Commercial systems: state of the art in 1999

John Hutchins
(University of East Anglia, Norwich NR4 7TJ, England)
(E-mail: WJHutchins@compuserve.com)

INTRODUCTION

In a general overview of the availability and potential usefulness of commercial machine translation (MT) systems and translation tools[1], it is important to distinguish three basic types of translation demand: dissemination, assimilation, and interchange.

The first, and traditional one, is the demand for translations of a quality normally expected from human translators, i.e. translations of publishable quality – whether actually printed and sold, or whether distributed internally within a company or organisation. The use of MT for **dissemination** purposes has been satisfied, to some extent, by machine translation systems ever since they were first developed in the 1960s. However, MT systems produce output which must invariably be revised or 'post-edited' by human translators if it is to reach the quality required. Sometimes such revision may be substantial, so that in effect the MT system is producing a 'draft' translation. As an alternative, the input text may be regularised (or 'controlled' in vocabulary and sentence structure) so that the MT system produces few errors which have to be corrected. Some MT systems have, however, been developed to deal with a very narrow range of text content and language style, and these may require little or no preparation or revision of texts.

In recent years, the use of MT systems for these purposes has been joined by developments in translation tools (e.g. terminology databases and translation memories), often integrated with authoring and publishing processes. These 'translation workstations' are more attractive to human translators. Whereas, with MT systems they see themselves as subordinate to the machine, in so far as they edit, correct or re-translate the output from a computer, with translation workstations (or workbenches) the translators are in control of computer-based facilities producing output which they can accept or reject as they wish.

The second basic demand is for translations at a somewhat lower level of quality (and particularly in style), which are intended for users who want to find out the essential content of a particular document or database resource – and generally, as quickly as possible. The use of MT for **assimilation** has been met in the past as, in effect, a by-product of systems designed originally for the dissemination application, since some users found that they could extract what they needed to know (e.g. for screening and/or information gathering) from the unedited MT output. They would rather have some translation, however poor, than no translation at all. With the coming of cheaper PC-based systems on the market, this type of use has undoubtedly grown substantially.

---

[1] This version of the paper for the collection "Computers and translation" edited by Harold Somers (Amsterdam: John Benjamins, 2003) was completed in September 1999. It includes references to examples of then currently available systems, which were excluded from the later published version (revised in summer 2002) because it was recognised that the rapid changes in the commercial scene would render all these references out of date as soon as the article was published. This version is being made available as a historical record of the 'state of the art' in mid 1999.

Related to this application is for translation within multilingual systems of information retrieval, information extraction, database access, etc. Here MT systems operate as components of **information access** systems, i.e. translation software is integrated into (a) systems for the search and retrieval of full texts of documents from databases (generally electronic versions of journal articles in science, medicine and technology), or for the retrieval of bibliographic information; (b) systems for extracting information (e.g. product details) from texts, in particular from newspaper reports; (c) systems for summarising texts; and (d) systems for interrogating non-textual databases. As yet, however, there are few commercial systems available in this area.

The third type of demand is that for translation between participants in one-to-one communication (telephone or written correspondence). In this **interchange** use, the situation is changing quickly. The demand for translations of electronic texts on the Internet, such as electronic mail and 'chat' lists, is developing rapidly. In this context, human translation is out of the question. The need is for immediate translation in order to convey the basic content of messages, however poor the input. MT systems are finding a 'natural' role, since they can operate virtually or in fact in real-time and on-line and there is little objection to the inevitable poor quality.

Another context for MT in personal interchange is the focus of much research. This is the development of systems for **spoken language** translation, e.g. in telephone conversations and in business negotiations. The problems of integrating speech recognition and automatic translation are obviously formidable, but progress is nevertheless being made. In the future – still distant, perhaps – we may expect on-line MT systems for the translation of speech in highly restricted domains.

The focus of this presentation will be the development and use of commercially available systems for dissemination, i.e. for aiding the production of 'publishable' quality translations. Other applications will be treated more briefly. (It does not need stressing that such are the rapid changes in this field that any survey of commercially available MT systems is almost out of date within three or four years, and no more than of historical interest after ten years.)

