

Language Tools at the EC Translation Service: The Theory and the Practice

Dimitri Theologitis¹
Computer Aids for Translation
European Commission Translation Service

CCE JMO B2/21
L-2920 Luxembourg

dimitrios.theologitis@sdt.cec.be

Language technology brings new tools to the translator's desktop: full-text retrieval systems, terminology systems, translation memories and machine translation. European Commission's Translation Service, SdTVista is used for full-text search of reference documents, the powerful aligner in Euramis creates translation memories which are stored in a central Linguistic Resources Database, sentences not found in the memory are automatically sent to the EC Systran machine translation system, and relevant terminology is retrieved in batch mode from the Eurodicautom termbase. The resulting resources are brought together on the Translator's Workbench².

These tools are being used extensively today by about one third of SdT translators. While it would be premature to cite global statistics, there are strong indications that substantial gains in productivity can be achieved. The 1996 machine translation survey, the experience of the Translation Workshop between 1996 and 1998, and the 1998 translation memory tools survey confirm this. The crucial factor is the assimilation of translation technology in the workflow of a large organisation — and this constitutes a real challenge for management.

THE THEORY

The Context

The European Union today has 15 Member States and uses 11 official languages. There are 7 European Institutions and Organs, each one of which has its own translation service to provide for its multilingual needs. The European Commission alone has some 16 000 officials, 2 000 of whom work for the Translation Service (referred to from now on as *SdT* for *Service de traduction*). Over 1 000 000 pages are translated per year (some 275 000 000 words).

Basic Assumptions

Repetition

As a result of the way the European Commission works, there is an inherent degree repetition in the documents that it produces and that the SdT translates. A high-level political decision taken at, say, a European Summit meeting will normally spawn legislative action such as Regulations or Directives. For this, the Commission will

¹ The author wishes to thank JOHN BEAVEN for feedback and comments.

² A complete description of the language tools used by the EC Translation Service can be found in issue 1.1998 of "Terminologie & Traduction", the professional journal of the language services of the European Union's Institutions, http://europa.eu.int/comm/sdt/bulletins/t&t/index_en.htm

submit proposals to the European Parliament and the Council of Ministers, based on working documents. All of this written material will be quoting and re-using text on the same subject matter. This repetition process is reinforced if one takes account of the several iterations foreseen in the legislative process — let alone the exchanges and the modifications between Institutions.

The Need for Terminological and Phraseological Consistency

Almost half of the translations (46%) carried out in the SdT arise from the Commission's legal obligation to publish in those languages, against 26% which are translated because of an operational obligation and a further 28% for which there is a political obligation. Particularly in the case of legislative texts, passages which have already been translated need to be *quoted*, and certainly never *re-translated* with the possible risk of producing diverging translations for the same source text. Terminology has to be consistent throughout whole series of documents about a given subject — stylistic variations are not welcome in this context!

The Needs of the Translator

In view of the above and for the purposes of this presentation, the basic needs of the translator can be summarised as follows:

Originals and reference documents are required in electronic form. Paper is not sufficient any more as it cannot be subjected to electronic treatment.

- **Relevant terminology.** Translators are often swamped in terminological noise — the problem of too much choice could be as important as the absence of terminology information.
- **Relevant phraseology.** Previously translated sentences are the smallest (nearly) autonomous linguistic building block.

And last but not least, if the translator can find someone else to do the job, in this case a machine, then so much the better! We will come back to machine translation later.

Major Systems in SdT

Based on the above needs, the EC Translation Service went about constituting a comprehensive computing environment adapted to translation work. This was possible thanks to a high-level office automation infrastructure comprising:

- High-end PCs linked to a wide area network between Commission buildings in both Brussels and Luxembourg.
- A backbone of Unix servers local to the Translation Service in two departmental computing centres in Brussels and Luxembourg.
- Ready access to the full power of the Commission's Computing Centre, including Unix and proprietary mainframe computers. This also means that a 24-hour, 7-days per week service including backups is available for key applications.

The major SdT systems fall into three categories:

Administrative Systems:

- POETRY (Processing of Electronic Translation Requests) which is used by requesting departments to transmit the text to the SdT.

- WINSUIVI (follow-up) for basic workflow management within SdT.

Document Management Systems

- SDTVISTA, a document storage and full-text retrieval tool.
- A project has just been launched for a new system to allow easy one-stop access to major electronic documentation sources across in-house and external documentary databases, CD-ROMs and the internet.

