EUROPEAN COMMISSION
TRANSLATION SERVICE

Directorate of Resources and Language Support
Analysis of Needs and Multilingual Tools

EC SYSTRAN: THE COMMISSION'S MACHINE TRANSLATION SYSTEM

by Angeliki Petrits
MT Management Team
European Commission Translation Service (SDT)

- Revised 30 August 2001 -

Adapted from
SYSTRAN Development at the EC Commission from 1976 to 1992
by Ian Pigott

TABLE OF CONTENTS

1. Introduction

For more than two decades the European Commission has been developing and adapting a multilingual machine translation (MT) system for internal purposes. In its current state, the Commission's system – EC SYSTRAN - contains 18 language pairs, in which English, French, German and Spanish play key roles (see *Figure 1)*.

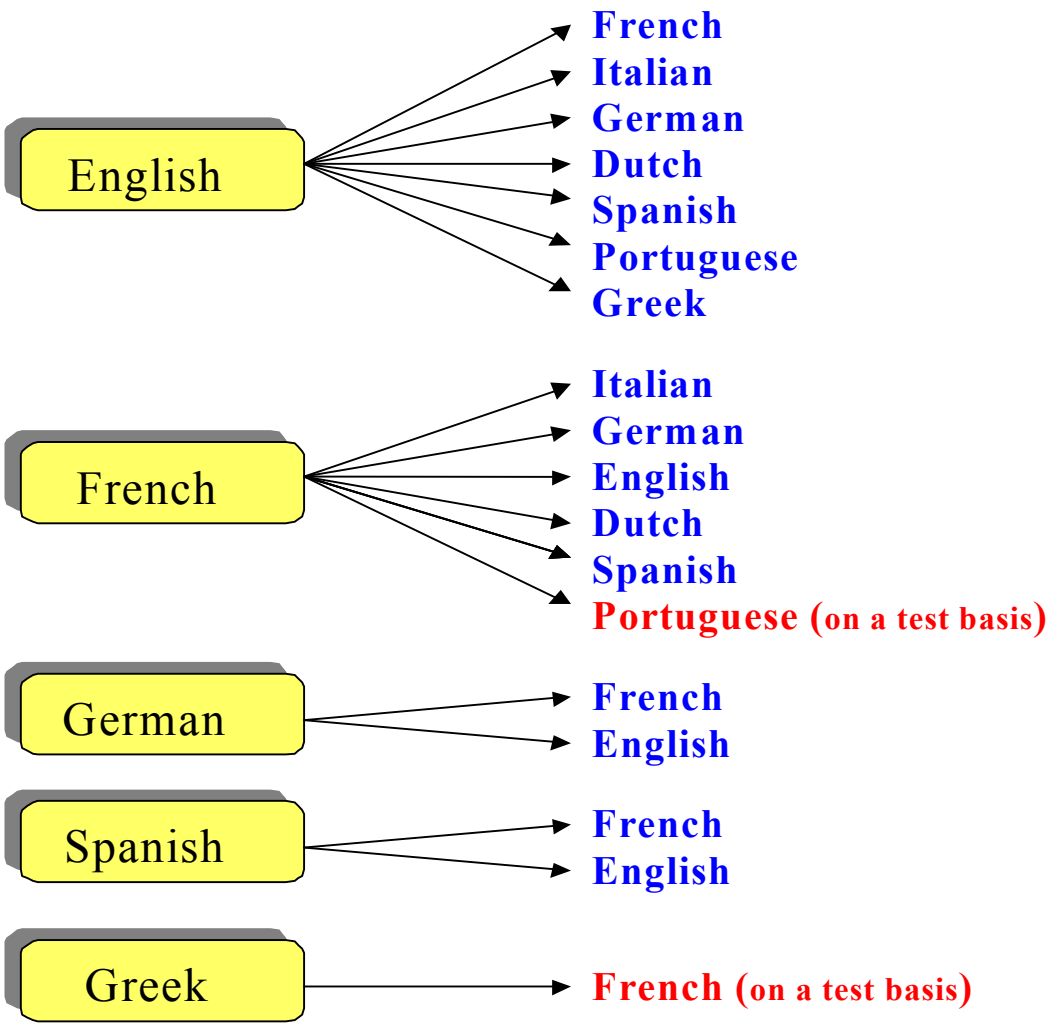

Figure 1 - Language Pairs

These pairs provide translation from English into French, Italian, German, Dutch, Spanish, Portuguese and Greek; from French into English, German, Dutch, Italian, Spanish and Portuguese (on a test basis); from German into English and French; from Spanish into English and French; and from Greek into French. Test versions of the youngest language pairs - Greek into French and French to Portuguese - have been available since 1997 and 1998 respectively.

Traditionally funded by the Directorate-General for the Information Society under the research budget, EC SYSTRAN has since 1998 been run as an operational concern by the Translation Service.

2. Historical Background

A product of the Cold War, SYSTRAN (an acronym for *System Translation*) was invented in the USA in the late fifties by the Hungarian-born American Peter Toma. After moving to California in 1956, Toma put his convictions into practice for the development of a pragmatic approach to machine translation. Unlike most of his contemporaries, he did not believe that linguistics could provide a satisfactory solution to the computerisation of language, but was convinced that the analysis of language had to fit the existing computer technology.

The first operational language pair, Russian-English, was installed by the USAF in 1969. In 1973, NASA commissioned the development of English-Russian for the Apollo-Soyuz project, then in 1974 Toma's group applied the results of the English analysis to an English-French prototype. That's where the European Commission came in.