Types of systems

Today we may distinguish the following types of systems and their most appropriate areas of application:

a) mainframe, workstation and/or client-server systems on intranets of large organisations

b) MT systems for professional translators

c) translator workstations for professional translators operating on company intranets or independently

d) computerised translation tools: dictionaries, terminology management software, translation memory systems

e) MT systems for occasional users and/or casual home use

f) systems designed for Internet use and/or for translating Web pages, either for company or individual use

g) MT services on the Internet providing translations on demand

Traditionally, MT systems are divided according to techniques: direct translation, transfer-type, interlingua-based, statistics-based, example-based, etc., but these distinctions are largely irrelevant to and hidden from users, and they are ignored in this paper. In general, however, it may be pointed out that commercial systems are based usually on well-tested approaches – for obvious reasons – and these tend to be based on the older traditional

(linguistics rule-based) strategies developed from the 1960s to the late 1980s. More recent developments in MT research in the 1990s based on text corpora – the statistics-based and example-based approaches [see chapter ?] – have not yet had much impact on the commercial MT scene, although some systems do incorporate example-based methods to a certain extent, and most of the translation workstations provide statistics-based facilities for the use and creation of 'translation memories', i.e. bilingual corpora of previous translations and their originals.

As throughout the computing industry, there has been a *de facto* standardisation of hardware, operating systems and inter-compatibility. For the larger systems, the standards are Sun systems, Solaris and Unix. For the smaller systems, the standards are PC compatibles, Pentium CPUs, Microsoft Windows 95/98 or NT. Some older versions are still available for Microsoft DOS systems, and increasingly rarely there are versions for Macintosh equipment. As for Internet access, nearly all systems either include or run with Netscape Navigator or Microsoft Explorer. MT products for Japanese, Chinese, and Korean generally require additional software (e.g. Japanese Windows, Japanese Language Kit), and occasionally run only on proprietary hardware.

Fuller details about all the systems mentioned here (and many others not mentioned) may be found in the *Compendium of translation software* produced by the European Association for Machine Translation.[2] This is a listing of current available commercial MT systems and computer-aided translation support tools (including translation workstations, terminology management systems, electronic dictionaries, etc.)

Mainframe, Client-server and workstation systems

The oldest MT systems are those developed originally for mainframe computers, e.g. the Systran, Logos and Fujitsu (Atlas) systems. The Systran system, originally designed for translation only from Russian into English, is now available for a very large number of language pairs: English from and into most West European languages (French, German, Italian, Spanish, Portuguese), Japanese, Korean, etc. Likewise, Logos, originally marketed for German to English, is also now available for other languages: English into French, German, Italian and Spanish, and German into French and Italian. The Fujitsu ATLAS system, on the other hand, is still confined to translation between English and Japanese (in both directions).

In their mainframe forms – much improved from their earlier 1960s and 1970s designs – these systems are still available, primarily now it appears for large companies or organisations wanting to include a MT engine in already existing documentation systems.

For most purposes, however, the large-scale systems take the form of workstations or client-server systems operating over company intranets. As the workstation and/or server the popular choice has been the Sun SPARCstation, and many of the older larger systems are still available for this platform. However, the Japanese computer companies chose to develop MT software for their own equipment, and it is still the case that some of the larger Japanese systems are available commercially only in these versions. For example, NEC's PIVOT systems run only on NEC EWS hardware. In most cases, however, they are available both for their own equipment and for the ubiquitous PC Windows configuration, e.g. the HICATS systems from Hitachi and the ASTRANSAC systems from Toshiba.

Needless to say, the prices of these systems make them affordable only for large companies or organisations with large translation services. From the mid 1990s onwards, most of these systems have begun to appear in cheaper versions for personal computers (see

---

[2] European Association for Machine Translation, c/o ISSCO/TIM, 54 route des Acacias, CH-1227 Carouge (Geneva), Switzerland.

below) – although often with substantially smaller dictionary resources and without facilities for working in groups and networks.

The main customers and users of mainframe and client-server systems are the multinational companies exporting products and goods in the global market. The need is primarily for translation of promotional and technical documentation. Technical documents are often required in very large volumes: a set of operational manuals for a single piece of equipment may amount to several thousands of pages. There can be frequent revisions with the appearance of new models. In addition, there must be consistency in translation: the same component must be referred to and translated the same way each time. This scale of technical translation is well beyond human capacity. Nevertheless, in order to be most cost-efficient, a MT system should be well integrated within the overall technical documentation processes of the company: from initial writing to final publishing and distribution. Translation systems are now being seamlessly integrated with other computer-based systems for the support of technical writers, not just assistance with terminology, but also on-line style manuals and grammar aids.

There are numerous examples of the successful and long-term use of MT systems by multinationals for technical documentation. One of the best known is the application of the Logos systems at the Lexi-Tech company in New Brunswick, Canada; initially for the translation into French of manuals for the maintenance of naval frigates, the company has built up a service undertaking many other large translation projects. Systran has many large clients: Ford, General Motors, Aérospatiale, Berlitz, Xerox, etc. Users of Logos include Ericsson, Osram, Océ Technologies, SAP and Corel. The METAL German-English system has been successfully used at a number of European companies: Boehringer Ingelheim, SAP, Philips, and the Union Bank of Switzerland.

