at the bad Japanese. This is because

Japanese and English are such radically different languages that reading in the second language is significantly slower than in the first, even for experienced translators. Thus machine-aided translation is already allowing someone who would not even consider translating to be one of the world's most productive translators.

Despite his enthusiasm for MT, Hayakawa still thinks there is room for improvement. Since MT output is extremely consistent, many of the editing operations that consume the majority of Hayakawa's time are repetitive and could be mechanized. Some of them can be handled by simple search and replace in a text editor, but the MT system itself would be a better place to handle many of these repetitive problems. The different stylistic requirements of different customers could also be handled by user-selectable options. A problem that Hayakawa would like to see addressed is that LogoVista cannot differentiate English from programming languages and tends to blithely translate the samples of programming code that appear in a document. One surprising desire is for increased speed. Although the current version of LogoVista can translate a 400-page manual in three days Hayakawa's PowerPC-based Macintosh, which is over four times faster than the 30 to 40 pages per day the post-editing requires, Hayakawa would still like it to be faster. There is a psychological effect here, since there is a feeling of substantial progress having been made at the point the rough draft is complete. Thus he would really like to set the machine going and come back the next morning with the job ready to edit.

From the standpoint of quality translation, MT is problematic, especially for the Japanese to English case. Yet the reality is that MT, when seen as machine-aided translation, is already an extremely useful tool in the overburdened E to J technical translation market. In Japan, the press, the MT developers and the Japanese government bureaucracies that backed AI and the MT projects have hyped MT to the point that many people think that MT systems can produce usable language. This is a problem because English differs so radically from Japanese that it is virtually impossible for Japanese end-users to evaluate J to E translations.

Users of MT systems must be aware, as is Hayakawa, that MT systems have no understanding of the subject of the translation and are merely playing word games. This means that when a user is concerned with quality and accuracy, he/she must only accept a translation if he/she is sure that it has been checked against the original by someone who understands both languages and the subject of the document. Since the idea of delivering MT output as such to a client would never even cross Hayakawa's mind, he was essentially speechless when I asked him if he had ever thought about this issue. Although for Hayakawa MT is an indispensable tool that has helped him triple his business over the last two years, MT only becomes interesting when it is compared to the very worst human translators.

Despite having to rewrite every sentence produced, LogoVista E to J still manages to be a great improvement over the hideously incompetent freelance translators upon which Hayakawa was dependent prior to acquiring LogoVista. As Hayakawa puts it, "whether LogoVista is good or bad is not the question. Rather, it is simply the case that translation of the required volume would be impossible without it." For Hayakawa, LogoVista E to J is nothing less than a salvation.

David J. Littleboy (70751.1562@compuserve.com) is an American translator living in Tokyo.