The Commission's Translation Service (or *SDT*, for *Service de Traduction*) is the largest in the world, with some 1,300 translators and terminologists translating over a million pages per year. The growth in the number of texts and languages, and in the technicality of the subjects handled, has led the Translation Service to seek ever more advanced technical solutions to enable it to fulfil its mission.

The Commission therefore acquired certain rights for SYSTRAN in 1976 from WTC (World Translation Center), Toma's company in La Jolla, California. It chose Toma's system because at the time it was the only operational fully automatic system available for English-French, and it was with this language pair that *EC* SYSTRAN took its first faltering steps in 1976, in the shape of a pilot project. On the basis of what were considered encouraging results, a French-English version of the system was created the following year, and in 1978 English-Italian was introduced.

The dictionaries were gradually extended to cover the main areas of European Commission activity. Moreover, with the 1980s came the arrival of new language pairs. There are now a total of 18 pairs available to staff throughout the institution.

Figure 2 over the page provides the year in which development started for each language pair and the number of translations in the one-word dictionary (STEM) and the expressions dictionary.

Year Development Started	Language Pairs	STEM Dictionary	Expressions Dictionary
1976	English-French	**78,821**	**69,824**
1977	French-English	**66,546**	**114,327**
1978	English-Italian	**68,626**	**54,017**
1982	English-German	**61,892**	**33,324**
1982	French-German	**53,392**	**57,269**
1984	English-Dutch	**47,002**	**11,606**
1984	French-Dutch	**33,987**	**23,860**
1985	English-Spanish	**90,974**	**41,974**
1985	English-Portuguese	**47,545**	**7,368**
1988	English-Greek	**65,693**	**21,345**
1988	German-English	**136,369**	**25,673**
1988	German-French	**82,055**	**17,031**
1989	French-Italian	**39,919**	**19,236**
1990	French-Spanish	**77,166**	**31,504**
1990	Spanish-English	**32,467**	**6,239**
1991	Spanish-French	**32,532**	**1,897**
1993	Greek-French	**29,014**	**1,519**
1997	French-Portuguese	**32,702**	**2,467**
		1,076,702	**540,480**

Figure 2 - Number of Meanings per Language Pair (December 2000)

But why has the Commission chosen to develop these particular 18 language pairs as opposed to some of the others from the 110 possible combinations of the current 11 official EU languages? There are three main criteria.

1) **Internal needs of the Commission.** Since English and French are the main working languages, it was natural to start with these two languages as both source and target and to add later Italian and German.

2) **Translation quality expected from related languages.** As a result, there has been a strong preference for developing language pairs involving combinations of either Romance (French-Italian) or Germanic languages (English-German), because we predicted that with a minor effort satisfactory results could be obtained - which proved to be true, at least for the Romance languages.

3) **Budgetary restrictions.** Financially, it was impossible to develop 110 language pairs, which would have been ideal. So, if a Member State was willing to co-finance the development of a particular language pair, the Commission would give greater priority to that development. This was the case with Greece for the development of English-Greek in 1988 and Greek-French in 1993. Since then, an MT office has been set up in Athens to provide translation services to the Greek public sector, and in 2000 development was extended to Greek-English and French-Greek, with support from the Commission's MLIS (*Multilingual Information Society*) programme, which promotes collaboration on multilingual projects between public and private organisations.

At the beginning of 2001, MLIS agreements were also signed with the Portuguese, Dutch and Flemish authorities for the development of Portuguese-English/French and Dutch-English/French. The private-sector partner for all 3 projects is the SYSTRAN group. The new products acquired after these developments can be used by the Commission and the authorities involved with no limitation in time. Cooperation also covers enhancement of the existing language pairs involving Greek, Dutch and Portuguese.

With respect to the Nordic languages, MLIS is also co-financing Finnish-English / English-Finnish developments using the translation and parsing technology of Kielikone and Conexor respectively. Several proposals have been made concerning collaboration on Danish, but so far nothing has come to fruition. As for Swedish, the government has commissioned a study of the MT industry with a view to considering cooperation. The results of the study (in Swedish) can be found at: http://europa.eu.int/comm/translation/en/eyl/nutek.pdf.

Finally, looking ahead to the enlargement of the EU, projects have started for English-Hungarian/Polish and Polish/Hungarian-French under the successor to MLIS, the HLT (or *Human Language Technologies*) programme. The work involves Polish, Hungarian and French universities together with the SYSTRAN group.

3. The MT Dictionaries

The performance and reliability of the Commission's MT dictionaries are now widely recognised. They have been built up over the past 25 years mainly on the basis of real texts and real translation problems encountered in the Commission's documents. Moreover, in 1994 MT dictionaries were enriched with the terms contained in EURODICAUTOM, the Commission's terminology database.

Initially the dictionaries were bilingual (English-French, English-German...). However, as new target languages were added to the same source language, the need for modularity grew. The mono-source/multi-target approach is now the norm (English-French/German/Spanish...) and has done much to improve the performance of the various language pairs.

This section presents an overview of the evolving methodology used for dictionary work as well as a fairly detailed description of the principal features of the various dictionary files.

3.1 Methodology

As with many new technologies, there were few ground rules on which to base MT dictionary coding in the beginning. The initial task in 1976 was to adapt the system for the translation of Food Science and Technology Abstracts in the context of improving access to documentary databases.