A pre-requisite for successful MT installation in large companies is that the user expects a large volume of translation within a definable domain (subjects, products, etc.) The financial commitment to a terminology database and to dictionary maintenance must be justifiable. Whether produced automatically or not, it is desirable for company documentation to be consistent in the use of terminology. Many companies in fact insist upon their own use of terms, and will not accept the usage of others. To maintain such consistency is almost impossible outside an automated system. However, it does mean that before an MT system can be installed, the user must have already available a well-founded terminological database, with authorised translation equivalents in the languages involved, or – at least – must make a commitment to develop the required term bank.

Most large-scale MT systems have to be customised, to a greater or lesser extent, for the kind of language found in the types of documents produced in a specific company – in effect there is no 'standard' Systran or Logos system. This can be the addition of specific grammatical rules to deal with frequent sentence and clause constructions, as well as the inclusion of specific rules for dealing with lexical items, and not just those terms unique to the company. A further step is the introduction of a company-specific 'controlled language', not just for standardisation but for reducing well-known problems of MT such as lexical and structural ambiguities in source texts. The amount of work involved in such customisation and in the pre-editing control of input may not be justifiable unless output is in a number of different target languages.

In appropriate circumstances, large savings are reported by many companies that have installed MT systems, in some cases there have been gains in the costs of producing finished translations of up to 40 or 50%, and nearly always companies report faster throughputs. However, it must be stressed that it is only the larger organisations dealing with 100,000 pages a year or more that can expect such dramatic savings. Smaller companies and translation services may gain only in terms of speed of production and experience no cost

savings. The situation is, however, changing rapidly, and cheaper more powerful MT systems, combined with cheaper and more powerful publishing and authoring systems, will bring comparable savings to a wider range of companies and services.

Multinational companies are inevitably distributed at many locations in different countries, linked by internal networks (intranets). In this environment, translation jobs can be passed easily in electronic form from one office or branch of the organisation to another. In particular, a document may be authored in one location, sent for translation at another, and printed and distributed at a third. Most of the major MT system vendors market client/server systems, e.g. ASTRANSAC C/S (for English to and from Japanese, from the Toshiba company), Atlas (for English to and from Japanese, from the Fujitsu company), the Logos Multilingual Documentation Translation Software (for English to French, German, Italian and Spanish, and from German to English, French and Italian), Systran Enterprise (for English to and from French, German, Italian, Japanese, Korean, Portuguese, and Spanish, and from Chinese and Russian to English), and T1 Corporate (for English to and from German, from the GMS company (now part of Lernout & Hauspie), based on the previous METAL system). Other more recent systems do not have a 'mainframe' ancestry, e.g. LMT (for English to and from German, from IBM), Intranet Translator (for English to and from French, German, Italian, Spanish and Portuguese, from Globalink), Enterprise Translation Server (the system based on Transcend, for English to and from French, German and Spanish, and also from English to Italian and Portuguese), Tsunami Server (for English to Japanese, from Neocor Technologies), TranSphere (for English to Arabic, from AppTek), and TranSmart (from Kielikone Ltd.) for Finnish to English, developed originally for Nokia Telecommunications.) Mention should also be made of companies that develop translation software for specific customers, nearly always large government organisations or multinational corporations and invariably for client-server systems. The longest extant is Smart Communications Inc. of New York, which has built MT systems for Ford, Citicorp, Canadian Department of Employment and Immigration, etc. Other companies include Cap Volmac Lingware in the Netherlands, ESTeam Inc. of Greece, and LANT n.v. of Belgium

Translation workstations

In the 1990s, the possibilities for large-scale translation broadened with the appearance on the market of translation workstations (or translator's workbenches). Translation workstations combine multilingual word processing, means of receiving and sending electronic documents, facilities for document scanning by optical character recognition (OCR), terminology management software, facilities for concordancing, and in particular translation memories (TM). The latter facility enables translators to store original texts and their translated versions side by side, i.e. so that corresponding sentences of the source and target are aligned. The translator can thus search for a phrase or even full sentence in one language in the translation memory and have displayed corresponding phrases in the other language. These may be either exact matches or approximations ranked according to closeness.

