Realities of Development in Market-oriented MT

Winfield Scott Bennett
Logos Corporation

Introduction

After many years of uncertainty and false starts a genuine market is emerging for machine translation and related computer technology. This emerging market has brought demands that were unforeseen in the past. Those of us who are in development in market-oriented MT must strive to meet these demands.

This presentation seeks to explore briefly a number of the realities of such development. My focus is entirely on commercial systems which are marketed for corporations and their service providers.

Market Realities

The need for rapid localization of products is undoubtedly the major reason for the sudden and increasing interest in machine translation and related technology. In the past five years localization has evolved to the point where it is essential for all products sold outside the home country. Delays due to translation times were once acceptable when they amounted to months; now such delays are barely acceptable when measured in days. Corporations face significant losses in revenue and possible loss of market share, if they cannot deliver their products to all their local markets at virtually the same time. Time-to-market is the key concept; simultaneous release of fully localized products in all markets has become the goal.

In these circumstances any means to speed up the process of document production and translation are considered to be worthwhile. In the past documentation was generally written in one language and delivered in final form for translation; now the process often involves multiple drafts given to translators from very early in the product cycle. This means that translators must be able to handle revisions efficiently ami effectively. The reuse of translations from previous jobs is clearly desirable, if not mandatory. This is hardly new, simply more urgent In the modern era translation is a process within a larger process, not a single event

Translators also face documentation other man printed matter. On-line help files and text within source code also must be translated. There are even predictions mat electronic documentation will become the principal source of information within the next few years. Means must be found to enable translation staffs to deal with texts embedded in source code and with hyperlinked files.

It is no real surprise that much translation work is into multiple languages and may be by several translators or groups of translators at the same time. Process management and communications are central issues in localization.

Finally, the market clearly prefers "one stop shopping", i.e., getting all the needs met in a single commercial source. Machine translation software suppliers are called upon to provide solutions for the whole range of market demands in an integrated package. Development for market-oriented MT is involved in producing a suite of tools and functionality for translators. Simply focusing on the machine translation system itself is not a realistic approach to commercial development in the 1990's.

Development

The market realities touched on above provide the basis for the development of machine translation and related tools in the 1990's. While the focus for commercial MT system development in the 1970's and 80's was principally on the MT engines themselves, more and more development emphasis must now be focused on meeting the broader needs. This does not mean abandoning work on the engines; clearly the better the output the less time and work needed for a finished translation. However, the other parts of the process demand their share of the developers' time.

Not surprisingly developers still must devote time to machine translation systems themselves. While improving the quality of output and the performance of the MT engines remain issues, accessibility and user-friendliness of the systems also demand attention. The market clearly shows us in commercial MT that long-standing metrics (e.g., performance or quality) are only part of the equation. Users need to be able to *use* the system as easily and as productively as possible.

The Client/Server configuration is becoming the standard, because such architecture allows a clear separation of the day-to-day user tasks (e.g., submitting and retrieving documents, dictionary work and postediting) from system administration. Freed of the need to deal with non-translation system issues, the translator can devote her or his time to the job itself.

Client/Server architecture is a start; another issue involves interfaces. The era has passed when users would endure unfriendly computer systems; windows environments have set the standard. Translators using MT systems want "windows look" environments; developers must respond or face a market unwilling to use their products. LogosClient, for example, provides just such an environment by allowing the user to work within the word-processing or DTP package of her or his choosing. Access to the machine translation system becomes merely a part of a familiar environment, another option.

Since no one produces unformatted ASCII text for documentation, any machine translation system must be able to handle formatted text routinely and effortlessly. Filter technology thus becomes crucial for market-oriented MT; development and maintenance of filters to handle all major formats are necessary in commercial machine translation. Along with all the major formats help files and source code must be accessible to the translation technology. Such filter work requires significant resources.

The centrality of MT dictionaries coupled with the need to reduce the time needed for database updating has made dictionary management an important market issue. Ease of use and speed of operation are guiding factors. Facilitating all aspects of dictionary creation and

maintenance has led to the constant development of dictionary management systems with user-friendly interfaces and to semiautomatic means of updating dictionaries.