Linguistic Systems

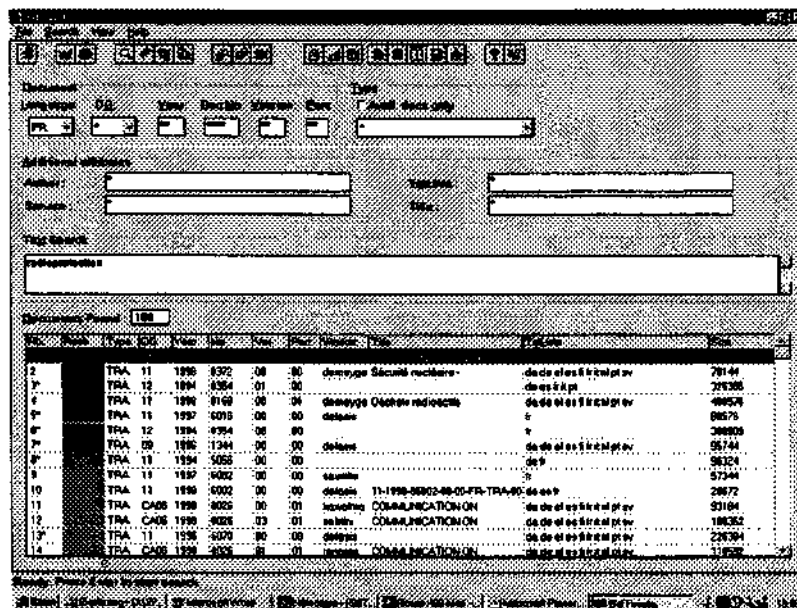
- Terminology: EURODICAUTOM, a central terminology database and TRADOS' MULTITERMPLUS!, used as its local feeder software. A One-Stop Terminology Shop is also being set up to facilitate searches through multiple resources.
- Translation memory: EURAMIS, a central linguistic resources database and TRADOS' TRANSLATOR'S WORKBENCH (TWB), used as a local translation memory front-end.
- Machine translation: EC SYSTRAN, the Commission's "flavour" of the well-known SYSTRAN machine translation software.

We shall now take a closer look at SDTVISTA and the linguistic systems.

SDTVISTA

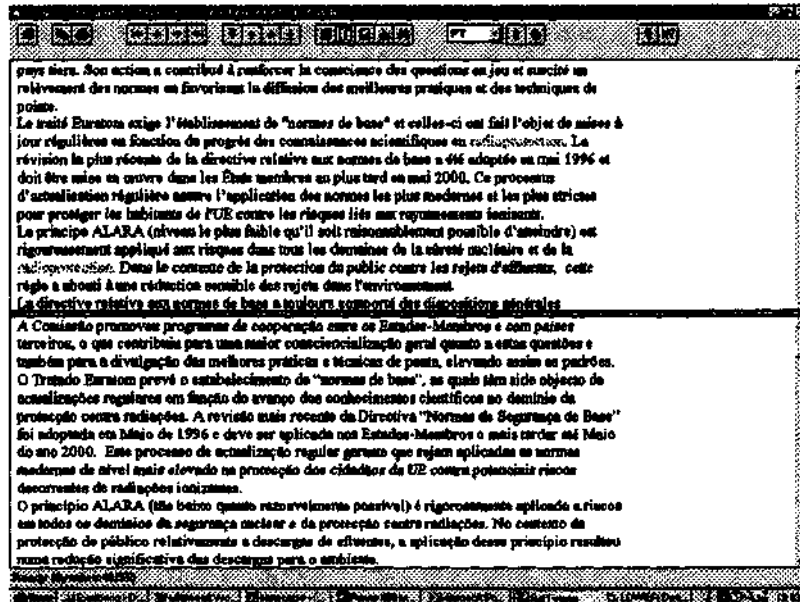
This is the major database for storing originals and translations carried out in SdT. As far as originals are concerned, more than two thirds arrive in electronic format. This includes nearly all documents produced by the internal Commission departments. Documents for translation coming from Member States' administrations however only arrive in electronic form if there is continuous exchange of information between the Commission and those services.

Currently, there are four years of text on-line. A full-text index has been created using "Search-Tools" by Fultext (Fulcrum), meaning that any translator can search for any word in any document. The system resides on the Computing Centre's computers and searching is astoundingly fast.



SDTVISTA Search

A facility to display two language versions of the same document in parallel, albeit only using somewhat crude positioning criteria (not aligned texts), is highly appreciated by translators. The immediate advantage of such a system is that it doubles as a powerful terminology utility, frequently used in the absence of any automatic or semi-automatic tool that would extract terminology by parsing source and target language documents.



SdTVISTA Parallel Scroll

The downside to it is that all imperfections in translation become immediately apparent and would have to be dealt with by the special SdT terminology unit which, together with the translating colleagues, has to assume a normative role and indicate preferred translations. This mechanism, although possible in theory and could be supported by appropriate tools such as MULTITERM and EURODICAUTOM, is in practice implemented only rather slowly.