It is often the case in large companies that technical documents, manuals, etc. undergo numerous revisions. Large parts may remain unchanged from one version to the next. With the translation memory (TM), the translator can locate and re-use already translated sections. Even if there is not an exact match, the versions displayed may be usable with minor changes. There will also be access to terminology databases, in particular to company-specific terminology, for words or phrases not found in the translation memory. In addition, many translation workstations are now offering full automatic translations using MT systems such as Systran, Logos, and Transcend. The translator can choose to use them

either for the whole text or for selected sentences, and can accept or reject the results as appropriate.

The translation workstation has revolutionised the use of computers by translators. Translators have now a tool where they are in full control. They can use any of the facilities, or none of them, as they choose. As always, the value of each resource depends on the quality of the data. As in MT systems, the dictionaries and terminology databases demand effort, time and staff resources. Translation memories rely on the availability of suitable large corpora of authoritative translations – there is no point in using translations which are unacceptable (for whatever reason) by the company or the client.

The original ideas for integrating various computer-based facilities for translators at one place go back to the early 1980s, in particular with the systems marketed by ALPS, but they were apparently not a commercial success and were withdrawn.[3] There are now four main vendors of workstations: Trados (with their Translators Workbench), probably the most successful company, STAR AG in Germany (Transit), IBM (the TranslationManager), and LANT in Belgium (the Eurolang Optimizer, previously sold by SITE in France).

Localisation support tools

One of the fastest growing areas for the use of computers in translation is in the industry of software localisation. Here the demand is for supporting documentation to be available in many languages at the time of the launch of new software. Translation has to be done quickly, but there is much repetition of information from one version to another. MT and, more recently, TM in translation workstations are the obvious solution. Among the first in this field was the large software company SAP AG in Germany. They use two MT systems: METAL for German to English translation, and Logos for English to French, and plan to introduce further systems for other language pairs. Most localisation, however, is based on the TM and workstation approach – mainly the Trados Workbench and the IBM TM/2. Typical examples of major users are Corel, Lotus, and Canon.

Localisation companies have been at the forefront of efforts in Europe to define standardised lexical resource and text handling formats, and to develop common network infrastructures, e.g. the OTELO project coordinated by Lotus in Ireland. The need for a general translation and management support environment for a wide variety of translation memory, machine translation and other productivity tools is seen as fundamental, and a number of companies are producing 'localisation support tools', e.g. Corel CATALYST for managing and routing localisation among translators, software engineers, project managers, for efficient use of different tools during overall processes, for automated updating, unified file tracking, etc.

The translation and management requirements of software localisation have been sufficiently distinct for the creation of a dedicated organisation (Localisation Industry Standards Association, based in Geneva), which holds regular seminars and conferences throughout the world.

Systems for independent professional translators

For the independent translator, the translator workstation may be no more affordable than the larger MT systems. Professional translators not employed by large organisations have currently two options: (a) relatively powerful systems capable of running on widely available computer equipment, e.g. Windows-based PC systems, (b) translation aids such as terminology management systems and TM programs.

---

[3] The company subsequently became a translation service under the name ALPNET, using its proprietary computer-based translation aids.

Most vendors of client-server systems also have systems on the market designed primarily for the demands of the professional user, i.e. systems that have facilities for post-editing and publishing, and that can be used with terminology databases and sophisticated word processing facilities. In origin, these systems are either downsized versions of mainframe (or client-server) systems or enhanced versions of cheaper PC systems. In the case of the former, often the same range of languages is covered as for the larger intranet versions. This is the case, e.g., for Systran Professional (the PC version of the client-server Systran Enterprise), for the Atlas for Windows (a version of the Unix-based Atlas system from Fujitsu), for T1 Professional, and for the two systems from the Pan American Health Organization (SPANAM Spanish to English, and ENGSPAN English to Spanish). What these systems lack in comparison with the intranet client-server systems are generally the wide range of document formatting and conversion facilities and sometimes the complete range of text processing compatibility. However, even this situation is changing as standalone computers become more powerful, and as users' demands become clearer, so that increasingly these 'professional' systems for the independent translator are acquiring the range of facilities found previously only in the largest mainframe and client-server systems.