Along with dictionary management there is a significant market demand lor means to manage terminology databases and glossaries. The intention is to permit users' access to dictionary information for their use along with the system's access. The capability to submit simple glossary files for translation is also desirable. Although such functionality may be construed as outside the realm of MT development, it clearly meets a market need.

In recent years translation memory (TM) has emerged as a significant and effective aid in the translation process. In various forms (e.g., Eurolang's Optimizer, GlobalWare's XL8, IBM's Translation Manager) mis technology allows the user to reuse her or his translations once they have been stored in a database. The idea is simple enough: store matched pairs of source and target language sentences and reuse them when and where they occur in any text. Most translation memory systems permit use of partial ("fuzzy") matches in the process based on thresholds established by the user. The obvious benefit is that reusing translations prevents translators from redoing work even in revised or updated documentation; other benefits (e.g., sharing translations and translation consistency) are also significant

While translation memory has been present in the market long enough to be a factor, the combination of TM and MT is quite new. Fully integrated TM-MT systems provide the benefits of bom: the translation memory package handles all sentences which are in its database and the machine translation system handles those which are unique; both systems are brought into play for the fuzzy matches, normally providing the user with two possible translations. Integration of complex systems, as we at Logos have done with both XL8 and Optimizer, involves considerable development effort, but clearly meets a very real market need.

The lessons of the current market make it quite clear that translation is not an event, but a process which is, in turn, part of a larger process. This fact challenges development groups to find solutions for the effective management of the entire process, including documentation revision, multiple translators and language pairs and various service providers (translation bureaus and freelancers). Since deadlines are vital, translation managers must also have an accurate knowledge of the completion status of the entire process. While such matters are rather far removed from the translation engine, it is a genuine need of the market and demands the attention of developers of commercial MT and related technology.

Since postediting is part of the reality of computer translation, the users need intelligent postediting environments. The functionality provided in the average word-processing and DTP packages is simply not powerful enough. While WYSIWYG seems desirable, other factors are equally at issue. Independent neutral postediting environments may be preferable to an established word-processing or DTP package.

Related to postediting is the need to capture either sentence pairs (for storage in a translation memory database) or terminology (for dictionary and terminology database updates). The relationship to postediting is clear: while postediting a text, the translator is most aware of what she or he wants to store for future reference. Such tools are offered in several TM packages (e.g., Optimizer and XL8).

Beyond the other demands there is a decided need for an assortment of tools to aid in translation. Such tools range from sophisticated utilities, such as pattern matchers and grammar and spell checkers, to simple macros within larger environments, e.g., a macro which allows multiple search and replace operations or dual synchronized editing windows. The key in such development is that the translator is the only one who can provide the insight into what is needed; the developer's task is to fulfill the need.

The modern translation market allows a variety of modes of operation. Members of a translation team may be in the same location or widely separated. They may be linked by LAN or WAN or not linked at all, operating remotely. To provide solutions for the various possibilities, it is necessary for commercial MT to be available in the various modes. Client/Server configurations provide a suitable basis for this. Providing remote access to the MT engine becomes essential. Just as LogosClient allows the user to interact with the MT server via LAN or WAN, LogosClient also provides the same capability via dial-up. Again development efforts must be expended to meet a significant market demand.

Conclusion

Two points are clear as machine translation technology finds itself in its fourth decade. First, there is finally a genuine and growing commercial market. Second, that market is making significant demands on those of us who are engaged in market-oriented development. Many of these demands are somewhat removed from the areas of MT development work in the past; all are needed to meet the realities of a present and emerging market.

This is an exciting time for market-oriented MT and related technology. Those involved in development for the translation market are learning to respond quickly to a growing demand for solutions in the commercial translation process; those involved directly in the process are learning that commercial MT with related technology can provide means to facilitate the process.

Ultimate market reality is that we developers must rely on the users to tell us what they need and that the users must rely on us developers to meet these needs. The joint efforts of users and developers are already leading to the evolution of machine translation and related technology in ways unforeseen even five years ago.