The Linguistic Systems

Terminology

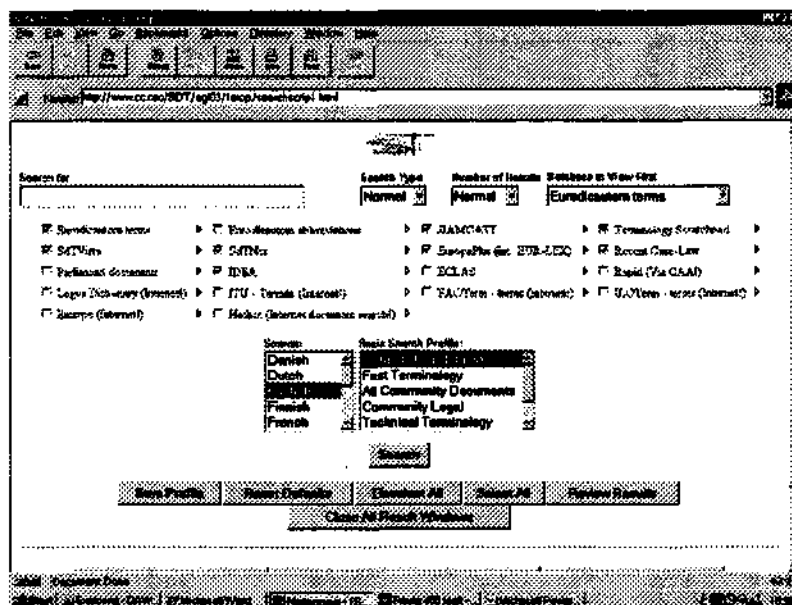
EURODICAUTOM

This is the Commission's central terminology database, accessible to all European Institutions and to the outside public on the internet (currently <http://www2.echo.lu/edic>, but this could change after September 1999). Started in 1976, it has consistently been updated with the terminology used in the domains of activity of the European Union. It contains some 5 200 000 terms and 300 000 abbreviations in 1 200 000 multilingual entries. At the moment it is undergoing major consolidation work to eliminate duplicates and retrieval "noise".

The One-Stop Terminology Shop

A prototype web-based search solution has been set up to help the translator with searches through multiple resources. For the moment it is mostly a set of search scripts, but it is scheduled to evolve into a full-fledged solution with subsequent linguistic treatment of the retrieved information including multilingual display of keywords in context (KWICs) and of document parts. The following screenshot is only presented as

an indication of the wealth of information that can be tapped with a simple web interface.



The 1-Stop Terminology Shop

Translation Memory Systems

Translation memory systems have been available on the market for some time but have been targeted at the individual translator or the small-to-medium translation agency. With some 2 000 potential users, SdT decided to take over in 1996 the continued development of EURAMIS (European Advanced Multilingual Information System), a research & development project started by Directorate General XIII (Telecommunications, Information Market and Exploitation of Research) under the Multilingual Action Plan (MLAP) and headed by Jean-Marie Leick. The systems reside physically on several computers in the SdT and the Commission's Computing Centres.

EURAMIS is made up of two main parts:

The Linguistic Resources Database

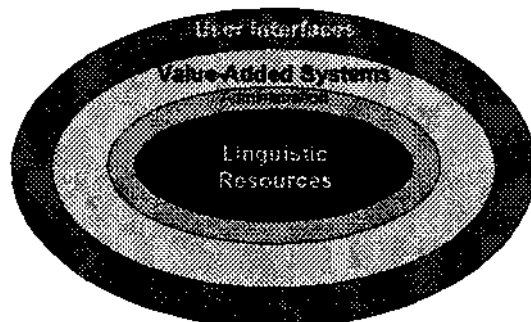
A database designed as a generic container of linguistic objects ranging from simple terms, through expressions and sub-sentence segments, up to whole sentences. For the moment it is mainly used as a central translation memory (and a database of sub-sentence segments), but will soon contain EUODICAUTOM and later the SYSTRAN dictionaries. The idea is that all systems can access the same linguistic data, thereby creating consistency and synergies. Depending on the type of system, a particular view of the data is presented to it.

The Multilingual Server

This contains the dispatcher and service sequencer, as well as the multiple retrieval facilities. Document format conversions are carried out in the dispatcher, and the sequencer comes into play whenever multiple products are requested. An example of this is the so called "Terminology from Text" search. In this, a document is scanned for terms and expressions using the SYSTRAN parser and morphology module as a lemmatiser to put the identified terms in canonical form. The term list is passed through EUODICAUTOM and the result of the search is formatted in either native

EURODICAUTOM format, as a Word document, or in MULTITERM import format. It is then sent back to the requester.