Other vendors of MT systems have 'professional' versions, which are in some cases enhanced versions of their cheaper systems for less-demanding (home) users [see below]. As just a few examples: Personal Translator PT Plus (for English to and from German), Power Translator Pro (for English to and from French, German, Russian, and Spanish, from Globalink), ProMT Professional (for Russian to and from English, French, German, Italian, and for other language pairs), and Winger Pro (for English to and from Danish and Spanish, for Danish into German and Russian, etc.). There are now a number of fairly powerful Japanese/English systems suitable for 'professional' use: LogoVista E to J Pro (for English to Japanese), Pivot (from NEC), Tsunami (for English to Japanese, from Neocor Technologies, now owned by Lernout & Hauspie), Typhoon (for Japanese to English, also from Neocor) Yakushi nyorai (from the CSK Corporation, for Japanese to and from English, and for Japanese into French and German)

Translation support tools

Just as large companies may well prefer translation workstations to fully-fledged MT systems, the individual professional translator may not want to purchase a MT system that may cover only some of the languages required. Since the mid 1980s there has been a wide range of translation aids, some designed originally for workstations in larger organisations, intended primarily for individual translators for use on PC-type equipment.

Electronic dictionaries (usually in CD-ROM form) are available from nearly all dictionary publishers, and from many companies supplying computer software. They can still be useful as lexical resources.

Terminology management software provides facilities for professional translators to create, update and revise their own lexical resources, whatever the languages concerned. Typical facilities include means for downloading from on-line or other electronic databases. Among the best known are the MTX system from LinguaTech International (previously called Mercury or Termex), Multiterm from Trados, TermStar from the STAR company, and Translexis from IBM Germany.

Software for translation memory in individual packages (as opposed to components of translator workstations) is being marketed by a number of vendors. These programs allow individual professional translators to build their own stores of searchable and aligned bilingual databases of original texts and their translations. Best known are the products Déjà vu (from Atril Software), MemCat (from Keck Software), WinAlign (from Trados), and specifically for the localisation industry, Loc@le from Accent Software, and SDLX from

SDL International. All these programs can cope with texts in any language written in Roman characters, and some with non-Roman scripts. One system specifically oriented alignment of Arabic and English is the Cat Aligner from Sakhr (Egypt).

Systems for non-professional (home) users

The basic need of the non-professional user of translation software is primarily as a means of access to foreign language texts, to find out what a text in an unfamiliar or unknown language is about. What matters is the message, in the most general terms. There is no necessity to have a perfect translation. Any of the systems already mentioned can serve this need; indeed in earlier years one of the main uses of mainframe MT systems was the provision of rough translations, i.e. the unedited crude output, for the purposes of intelligence analysis or for scientific and technological reviews. At the European Commission, one of the principal uses of the Systran system is still the production of crude (sometimes lightly edited) translations for rapid surveys of documentation.

Software for personal computers began to appear in the early 1980s from ALPS and from Weidner. Their output was at a level of quality suitable only for information assimilation use, but they were too expensive for the casual home user. In fact they were bought mainly by professional translators, who found them frustratingly unsuitable for their needs. This experience may have convinced professional translators that PC translation software would always be useless for their purposes, but the more recent 'professional' systems described above are changing this perception.

For the 'non-professional' user it was not until PC equipment and software were much reduced in price during the early 1990s that this large potential market was opened up. Earliest in the field were the Japanese computer companies, selling systems, usually for Japanese-English or English-Japanese, and designed to run on their own microcomputers. In the United States one of the earliest vendors was Linguistic Products with its series of PC-Translator systems. Over the years, many language pairs have been produced and marketed, with some success as far as sales are concerned. Another early vendor was Globalink, which after merging with MicroTac, a company which had been very successful in selling its cheap Language Assistant series of PC software, produced in the early 1990s its now well-known 'Power Translator' series for translation of English to and from French, German, Italian and Spanish. This has now recently been partly superseded by the 'Telegraph' series for the same language pairs.

Nearly all the vendors already mentioned have versions of their systems intended for home or non-professional use, but not always with the same range of language pairs (e.g. Systran Personal lacks Chinese and Russian). Examples of such 'home use' systems are T1 Standard, Personal Translator PT, ProMT Home, Winger, and numerous Japanese systems.

Finally, it may be noted that there is a proliferation of particularly cheap products, marketed as 'translation systems' but which is fact are little more than electronic dictionaries. They sell presumably because of the widespread belief among those unfamiliar with translation, that all that is needed in order to translate something is a bilingual dictionary.

Sales of PC translation software have shown a dramatic rise during the 1990s. There are now estimated to be some 1000 different MT packages on sale (when each language pair is counted separately.) The products of one vendor (Globalink) are said to be present in at least 6000 stores in North America alone; and in Japan one system (Korya Eiwa from Catena, for English-Japanese translation) is said to have sold over 100,000 copies in its first year on the market.