One of Peter Toma's representatives spent two months in Luxembourg in 1976 explaining basic dictionary coding techniques. Among the most useful pieces of advice she gave was the suggestion that the dictionary should be based primarily on the frequency of occurrence of words and phrases in context.

We therefore proceeded to make wide use of frequency listings combined with the KWIC (*Key Word in Context*) approach.

The result was several thousand pages of printout which lined all four walls of the library at the Data Centre. From about 3,000 pages of running text, the KWIC concordance displayed each word in alphabetical order together with the immediate context in which it occurred. It was thus possible, for example, to see how often the word *plant* had been used to denote a factory or installation and how often it came up in its biological sense.

The idea was that the most common meaning should be coded as the dictionary default entry while the exceptions should be handled by some special mechanism.

As time went by, we began to deal with an ever wider variety of subject areas. While the frequency principle was still important, it was applied increasingly to the use of words in written texts as a whole rather than in just one document type or subject area.

Exceptions were covered either by subject-field codes (so-called *Topical Glossaries*) or, preferably, by contextual rules or string expressions.

> *Example*: the default translation of French *coeur* is *heart* in English. However, if a user indicates that his/her text deals with the subject field *energy*, *core* will be offered instead. There is just one snag to using *Topical Glossaries*: having said that your document concerns energy, the machine will *always* translate *coeur* as *core*, even if the more basic meaning also appears in the same text. A more refined answer to this problem is for developers to write a contextual rule – stating, say, that *core* should only be used if words with a semantic affinity (such as *reactor* or *fuel rod*) appear in the same sentence.

Potential conflicts in dictionary work could often be detected by using KWIC indexes, not just of texts, but also of the actual dictionary files for a given language pair. These displayed all words in alphabetical order, irrespective of their position in an entry, thus enabling the coder to avoid introducing a new rule which would override a valid existing rule.

Many of the improvements in the dictionaries have of course stemmed from years of translation tests conducted on different types of document. In the early days, the priority seemed to be to build the dictionaries up quickly, sometimes to the detriment of the target-language equivalents. Recently, more careful choice of the most suitable meaning based on general or technical usage has led to considerable improvements in quality.

Furthermore, increased efforts have been made in recent years to eliminate as much *noise* as possible from the dictionaries. By noise, is meant information which may have been introduced subjectively or intuitively by an individual coder without due reference to representative corpora. This often led to the inclusion of incorrect syntactic or semantic information and could only be eliminated by careful checking of the validity of each code against examples from running text.

The corpus-based approach has now become the norm and will certainly continue for some time as the most reliable basis on which to select lexical information and target-language equivalents. However, the importation *en masse* of specific terminology (as was the case with EURODICAUTOM) is not excluded.

3.2 Guidelines for Dictionary Work

The methodology can be summarised in the form of a few guidelines.

1) Coding should be based as far as possible on a representative set of occurrences of each word or term as shown by frequency information from pertinent text corpora.

2) In view of the numerous subject fields and document types covered by the Commission and the other EU institutions, the default meaning of any term should be carefully chosen so as to be applicable to as many contexts as possible.

3) While subject-field coding (Topical Glossaries) is an available option and can be used for certain predefined environments, the most reliable method of catering for exceptions to the rule is by means of string expression coding or rule-based contextual entries.

4) The subjective introduction of syntactic or semantic codes can be reduced to a minimum by carefully checking the validity of each code in the light of large, representative corpora.

5) Future enlargement or improvement of dictionaries could well be based on parallel corpora, that is, the alignment of a source text with its (human-produced) translations and the semi-automatic exploitation of the terms contained therein.

These rules can only serve as general guidelines in cases where a sufficient number of pertinent examples can in fact be identified from text-corpus analysis. In practice, perhaps half the words and terms to be coded fit into this category. The others, in the absence of supporting examples, must initially be left to the expertise of the linguists and translators responsible.

3.3 The Dictionary Structure

In EC SYSTRAN jargon, the dictionaries are divided up into STEM and IDLS (*Idiom/Limited Semantics*) dictionaries - STEM for individual words, and IDLS for expressions and contextual entries.

In the actual translation process, there is constant interaction between these two dictionary types. Indeed, every entry in the IDLS dictionary is cross-referenced by means of a digital address to the corresponding principal word (or headword) in the STEM dictionary.

While most initial dictionary coding work is handled on the basis of a given source and target combination, the source files are in fact physically separate from any particular target. From the point of view of screen display or printout, it is possible to view the source file alone, the source with any target language or the source with all target languages.

The dictionaries are not reversible, mainly owing to the fact that the source language dictionaries contain far more information than the target ones. It has nonetheless proved possible to use the target dictionaries as a bridge for the rapid development of new bilingual dictionaries. For example, the French-Italian version was based on a merging of the French-English and English-Italian dictionaries and has proved a considerable success.

In general, the dictionaries are updated every month, but can be refreshed every 24 hours if necessary. There is a comprehensive update of both dictionaries *and* programs four times a year. The procedure is as follows:

- bilingual updates for each language pair individually;
- merging of the STEM and IDLS dictionaries on a bilingual basis;
- incorporation of all bilingual updates into a mono-source/multi-target structure;
- final release of new multi-target dictionaries after thorough checking and testing.

While this procedure is fairly demanding on computer capacity, it has proved reliable as errors can be detected and corrected by all those working from a given source language. In practice, a dictionary coordinator ensures consistency and is able to make decisions if human conflicts occur.