Ling Resources Philosophy

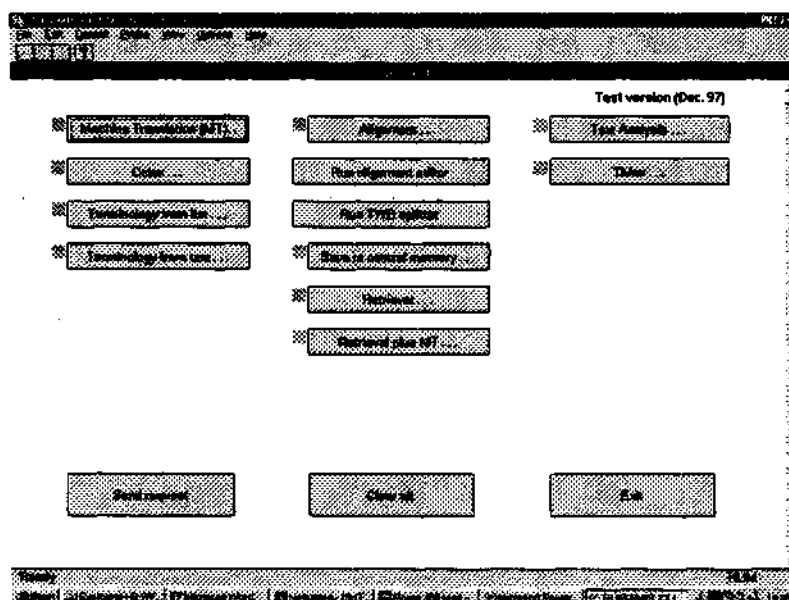


SdT

SdT Linguistic Resources and Systems

As far as the retrieval functions are concerned, the following screenshot gives an idea of the products available. It also gives an indication of the problems faced by a service which is not a software developer and as such does not have the necessary expertise in user-interface design!

EURAMIS is a batch treatment interface, i.e. there is never any direct connection with the system. The interface creates an SGML command file with instruction strings to which the texts to be treated are attached. The message is sent by e-mail to the dispatcher and the results are returned to the requester in the same way.



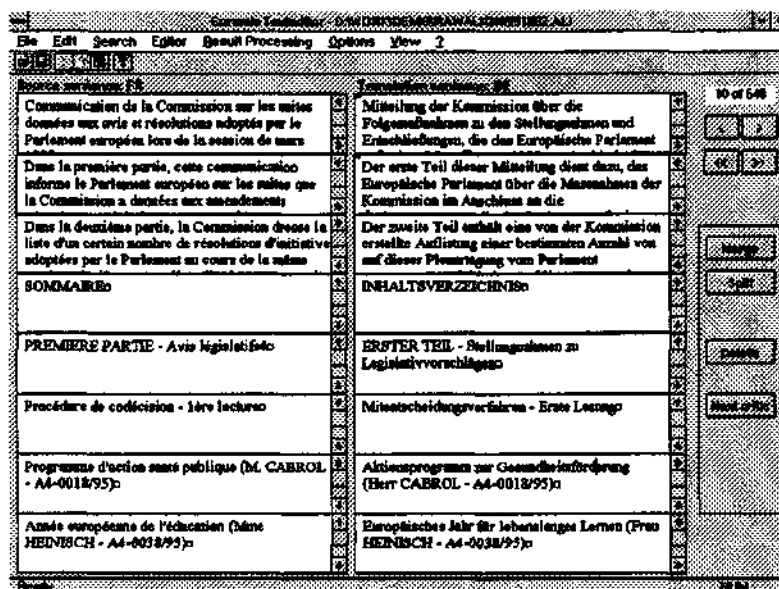
EURAMIS Interface

The leftmost column deals with SYSTRAN-related requests, ranging from raw machine translation through terminology from text (see above) to CELEX (EU legal database) retrieval. In this services, a parser is used to put references to EU legislation are put in canonical form. The relevant document titles or even the full documents themselves are then extracted from CELEX. The titles are certainly useful for citations, but, more

importantly, the documents can be aligned in the background and the user can receive a temporary translation memory made up from the legal documents referenced in the text.

The rightmost column launches the text analysis modules (more about this below).

The middle column provides the translation memory functions. At the time of the design of the system, commercially available alignment software was of limited functionality and so DG XIII and SdT decided to build yet another aligner, in the process making it more suitable to the types and formatting idiosyncrasies of their documents. Quality remains more than competitive even compared to the latest crop of commercial aligners, especially now that it can take into account "anchors" in the source and target texts. An alignment editor was also added providing all necessary functions including editing, merging, splitting of segments, and search and replace operations.



EURAMIS Aligner

The functions "Retrieval" and "Save to central memory" are self-explanatory; "Run TWB splitter" calls a small routine to split Trados translation memories according to particular sets of attributes. One particularly interesting menu point is however "Retrieval plus MT". There, the system accesses the translation memory and all matches above a user-defined fuzziness threshold are retrieved. For the remaining segments, a SYSTRAN machine translation is preferred. This then combines the best of both worlds: human revised translation is preferred wherever available and machine translation is only used for the rest.

As an added feature, results can be presented in different formats, such as Trados translation memory import format, MULTITERM import format, and Microsoft Word. The latter is quite useful in cases where the translator does not want to work in the TRANSLATOR'S WORKBENCH environment. The retrieved resources are then incorporated in a hybrid Word document and marked in different colours according to their origin (full or fuzzy translation memory matches, or machine translation).