Though it is difficult to establish how much of the translation software sold in large numbers is used regularly after initial purchase (some cynics claim that only a very small

proportion is tried out more than once), there is no doubting the growing demand for 'occasional' translation, i.e. by people from all backgrounds wanting gists of foreign text in their own language, or wanting to communicate in writing with others in other languages, however poor the quality. It is this latent market for low-quality translation, untapped until very recently, which is now being exploited. As a consequence, many products have to be treated with caution – they may not even meet minimal standards for crude 'information-only' translation. There is the clear danger that these products could give MT as a whole a name for poor quality in general.

MT for the Internet
　　　　The largest area of growth for translation demand is now undoubtedly based on use of the Internet. There is the need of the 'occasional' user for software to translate Web pages, electronic mail and other Internet resources and texts, either off-line or on-line, and the availability of on-demand Internet-based translation services for companies.

　　　　There has been a rapid increase in MT software products designed specifically for translating Web pages. Japanese companies have led the way, primarily with systems for translating from English into Japanese: nearly all the companies mentioned above have a product on this lucrative market. Some examples are: TransLinGO! (from Fujitsu), ASTRANSAC for Internet, Dr.Surf (from Media Vision and Kyushu Matsushita Electric), Translation Adaptor, Instant Translation (from Hitachi), King of Internet (from IBM Japan), Netsurfer/ej (from Nova). The Japanese companies were followed quickly elsewhere, e.g. Globalink with its Web Translator, Transparent Language (vendor of Transcend) with its Easy Translator, ProjectMT with its ProMT Internet.

　　　　As well as these systems for off-line translation, we are now seeing Internet gateway services adding translation facilities: the most recent example is the availability on AltaVista of versions of Systran for translating French, German and Spanish into and from English.

　　　　Equally significant has been the use of MT for electronic mail and for 'chat rooms'. The use is not simple curiosity, although that is how it often begins. Usage has rocketed, and soon CompuServe will offer MT as a standard for all its e-mail. As for Internet chat, Globalink has joined with Uni-Verse to provide a multilingual service. (Of course, none of these systems have been designed specifically for translating the kind of colloquial, jargon-filled, and often 'ungrammatical' language found in electronic mail and on-line discussion forums – and the danger of poor quality for the reputation of MT is real.)

　　　　At the same time, many MT vendors are providing network-based translation services for on-demand professional-level translation, generally with human revision as an option. In some cases these are client-server arrangements for regular users; in other cases, the service is provided on a trial basis, enabling companies to discover whether MT is worthwhile for their particular circumstances and in what form. Such services are provided, for example, by Systran, Logos, Globalink, CompuServe, Fujitsu, JICST and NEC. In most cases, clients have the option of receiving unedited translations or versions post-edited by the suppliers' translators.

　　　　Some companies have now been set up primarily for this purpose: LANT in Belgium is an example, based on its rights to develop the METAL system and on the Eurolang Optimizer, which it also markets. Its speciality is the customisation of controlled languages for use with its MT and TM systems. In late 1997 it launched its multilingual service for the translation of electronic mail, Web pages and attached files. And in Singapore, there is MTSU (Machine Translation Service Unit of the Institute of Systems Science, National University of Singapore), using its own locally-developed systems for translation from English into Chinese, Malay, Japanese and Korean (with Chinese its main strength) and with editing by professional translators. The service is providing large-scale translation over the

Internet for many customers world wide (mainly multinational organisations), and including much of the localisation needs for software companies in the Chinese-language markets. Finally, mention should be made of Alis Technologies, which provides a service based on the integration of disparate language products including the use of commercially available MT systems, documentation software, providing pre- and post-editing facilities and covering all types of text including internal documentation, email, and Web pages.

MT for information access

The growing use of the Internet is highlighting the need for systems that combine translation with other language-oriented facilities, in particular database searching, information retrieval and summarisation. As yet, however, there are few such systems available. The Systran company has recently released its InfoRaptor cross-lingual information retrieval software enabling users to search for information in another language; and it also allows users to search for sentences with certain linguistic and/or semantic characteristics. Most of the Webpage translators could be used for this purpose, although few enable search terms to be formulated and translated before searching the Web.

It is to be expected that in future this will be one of the main growth areas. Several research projects supported by the European Union combine MT with information access, information extraction, and summarisation software. Many are intended eventually as products for the information marketplace.

Systems by Language Pairs

Western European languages

From the very beginning of the commercialisation of MT systems, the major European languages have been well covered. Translation from English into French, German, Italian, Spanish, and Portuguese, and from these languages into English, is available from all the main vendors. This software can be found on sale by many retailers of computer software.