3.4 The STEM Dictionary

The designation of the one-word dictionary file as the *STEM* is somewhat misplaced as in reality the English source language file contains not stems (morphological roots) but full forms. The term originated with the Russian-English system where roots were in fact used and is more specifically applicable to other source languages such as French, German and Spanish where the morphological approach is also used even if, for ease of reading, full forms are in fact printed out.

The source-language side of the STEM dictionary contains the following types of information:

- the headword (e.g. *work*);
- basic grammatical information (common verb, 1st & 2nd person singular and 1st, 2nd & 3rd person plural, present tense);
- homograph code (homograph with infinitive & noun);

- semantic-syntactic codes (concrete object, concrete subject, human subject, past participle non-modifier when following noun, usually intransitive).

In some cases, semantic primitives can also be added (in the case of *work* as a noun, the code *process* is used). There is also provision for including information on prepositional and, for English source, adverbial government.

The above example is typical of the type of information to be found in the English source dictionary. Syntactic and semantic codes have been refined over the years in an attempt to improve the performance of the linguistic routines. In particular, syntactic codes of government (e.g. *X can govern object plus present participle*) as well as homograph preference codes now play an important part in disambiguation.

The semantic primitives, which are used first and foremost with nouns, have also proved very useful. In the Commission's version of SYSTRAN, the number of semantic codes has been reduced to about 40 with the result that they can be successfully applied in practice. Typical examples of these are "month", "device", "cities", "profession" and "property".

The use of dictionary codes, particularly the syntactic ones, depends to some extent on the source language to be analysed. Full lists with clear criteria for applicability are to be found in the various dictionary coding manuals.

Here again, it should be stressed that while the actual codes and code types may appear similar to those used in other machine translation systems; the difference is that they have been used for many years in practical translation runs. The degree of reliability is thus very high.

3.5 The IDLS (Expressions) Dictionary

3.5.1 Idioms

The first and simplest type of entry to be found in the IDLS, or *Idiom/Limited Semantics* dictionary, is the idiom.

In the Commission's MT system, the term *idiom* usually designates prepositional, conjunctive or adverbial phrases such as *with respect to*, *in order that* or *over the medium term*. The preferred way of handling these is to introduce them as so-called *idiom replaces* which has the effect of reducing multi-word strings to a one-word entry. Thus *with respect to* becomes *with.respect.to*.

The inclusion of such idioms in the dictionary has various advantages. First, the string is never misanalysed as having some other syntactic function; second, the analysis is facilitated by reducing several words to just one; and third, the target language meaning is easier to pinpoint.

3.5.2 SLS Expressions

The next level of expression coding is represented by the *Straight Limited Semantics* (SLS) feature which provides for the coding of noun phrases (often technical terms) with or without a translation.

On the one hand, an SLS entry assists the analysis in that the grammatical homograph possibilities are reduced to nouns or adjectives. For example, if *vegetable oil* is included in the SLS dictionary, the word *oil* will no longer run the risk of being resolved as a verb even if no target meaning is appended. For languages requiring more than a word-for-word

translation, a target meaning such as the French *huile végétale* can of course be added. An SLS differs from an idiom replace in that it is not necessarily an *invariant* string – *vegetable oil* could also appear in the plural, for example – and thus cannot be treated as a single word.

3.5.3 CLS Rules

One of the most powerful features of this MT system is the *Conditional Limited Semantics* (CLS) rule as it allows for target meanings to be introduced on the basis of predefined contexts.

Rules may be as simple or as complex as required, the aim being to cover exceptions to the default meaning provided by the STEM dictionary.

Thus, if the default meaning of the verb *work* is *fonctionner*, a rule could be written to obtain the translation *travailler* when the subject of *work* is human. (And whether the subject is human or not is precisely the kind of data that should be stored in the STEM dictionary.)

Several conditions can be used together. For example, to cater for one of the various possible translations of *content* in order to obtain the meaning *table des matières*, the following conditions could be established:

- the part-of-speech value must be a noun;
- the number must be plural;
- the word must begin with a capital letter.

Conditions of this type play a vital role in obtaining the right meaning in context. Some of the more ambiguous terms may be covered by dozens or even hundreds of CLS entries.

Their usefulness can be seen from the proportion of CLS entries accessed in relation to other types of IDLS expression in actual translations. It usually runs at more than two to one, particularly for the more mature language pairs.

3.5.4 Homograph and Parsing Limited Semantics (HLS and PLS)

The system's linguistic routines will normally provide correct results when parsing the source language texts. They are, however, usually based on linguistic or paralinguistic phenomena rather than on actual words (except for the specific cases covered by lexical routines – see 4.3).

At times, though, it is necessary to alter the parse for a specific lexical context. For example, the program might opt for a verb rather than a noun after the sequence *go to xxx* if *xxx* is a noun-verb homograph. The program would therefore analyse *go to* ***help*** correctly (help = verb) but would fail on *go to* ***school***. It is possible to cover this exception in the form of an HLS rule which will ensure that *school* is treated as a noun in this particular context.

Similarly, exceptions to parsing rules can be handled by PLS entries. The normal program might fail to establish the correct dependency between *put* and *off* in *he put the light off the next time he came home*. If this type of context occurred sufficiently frequently, it might be well worth while to set the correct dependencies by linking *put* and *off* rather than *off* and *time*. This could be done by means of a PLS entry.

One of the additional advantages of HLS and PLS entries is that they usually apply to the source language irrespective of the target and thus lead to improvements in all language pairs based on a given source language.