T-Man

Initially a batch-replacement tool developed in-house by Roger Bennett for sub-sentence level segment replacement, including regular expressions, using the "sed" stream editor. These functions have now been integrated into EURAMIS. Constitution of

the segment database is basically a manual operation, but some routines have been developed for automatic repetitive string identification. An editor showing parallel sentences helps introduce the translations of the repetitive strings.

The translation of the monthly "Bulletin of the European Union" (<http://europa.eu.int/abc/doc/off/bull/en/welcome.htm>) has been rationalised in this way for the past five years, with great savings in time and translation consistency.

EC SYSTRAN Machine Translation

Much has been written elsewhere about the European Commission's successful implementation of the SYSTRAN machine translation system. For the purposes of this presentation let it be said that SYSTRAN has been in use at the EU since 1976, is available to every official of every European Institution over e-mail and the interinstitutional intranet, has processed some 260 000 pages in 1997 and is expected to do at least 30% more in 1998. 17 language pairs are available:

- Spanish into English and French
- German into English and French
- Greek into French
- English into Spanish, German, Greek, French, Italian, Dutch and Portuguese
- French into Spanish, German, English, Italian and Dutch.

Below is a short example of the output that one can expect a well-tuned system to produce:

Original French Text

Pour éviter ces limitations, certaines administrations utilisent l'approche du développement humain du PNUD (I, IRL).

Tous les participants ont également exprimé l'opinion selon laquelle l'approche et les méthodologies, notamment la méthode des «poverty assessment», appliquées par la Banque Mondiale depuis 1990, constituent des instruments d'analyse très utiles pour saisir les situations de pauvreté dans les pays bénéficiaires et pour bien cibler les programmes d'assistance qui visent à lutter contre cette même pauvreté.

EC SYSTRAN Raw English Translation

To avoid these limitations, certain administrations use the approach of the human development of the UNDP (I, IRL).

All the participants also expressed the opinion according to which the approach and the methodologies, in particular the method of «poverty assessment», applied by the World Bank since 1990, constitute very useful analysis instruments to grasp the situations of poverty in the beneficiary countries and well to target the assistance programmes which aim to fight against this same poverty.

MT Developments

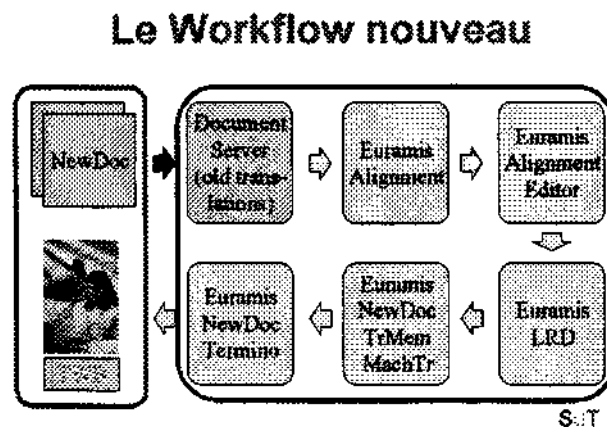
After a spell of rather cool relations, Systran S.A. is once more working together with the European Commission for the further linguistic development of the system in accordance with the Commission's linguistic specifications — but also to port EC Systran to a new Unix platform. This should allow much better integration with the rest of the computer environment of the Commission, such as easy on-the-fly intranet-based

translation à-la ALTAVISTA (<http://babelfish.altavista.com/>), but also real-time translation from within the word-processor over the LAN. It will also mark the end of the Assembler-code era (SYSTRAN has been running on the same IBM-compatible mainframe environment for over 20 years...).

The New Workflow

All the tools available and the various integrations between them constitute a powerful translation technology environment, which can greatly help the translators in their work. Previously, a translation job would arrive from the requesting department to the translator's desk in "free fall", so to speak, through all the administrative stages but with hardly any check on the content of the document. The translator then would have to go back up the whole chain and ask for reference documents, terminology and other available documentation. All this multiplied by ten, of course (or by as many target languages as were requested).

Now, using computer technology, it is possible to rationalise document preparation to a great extent. It is also possible for the same person to do this for multiple languages, thus further contributing to the savings. What follows is a possible scenario, which, while it might appear maximalist is by no means rare in today's SdT.



The New Workflow

Constituting a Translation Memory

Using SDTVISTA it is easy to search for reference documents and their translations in the SdT archive. This search is also scheduled to be automated in the future, using segments from the new text and matching them with the contents of the archive.

The retrieved documents are then aligned and the output edited using the EURAMIS aligner and alignment editor. The resulting memory is saved in the EURAMIS Linguistic Resources Database.