Thanks to its many years' pedigree, Systran offers probably the most powerful systems. Based on the mainframe systems (which are still available) there are now versions designed for different categories of users. At the 'top' end are the Systran Enterprise client-server systems designed for professional use in independent or company translation services, and the Systran Pro systems for PC and Windows. For the less demanding cheaper end of the market there are the Systran Classic and Systran Personal systems, both designed primarily for individual home use (and may not be completely satisfactory for the professional translator.)

The principal rival to Systran in the mainframe and client-server market has long been the Logos Corporation. Its Multilingual Document Translation Software is designed particularly for the large multinational company users wanting high-quality customised systems.

Two long-standing commercial vendors are Globalink and Linguistics Products. The former's Power Translator is available in versions for both home users (recently superseded by Telegraph) and for professionals. Globalink (now owned by Lernout & Hauspie) also markets a system for company intranets, and its MT systems are accessible directly over the Internet through its Comprende on-line service. The PC-Translator from Linguistic Products Inc. has a smaller range of systems type, basically only suitable for occasional home use and its output quality has a lower reputation.

The Transcend systems (originally from Intergraph and now from Transparent Languages Inc.) also cover French, German, Spanish, Italian and Portuguese. The standard range versions are priced at the mid point and aimed primarily at professional users.

However Transparent Languages has recently introduced Enterprise Translator, a client-server system for company translation services, and Easy Translator, a cheap model designed for the home market. Both are based on the same MT engine as Transcend.

In addition there are a growing number of cheap Windows-based systems, definitely for the less-demanding home user, some specifically for Web use.

German only

The Metal system for German-English and English-German translation, developed originally marketed by Siemens, is now owned by GMS (a division of Lernout & Hauspie) and available through the publisher Langenscheidt as the T1 range of systems. As with Systran and Globalink there are client-server versions (T1 Corporate and T1 Workgroup), systems for the professional user (T1 Professional) and systems for less-demanding users (T1 Standard and Standard Plus).

The rival to T1 in the market for German/English only systems is the Personal Translator PT from IBM, sold by the von Rheinbaben & Busch Verlag. Originally developed by IBM as LMT for its own internal use, the Personal Translator is, as yet, available only as a top-market system for professional.

Spanish only

There are a particularly large number of systems for Spanish only. Oldest are the two systems from the Pan American Health Organization, SPANAM and ENGSPAN. Originally designed as internal mainframe systems, both are now available as PC Windows and intranet systems. Langenscheidt markets a version of T1 Standard Plus for Spanish/English. In addition there are a growing number of cheaper US products, intended for less-demanding users, e.g. Spanish Amigo, Spanish Scholar, Translate Instant Spanish, etc.

Scandinavian languages

For many years the only systems for Scandinavian languages were the PC-Translator systems for Danish, Norwegian and Swedish which were of relatively low quality. Now there are also Winger (for Danish/English), NeuroTran (Norwegian/English), and in particular TranSmart (from Kielikone) for Finnish-English translation. Nevertheless, this market is still undeveloped in comparison with other European languages.

Russian

The commercially most successful systems for Russian come from Project MT Ltd. (St Petersburg). Its range of systems, originally called Stylus, are marketed now as ProMT, each version available for Russian/English, Russian/French, Russian/German, and Russian/Italian. The range includes systems for home use, professional use, and specifically for Internet (Web) use (ProMT Home, ProMT Professional, and ProMT Internet respectively). In addition, the company has developed systems for translation into French from English and from German; these are marketed under the name Reverso through Softissimo (Paris, France).

The main rival to the ProMT Russian/English range is the PARS system (from PARS Ltd. and Polyglossum Inc.), intended for professional use. More basic are the low-priced LTGold and Lexi-Trans.

Other European languages

Other languages of Europe have been neglected by the main vendors. There are no 'professional' quality translation systems for Greek, Hungarian, Rumanian, Serbian, Croatian, Catalan, or any of the Celtic languages. However, Polish is covered by Lexi-Trans, Czech by a system called SKIK, and Ukrainian by the PARS company, with systems both between English and Ukrainian and between German and Ukrainian.

Japanese

In the 1980s, nearly all Japanese computer companies developed and began marketing MT systems, predominantly between English and Japanese. In recent years, many

more systems have appeared, a large number specifically for Internet/Web use, which are obviously meeting a great demand in Japan.

The older mainframe or workstation systems are now marketed also in Windows versions for either English to Japanese or Japanese to English, or both together in one product. Examples are ATLAS from Fujitsu, HICATS from Hitachi, ASTRANSAC from Toshiba, Pensee from Oki Electric Industry Co., PIVOT from NEC, and Duet from the Sharp Corporation. Nearly all are now available in versions for client-servers, for professional use, for home use, and for Internet and Webpage translation (in the latter cases sometimes under new names, e.g. Toshiba's system for the Internet is Honyaku Internet, and Fujitsu's is called TransLINGO.)