3.6 Target-Language Coding

The type of information contained at the target level is far less complex than for the source language. It usually consists simply of a translation but will, in some cases, contain a *Special Meaning Code* (see 3.7 below) which can provide for specific requirements for the syntax of the target language.

Initially, target-language coding required the inclusion of digital morphological codes which would invoke the correct table of endings for verbs, adjectives and nouns. This feature has now been fully automated to the extent that the system will recognise which particular endings are required for any particular word. The approach is exhaustive, enabling the coder simply to add the full form as a trigger to the appropriate endings.

As regards choice of target meaning for the STEM dictionary, the most generally acceptable default is usually selected. For example, the best French target entry for *station* might well be *poste* rather than *gare* as it would probably apply to a wider variety of usages.

By contrast, in selecting meanings for SLS or CLS entries, the specific equivalent in context is required. Here, *railway station* could be given the meaning *gare* as an SLS and various CLS entries could be added to ensure that when *station* occurred in sentences with *trains*, etc., the meaning *gare* also appeared.

3.7 Special Meaning Codes

The function of the special meaning code is to provide additional information required for the relationships between words in the target language. For example, in French most adjectives appear after the noun that they qualify. The target routines will rearrange the word order accordingly unless the special meaning code "rearrange before" is used, in which case the adjective will be placed *before* the noun.

Other special meaning codes deal with phenomena such as government, auxiliary verbs or noun phrase composition. For example, "governs *de* plus infinitive", "use *être* rather than *avoir*", and "governs *en* in relationships with other nouns as in *teneur en fer*".

One special meaning code which has proved particularly useful in relation to SLS and CLS entries is "keep meaning" which ensures that the meaning selected will continue to be used as the translation of a given term throughout the text or until such time as a contradictory rule is executed.

For example, if *pétrole* is used to translate *oil* when it occurs in the phrase *mineral oil*, the "keep meaning" code could be used to ensure that other instances of *oil* in the same document (but without *mineral*) are also translated as *pétrole* rather than *huile*.

3.8 Topical Glossaries (TGs)

Like most other machine translation systems, the Commission's MT system has a feature for dictionary coding on the basis of subject fields or document types.

While this feature has proved particularly useful for some of our external users working, for example, in the nuclear or aerospace fields, it has been less useful for in-house users owing to the wide variety of fields covered, not to mention the switches between fields within the same document.

There have, of course, been exceptions. Use of a parameter for minutes of meetings has not only led to more appropriate terminology but has also served to trigger tense changes

between English and the other languages. (English minutes are written in the past tense, but the French, Spanish, and Italian convention is to use the present tense.) Benefits can also be obtained from topical glossary coding when a specific user group is targeted.

The danger, on the other hand, is that users will expect far better results when their own particular subject-field codes are used. Unfortunately, in view of the nature of many of the texts translated this is not always the case. Indeed, sometimes clear degradations occur as the user may well introduce inappropriate parameters and obtain equally inappropriate results.

The most reliable strategy is certainly to attempt to cover as many occurrences as possible on the basis of a good STEM default translation together with powerful SLS and CLS rules for selecting other meanings when the context requires. This approach will also have the effect of providing benefits to all users rather than just to those working on the basis of a given subject-field parameter.

In a similar spirit to topical glossaries, we are now also in a position to offer *user* glossaries, the contents of which are focused more on the needs of a particular user rather than on a specific subject field as such, although in practice, the two glossary types overlap.

3.9 EURODICAUTOM

In the mid-1990s, the MT dictionaries were reinforced by EURODICAUTOM, the Commission's terminology database. If the machine fails to find a word or expression in its own dictionaries, it will now check EURODICAUTOM data for the terms concerned. This development has quadrupled the number of translations which are in principle available to the system.

Tests carried out to measure the amount of progress obtained proved that higher translation quality is achieved in technical texts, whereas in general ones EC SYSTRAN provides better translation. The reason is that in the EC SYSTRAN dictionaries the most general meaning of a word is coded as the default entry, while EURODICAUTOM is highly specialised. Consequently, access is controlled: in order to trigger EURODICAUTOM, users *must* indicate a text domain when making a request. Then, if a EURODICAUTOM term is matched in that domain and there is no entry in the MT dictionaries, it will be taken.

3.10 CELEX

MT is also connected to the Commission's legal database, CELEX: references to EU legislation in source texts are extracted and looked up in the CELEX legislative database. The full title of the relevant piece of legislation is then retrieved for both source and target languages and placed at the end of the MT output.

3.11 Conclusions

The Commission's MT dictionaries have a number of features and qualities which go far beyond their counterparts in other systems.

In regard to the STEM dictionary, the development of syntactic and semantic codes over the years has now led to a set of items which can be reliably applied. In the case of syntactic codes, which do much to assist in the analysis of source text, a wide variety of patterns has been covered and comprehensively tested on the basis of text corpora from various sources.

With the semantic codes, too, we have found that different coders soon learn to apply them consistently in view of their clear definitions. While the code set is comparatively small, each code has been introduced for a specific purpose with the result that the quality of dictionary entry should in general be superior to that of systems which offer a large and often baffling number of codes, seldom correctly applied in practice.

Another special feature of the Commission's MT system is the CLS rule. While some systems do make provision for a limited number of grammatical dependency relationships (subject/verb, preposition/object), the Commission system's ability to cater for any combination of contextual rules is second to none. Indeed, it is this feature which is largely responsible for the correct choice of meaning in context.