Resources for the New Text

Using the EURAMIS interface, the new text is sent to the server for the following resources, according to the requirements described earlier in the paper:

- A relevant translation memory. Only sentences pertaining to the text at hand are retrieved.

- Relevant terminology. Using the "Terminology from Text" function, all the contents of EURODICAUTOM matching words or expressions in the text are retrieved, and the results transformed into MULTITERM format.
- For sentences not found in the translation memory, a suggestion from SYSTRAN. Titles of legislative acts from CELEX are also added.

The Actual Work

All these resources are sent to the translator via e-mail and are imported into the TRANSLATOR'S WORKBENCH to be available on first demand. Everything is at the translator's fingertips, with no need for standard documentation and terminology searches any more.

Finally, the translated texts are post-processed so that translation memories can be updated. Feedback is also sent to the machine translation team.

A wonderful world then? Well...

THE PRACTICE

Access to Systems, Potential and Actual Users

Systems are available throughout the Translation Service, some to the whole Commission, some to all the European Institutions while others are public:

- SDTVISTA, the translation and text archive, is available to the whole of the Translation Service (2 000 potential users, translators and secretaries). Through a web interface, individual Directorates General have access to the parts containing their own documents.
- EURODICAUTOM is open to the public through the web interface. It is being intensively used by translators and terminologists, some of whom have it open all day on their desktop. The Commission has a site license for MULTITERM, which is part of the standard SdT desktop. In practice, some 1 500 translators use those terminology tools regularly.
- Access to SYSTRAN is also generic for all the European Institutions, but the number of actual users is restricted by the available linguistic combinations. In practice, SYSTRAN is being used by some 600 translators and secretaries in SdT and by some 2 500 administrators and secretaries in the Directorates General. Usage by the other Institutions accounts for some 10% of the total number of requests. The Commission's SYSTRAN license covers also use by the European public sector and lately there has been a growing number of users from ministries and universities in Member States. It is too soon to produce figures, but at present the Commission gets some three new requests for access to MT facilities from outside every month.
- EURAMIS - TRANSLATOR'S WORKBENCH. While EURAMIS is a product developed in-house and thus everyone in SdT has access to it, the licensing scheme with Trados for the TRANSLATOR'S WORKBENCH foresees payment by individual licenses and so, for budgetary reasons, these have to be carefully dosed. EURAMIS and TWB are meant to be used together and are thus normally installed in tandem. At the end of 1998 there will be 500 licenses of the tandem installed, which covers more than one third of the translator population.

Productivity Gains?

Let it be said up front that at the present stage, no affirmations about the general impact in productivity can be made. There are positive indications, but since these tools, and especially translation memories, have been operational for only about one year, any interpretation of the statistics in that sense would be premature. Here are some data:

- Translation demand, which reflects general Commission activity, is more or less stable.
- The Translation Service has committed itself to having more texts translated by external freelance translators. In 1997 the percentage was 17% (16% in 1996).
- In anticipation of future enlargements, SdT has undertaken a reduction in the number of translator posts by some 2% a year.

As has already been stated, no trend is immediately apparent and any global rise in productivity could be due to a variation in demand and a higher internal workload. However, there have been two large-scale feasibility studies which included measuring of productivity and these are reported below.

Machine Translation Feasibility Study

When DGXIII had to withdraw from the financing of SYSTRAN development and maintenance, the Translation Service conducted a feasibility study in the summer of 1996 to examine whether it should take on the project. The study contained five points: a user survey, an examination of legal aspects, a market study for alternative MT systems, field testing of SYSTRAN and finally a cost/benefit analysis.

The Machine Translation Usage Survey

Response rates

Response rates were good overall (despite the fact that the survey was conducted over the summer period) so that results can be considered to be representative:

- SdT machine translation users: 71% (370 out of 520)
- SdT non-users: 34% (403 out of 1180)
- Users from other Directorates General: 30% (735 out of 2428)
- DG non-users: 34% (278 out of a random sample of 817)

In total 1 786 responses were received from a total surveyed population of 4 945 officials.

Results

From the SdT machine translation users:

- 67% consider machine translation useful. "Useful" here meaning that the translator saves time compared with pure human translation.
- 72% gain time by post-editing a raw machine draft.
- 27% of translations would have been late but for machine translation.

From the SdT non-users of MT:

- Two thirds do not use it because their language pair is not covered. Let it be said here that although the possible number of language combinations with 11 languages is 110 and EC SYSTRAN covers only 17 language pairs, these 17 cover two thirds of the total human translation requests in SdT. However, we consider that 8 of the available language pairs are really operational (the ones with Spanish, English, French and Italian) and these cover one third of human translation demand, and hence the figure of two thirds mentioned above.
- 45% claim not to know how to use it! In the meanwhile the machine translation team has caught up with an information campaign, but this figure illustrates how software is perceived by a large number of users: SYSTRAN as a software product has no real user interface; the user merely has to send an e-mail message to an address which can be found in the Commission's electronic address book. This shortcoming has been addressed with a simple dedicated Windows interface, and now the EURAMIS interface, as well as a web interface on the intranet of the Institutions.