Competition in the market of systems for professional use includes highly regarded products from companies of US origin. The LogoVista system for English to Japanese is a joint product of the Language Engineering Corporation and the Catena Corporation (Japan), which also markets cheaper PC systems for home use. LogoVista comes in version for all types of user, from the large translation company to the Internet home user; it is also one of the few system also available for Macintosh equipment. Neocor Technologies (now owned by Lernout & Hauspie) has developed two systems: Tsunami for English to Japanese, and Typhoon for Japanese to English, both with impressive ranges of useful facilities, and both with notebook versions on sale.

From Japan appear almost every month a new cheap system for English-Japanese and Japanese-English translation. Many of them are designed specifically for Internet use and for translating Web pages. How many of these systems will survive more than a few years cannot be said. In any case, few of them are obtainable outside Japan. Some examples are JEBank, and EJBank (Kamejima), Honyaku Kobo (Matsushita), J*London (Kodensha), Korya Eiwa (Catena), My Translator (Nagase), Transer (numerous variants, Nova), TransLand (Brother), The Translator (Catena), TransSupporter (Sanyo), etc.

Chinese

In contrast to Japanese, there are still few Chinese/English systems of reasonable quality. Most systems are intended for primarily non-professional use (interactive composition), e.g. the CITAC Translator, ESP Plus, Marco Polo, HansBridge (English-Chinese only), and TransPerfect. However, there are Power Translator systems from Globalink for both language directions, and Systran Enterprise and Pro systems for Chinese into English.

Korean

The situation is much the same for Korean/English. Systran Enterprise has versions for Korean (for translation in both directions), otherwise there are lower-quality systems (mainly, except for Enkor, merely electronic dictionary aids.) However, for Japanese/Korean a number of the Japanese manufacturers have developed versions of their systems, e.g. Hitachi markets HICOM-MT, NEC has a mainframe version of PIVOT, and at the cheaper end Kodensha sells J*Seoul.

Arabic

There have also been relatively few systems for Arabic, despite the obvious potential market. TranSphere from AppTek Inc. (now owned by Lernout & Hauspie) is a client-server system for English-Arabic translation, and has probably the highest reputation. Other newer products known to be available for English to Arabic are Al Mutarjim Al Arabey (ATA Software) for professional translation (with a companion Al-Wafi for home use), An-Nakeel El Arabi, and the CAT translator from SAKHR (also Arabic to English). Finally, there are the MLTS systems from Cimos (France) for English or French into Arabic, and a version of the Lexica system from South Africa (see below.)

Hebrew

Only one system for Hebrew/English appears to be marketed at present; this is Targumatic (available from World Language Resources), a system mainly for non-professional use.

Other languages

In recent years, the EPI-USE Systems (Pty) Ltd. (South Africa) has started marketing a number of systems for translation between English and various languages of Southern Africa (Afrikaans, Sotho, Swahili, Tswana, Xhosa, Zulu) The systems are known as Lexica in South Africa, and are marketed elsewhere as Pink Translator by Pink Computers. The Lexica range includes also software for English to and from Arabic, French, German, Italian, Portuguese, and Spanish.

Many languages are still not covered by commercial systems, e.g. Indonesian, Malay, Vietnamese, Thai, all languages of the Indian sub-continent (not even Hindustani, Urdu and Bengali), and, apart from those just mentioned, most African languages.

Translation workstations

In principle, most translator workstations are designed for use with a wide range of languages, made possible by the absence of programs for linguistic analysis and synthesis. Only the need for greater sophistication in the alignment programs of translation memory systems makes them less suitable for some languages, particularly non-European languages. Although designed initially and primarily for languages using Roman alphabets, more workstation vendors are producing versions suitable for use with languages such as Arabic, Chinese and Japanese.

Conclusion

After many years of development, commercial MT systems are now capable of serving well the demands of multilingual companies and professional translators seeking cost-effective production of good-quality translation for dissemination purposes. This is particularly the case for translation between the major languages of the global marketplace. There remain many gaps for 'minor' languages, including those of Eastern Europe, Africa, and India.

Systems for assimilation purposes (for the less-demanding 'occasional' user) are also widely available, with good language coverage on the whole. However, these systems often give poor quality output, even for well-written source texts, let alone the low-level writing on electronic mail and other Internet applications. There is clear need for improved quality in this area of commercial software, and for some consumer guidance for potential purchasers in order to minimise dissatisfaction.