Lastly, the homograph and parsing rules (HLS and PLS) which can now be coded provide a simple, straightforward way of modifying the analysis on a lexical basis without upsetting the program.

4. The Translation Process

4.1 General Approach

At the present stage of development, the Commission's MT system can be described as a modular, transfer-type system in that the source and target language components are almost completely separate. The actual *translation* is thus performed by linking the two by means of the transfer stage which, in particular, provides the target-language equivalents by drawing on the bilingual dictionary files specific to the language pair in question.

Like many other machine translation systems, it can thus be divided into source-language **analysis**, bilingual **transfer** and target-language **synthesis**. *Figure 3* below offers a simplified view of the process – more details are provided in the sections that follow. As you will see, MT is more than just a question of word replacement!

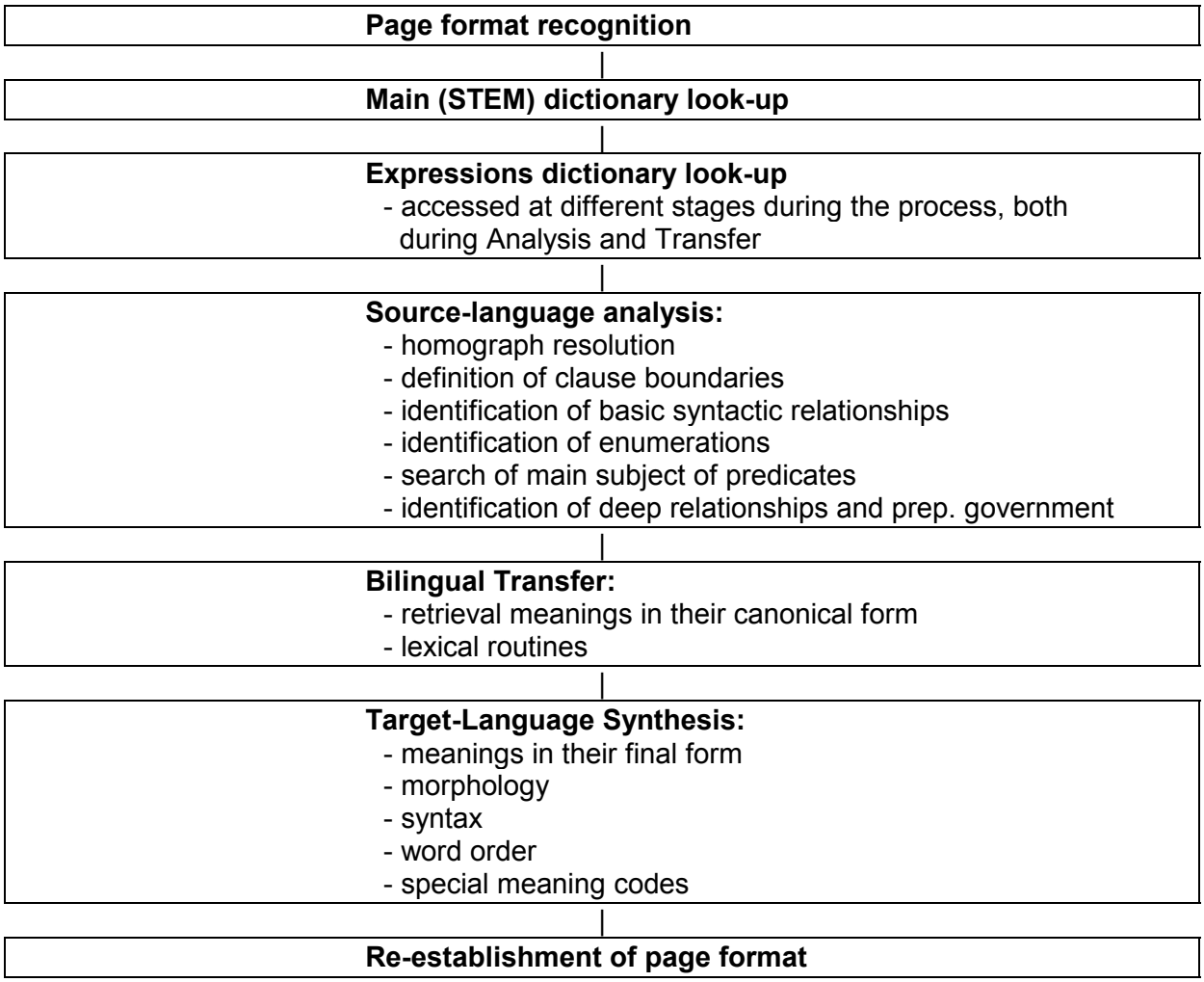

Figure 3 – Overview of EC SYSTRAN Translation Process

4.2 Source-Language Analysis

General users of MT can make a request in 2 ways: by means of their standard e-mail application or a user-friendly Web interface. Translators have a more advanced interface known as *EURAMIS*, which offers additional translation tools (see 5.1 below).

Once the document to be translated has been sent to EC SYSTRAN, it is submitted to pre-processing routines which separate formatting information from the actual text and provide a sound basis for establishing translation units, usually sentences, in the MT input format.

The actual linguistic analysis can then begin.

From the information appended to each word in the STEM dictionary and that created for any unrecognised words on the basis of their endings (French *–ité* and Spanish *–ción* suggest feminine nouns, for example), the six major parsing steps can be executed sequentially.