Users from the Directorates General:

- 53% consider more than half of MT output useful.
- 95% gain time thanks to machine translation.
- 59% would have sent more documents to SdT for translation but for MT.

Non-users from the Directorates General:

- 20% had no need for translation at all; the sample was random and inevitably included officials without any translation needs, such as couriers and drivers.
- 63% translate short texts themselves.

From the above results the Translation Service concluded that machine translation was useful both for its internal needs but also for those of the Institutions) and decided to take over the project.

The Machine Translation Productivity Survey

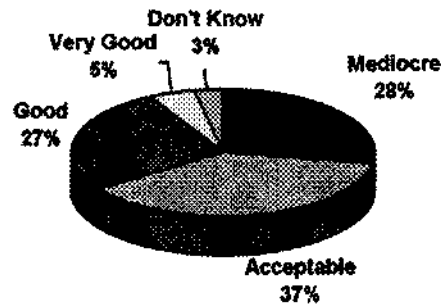
Since variations in global SdT productivity could be difficult to interpret, it was decided to conduct a large-scale experiment on the field by having MT users fill in a detailed report after every translation they did on the basis of a raw MT draft. A total of 104 users participated in the survey covering 12 language pairs, 3 162 pages were translated (some 700 000 words) and 285 reports were handed in.

Results

The results were conclusive: The average translation output using a raw machine draft as a basis was 60% higher than with using traditional methods. In other words, our translators reported that they could produce some 16 pages in the time it would normally take them to produce 10.

Subjective impressions

At the same time interesting remarks were made regarding the perceived quality of MT output:



Perceived Machine Translation Quality

Translation Memories

There are two sources of information about the impact of translation memory tools on the SdT:

- The experience of the so-called "Translation Workshop" (*Atelier de traduction*) from 1996 to mid-1998.
- The translation memory tools survey of June 1998.

The Translation Workshop

It was felt in 1996 that the introduction of translation memory tools that change fundamentally the way traditional translators work needed careful preparation. It would be necessary to identify the types of documents that lend themselves best to translation memory treatment. The tools themselves had to be integrated into the computer environment and in particular with the central translation memory tool, EURAMIS. New user guides adapted to the workflow of the Service as well as a new training scheme had to be drafted. The then top management of the SdT decided to go ahead with the operation and so a multilingual team of some 25 translators and secretaries was created in Brussels.

Productivity considerations

Judging by the Translation Workshop statistics, productivity gains using the tandem EURAMIS/TRANSLATOR'S WORKBENCH were the same as with machine translation: some 60%. No results are available yet for the combination of translation memories and machine translation (the "Retrieval plus MT" function of EURAMIS), but preliminary discussions indicate that the increased productivity will certainly not be the product of the two.

Initial Investment

One finding which came quite clearly out of the Translation Workshop experience is that initial investment in setting up the memories for a whole Institution can be high. Human resources need to be allocated to find reference documents, align them and store them in the memory. Co-ordination is also necessary to make sure that all 11 languages advance at the same pace and that memories are truly multilingual and have no gaps. This investment should diminish over a time-span of two years, when most of the repetitive documents, many of which are annual reports, have already appeared twice. Actual size of the EURAMIS memories is some 500 000 segments in 11 languages.

Training

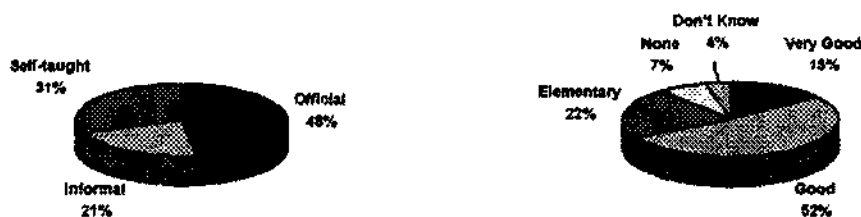
The Translation Workshop organised the training of translators and secretaries in translation memory tools. The approach was quite innovative: trainees in teams of 20 were asked to move into the Workshop for three weeks. There, 10 experienced colleagues took them through real hands-on training using the trainees' own documents as material. During these three weeks they also followed a number of short seminars of increasing complexity. One interesting result was that total translation output during these three weeks was little less than what would normally be expected from a comparable team of traditional translators counting trainees and trainers together!

Other translators in Luxembourg and Brussels were trained in a more traditional way and some even taught themselves. After the initial training period, a survey was carried out and here are the results.