These are:

- homograph analysis;
- clause boundary definition;
- establishment of basic syntactic relationships;
- creating links between items in enumeration;
- finding the main subject and main predicate (verb) of each clause;
- establishment of deep syntactic relationships.

4.2.1 Homograph Analysis

The function of this routine is to try to establish the exact part-of-speech value of each word in the sentence. All of the five source languages we have covered to date present grammatical homograph problems but by far the most difficult to handle is English, followed by French and German.

The pass begins from left to right. For words where no part-of-speech ambiguity exists, the dictionary information will not only give the pertinent part of speech but will provide other syntactic and semantic clues which will help resolve those which are homographs.

In the case of the homographs themselves, the various dictionary entries (e.g. the three entries on *light* for noun, verb and adjective) will be examined together with the syntactic and semantic codes attached to each one with a view to establishing the most likely solution.

The basic principle is that of working from left to right, establishing one by one the correct part of speech of each homograph. The more correct *hits*, the easier it is to continue successfully down the sentence. But this can be quite a complex matter, particularly when a sentence contains a large proportion of homographs, as is often the case in English and French. Moreover, if a serious mistake is made near the beginning of a sentence, many of the subsequent homographs will be incorrectly analysed, as the information required for establishing their value is itself incorrect.

For many years, we tried simply to improve the various routines, adding more and more tests. However, with over 80 homograph types for English alone this soon developed into a

mammoth task with some routines, particularly those for the more frequently encountered problems (verb/noun, participle/finite verb), extending to several thousand lines of program.

Not only did the program become difficult to manage, but the degree of progress became slower and slower from year to year.

After much trial and error, the decision was made to handle homograph resolution in two stages. First, all the reasonably straightforward forms would be handled; then, on the basis of the far richer and more reliable information for the majority of words in the sentence, a second pass could deal with the more difficult cases.

In addition, to support this approach, a number of new dictionary codes were introduced to take account of more pragmatic information found in corpus work such as the frequency of occurrence of different parts of speech in context. Thus, even in cases of apparent ambiguity, the correct value could normally be established correctly.

Finally, the HLS dictionary feature described in the previous chapter was introduced to facilitate resolution of real exceptions to otherwise reliable syntactic rules.

While the results of homograph resolution are now surprisingly good for all the languages covered, if a mistake does occur at this early stage in the parsing process, it is still likely to lead to serious mistranslations.

It is for this reason that work on improving homograph resolution continues to receive high priority.

4.2.2 Clause Boundary Definition

Although information established from preceding sentences can to some extent be used in the parsing process, MT continues to be based on a sentence-by-sentence translation.

Actual sentence definition is handled partly by the peripheral, pre-processing programs and partly by the system's own "get sentence" routine prior to homograph establishment.

As the remainder of the parsing process will deal primarily with established grammatical dependencies within each clause of the sentence, it is important to start by establishing clause boundaries.

The system differentiates between main clauses (of which there may be more than one in a sentence) and subordinate clauses.

While in many cases clause boundary definition can be handled without major difficulty, problems can occur in the following cases:

- One clause is embedded within another. (***An embedded clause may*** *if it is not correctly analysed* ***break the flow of the sentence****.)*

- No obvious clause opener is used, as is often the case with the so-called missing *that* in English. (***He maintained*** *clause boundaries were difficult to establish*.)

- The *sentence* is not complete or is incorrectly punctuated. This often occurs in administrative and legal documents when an introductory passage leads on to a series of indents.

Progress on all of the above continues to be achieved but, particularly for translation into the Germanic languages, errors in clause boundary resolution can still have disastrous effects on the structure of the translation.

4.2.3 Establishment of Basic Syntactic Relationships

On the basis of the dictionary information available for each word in a clause, it is now possible to establish relationships or dependencies between key syntactic items. These include:

- preposition and its immediate object;
- verb and its direct object;
- verb and adverb;
- noun and its qualifier (e.g. article or adjective).

Pointers are normally established in both directions, for example, information stored on the preposition will point to the word number (counted according to its place in the sentence) of its object (noun or pronoun) and the information on the object will now also include a pointer back to the preposition.

4.2.4 Enumeration

The importance of establishing enumeration in machine translation is often underestimated. Indeed, one of the most difficult tasks to handle in dealing with technical terminology or administrative jargon is how, whether, or when the various words are enumerated with each other.

Depending on how enumeration is established, the basic syntactic relationships will be correctly or incorrectly extended.

One of the main problems is to decide on the extent of coordination. For example, the dependencies in *old men and women* differ from those in *old men and dogs*. Here, it is the affinity between *men* and *women*, which is much stronger than that between *men* and *dogs*, which tells us which nouns *old* qualifies. Thus, the fact that *men* and *dogs* may both be the subject or object of the same verb (requiring one level of enumeration) does not necessarily mean that this affinity extends to adjectival qualification.

In our MT system, the decision can be made partly on the basis of the syntactic and semantic information stored in the dictionary and partly on the contents of the expressions dictionary.

In this connection, it is interesting to note that while the semantic primitives are used above all to cover a variety of dependencies in CLS coding, they have also proved useful in helping to establish enumeration.

Potential enumerations are normally indicated by co-ordinate conjunctions (*and, or, as.well.as*) or by use of the comma. They may occur not only between parts of speech of the same type (mainly nouns, verbs, adjectives and adverbs) but also between phrases or even clauses.