The Translation Memory Tools Survey

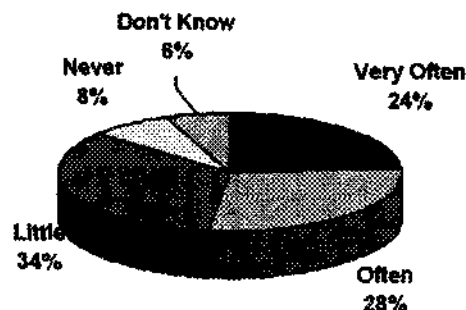
In June 1998 some 390 EURAMIS and/or TRANSLATOR'S WORKBENCH users were asked to reply to a questionnaire regarding their knowledge and use of TM tools. 263 responses were received (199 translators, 63 secretaries and 1 administrator), i.e. 67%.

The following graphs summarise the responses to the main questions.



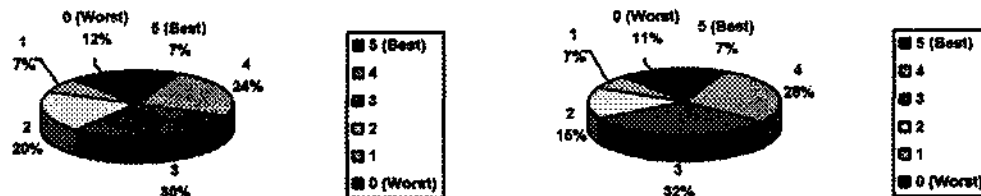
EURAMIS/TRANSLATOR'S WORKBENCH Training and Perceived Knowledge

It is interesting to note the next question, as to the usage of translation memory tools after the trainees returned to their translating units. This can be regarded as a test for the organisation, that is, how quickly can the organisational environment adapt to the introduction of newer technology. The results, for secretaries and translators alike, are encouraging: More than 80% use translation memory tools in their work.



Use of EURAMIS/TWB Licenses in Translating Units

The two questions that follow give the subjective impression of translators and secretaries as regards the speed and the user-friendliness of both EURAMIS and TRANSLATOR'S WORKBENCH. Although the two products are quite different, the first being basically an e-mail request interface while the latter a real-time TM manipulation interface, the results were nearly identical and are given together. A third question about ergonomics gives results very similar to the ones on user-friendliness and is not reproduced separately.



Speed – Translation Memory Tools Perception – User-Friendliness

As a last item, the follow-up tools of SdT are in the process of being adapted and in the future statistics will be available on the use of translation memory and machine translation tools.

Organisational Aspects of Translation Memory Tools

On the whole, installations and training are progressing according to plan. Whilst general progress is satisfactory, the single most important factor hindering the full-scale introduction of translation technology is the relative slowness of a traditional organisation to adapt. Translators and management alike need to adopt an attitude of sharing linguistic resources — something which in the eyes of some exposes them to criticism of their work. On the positive side, there is a strong push from translators themselves who see their work facilitated on the one hand, and on the other from top management which is obviously very sensitive to the qualitative and quantitative improvements possible.

The Translation Assistants

SdT used to have a tradition of one secretary for every four translators going back to the days of the typewriter and the dictaphone. This is not necessary any more as all members of the Service are equipped with their own PCs. Most translators now type their own texts, others dictate, have a secretary type the text for them and only do the revision on screen.

The TM survey however brought up another aspect: 75% of translators have asked a secretary to prepare a "dossier" for them, that is do all the necessary pre-treatment work, including the setting up of the translation memory and/or requesting a raw machine translation. As it is certainly more interesting for a person to do this type of work rather than sit all day in front of a PC typing, secretaries embrace the new technology all the more.

Indeed, what is happening is that we are witnessing the emergence of a new specialisation, that of the "Translation Assistant". This is someone who, while not having all the language skills necessary to produce a finished translation, still knows enough to take a text, with the help of computers, to a point where it just needs revising by a competent linguist.

Perspectives

As far as installation of technology is concerned, translators exert pressure to install tools — so these must be working after all! The best test of a tool is whether users miss it if it is not there anymore or if it does not work the way it used to; the calls to the help desks are a good indication that this is the case and that translators by now rely on the tools for their work.

On the other hand, the translation workflow starts to adapt. Numerous initiatives have been set up for central pre-treatment for all languages before the document reaches the translator or for "document spotting" to make sure that texts that can be dealt with technology do indeed get the treatment. These initiatives now need to be co-ordinated at a Service level. As mentioned earlier, workflow tools are in the process of being adapted to include technology treatment data.

Automatic global pre-treatment is also in sight. With computer hardware prices tumbling and systems becoming more reliable, discussions are underway to give all documents the maximum possible technological support, regardless of the solution that the translator is going to use. Statistical information will merely indicate the extent to which each tool contributed in the pre-treatment, such as the percentage of segments retrieved from the memory, the success rate of automatic reference document searches, and the percentage of hits from the machine translation dictionaries, so that an educated choice can be made.

Finally, there is so much technology available that the issue will not be whether to use it or not, but rather — and that is the challenge — to find the most cost-effective combination of human and machine input for a given document type and for a given linguistic quality requirement.