The relationships are again marked by word number pointers in both directions, thus further enriching the amount of information available on each word.

4.2.5 Subject and Predicate Establishment

The next stage in the analysis process is concerned with the identification of the main subject and main predicate of each clause.

This is very useful in establishing more general relationships across the sentence. Indeed, by including the word number of the subject and of the predicate in all the other words, it is possible to test various relationships in the transfer and synthesis routines.

It also provides an easy basis for writing CLS rules dependent on a particular type of subject or verb.

4.2.6 Deep Structure

The final step in the analysis sequence is to mark up the deep structure of the sentence.

The deep structure provides a means of drawing relationships between a verb and its semantic subject or object irrespective of mood. For example, the deep structure of *The ministry collected sales tax* and *Sales tax was collected by the ministry* will be the same although the surface structure is different.

Logic concerned with the subject or object of a verb will thus apply to either case, irrespective of whether the active or passive verb forms are used, provided the meaning remains the same. Participial structures will also be taken into account as in *The taxes collected...* or *The ministry collecting taxes....*

The results of this deep analysis are particularly useful in expression coding and in providing for tense and mood transformations between source and target.

The result of all this analysis work is a so-called byte area appended to each word in the text which contains a wide variety of information including:

- all the static information retrieved from the dictionary;
- the part-of-speech value decided by homograph resolution;
- a marker on each word identifying the clause to which it belongs;
- information on the article government of nouns;
- pointers between words which have a direct syntactic relationship with each other;
- pointers between elements of enumeration;
- pointers from all words to the subject and predicate;
- pointers establishing the deep structural relationships between subjects, verbs and objects irrespective of mood.

All this is contained in up to 180 bytes or boxes attached to each word. In practice, the average word may have from 20 to 50 useful pieces of information which can later be used at various stages in the transfer and synthesis programs as well as by CLS, HLS and PLS entries.

It should be stressed that Analysis works strictly on the source-language side, obtaining as much information as possible (syntactic, semantic) on the words in the text in order to

establish relationships between them. It is not aimed at providing a translation - *that* is handled by the Transfer and Synthesis stages.

4.3 Bilingual Transfer

Here EC SYSTRAN begins looking at the actual translation of the text. Target-language equivalents for the source words/expressions identified previously are now introduced, however they will not appear in their proper form (correct inflection, arrangement, agreement, etc.) until the next step in the procedure (Synthesis). For example, if you requested a translation into French of *I want a green house*, at this stage in the process the translation would look like this: *je vouloir vert maison*.

Apart from the retrieval of meanings from the various dictionary files, the transfer module consists largely of *lexical routines*.

These deal with translation problems which normally require changes of structure or the implementation of complex rules between the source and target languages.

Routines based on the semantic codes for cities and countries will, for example, ensure that the correct prepositions and articles are used in the target language while the routine for months will deal with date structures.

Lexical routines may also be written on or around a given word if major changes are required between source and target. For example, *expect* in English is often used passively (*he is expected to come*), whereas an impersonal active translation may be required in certain contexts (*On s'attend à ce qu'il vienne*...).

Nowadays, with a more powerful IDLS dictionary, it is often possible to cover the functions of lexical routines in the dictionary itself. Nevertheless, lexical routines continue to provide a useful means of grouping together most of the criteria for distinguishing between various translations for words which are particularly ambiguous or present complex structural problems.

4.4 Target-Language Synthesis

By the end of transfer, the source text will (hopefully!) have been correctly parsed and the applicable meanings will have been identified.

Three important operations still have to be carried out:

- inflections or morphological appurtenances need to be added to many of the target translations (particularly verbs, nouns and adjectives);

- articles, prepositions, infinitive particles and other target-related function words must be inserted according to the syntactic rules of the target language;

- word order must be adapted by means of rearrangement routines.

Expressed in this way, the process seems quite simple. It can, however, be a very difficult matter to obtain the correct article in French when no article at all is used in English. Even more difficult is the establishment of correct word order in German when translating from French or English, where the structure is very different.

The final stage of the translation process is re-establishment of the page format of the original text.

This is handled by post-processing routines which mirror the functions of pre-processing. They re-insert word-processing, formatting and pagination data which ideally allow for the text to be returned (by e-mail) to the requester's PC with all the display information of the original.

4.5 Conclusion

The fact that the sequential approach to translation processing in the Commission's MT system has survived without major changes over the years is evidence that the method is well founded.

Indeed, for the development staff, it appears much easier to deal with small component parts, each providing a result for the next stage, than with the rule-based approach which can produce more than one solution (sometimes dozens) without necessarily finding the correct one. Complexity here becomes a serious problem, not just in terms of computer capacity, but in terms of the human mind.

Moreover, the sequential approach is open-ended; in other words, new logic can be added without upsetting what already exists. Over the years, new parameters, programming macros and dictionary codes have been added, sophisticating the degree of treatment at each stage. The analysis area itself has been expanded to store additional information on each word and new features such as backtracking have been introduced.

Much more, of course, remains to be done. There are still a number of error types which require deeper analysis and, on the target side too, further enhancement of routines dealing with article assignment or word order is called for.

The challenge will be to increase performance while maintaining transparency, to increase translation quality without increasing the system's complexity.

5. Machine Translation Use

5.1 Gaining Access

EC SYSTRAN is available via the Web (see *Figure 4*) or electronic mail to all Commission officials who are equipped with a PC connected to the network. The system is also accessible to other EU institutions and public authorities in the Member States.

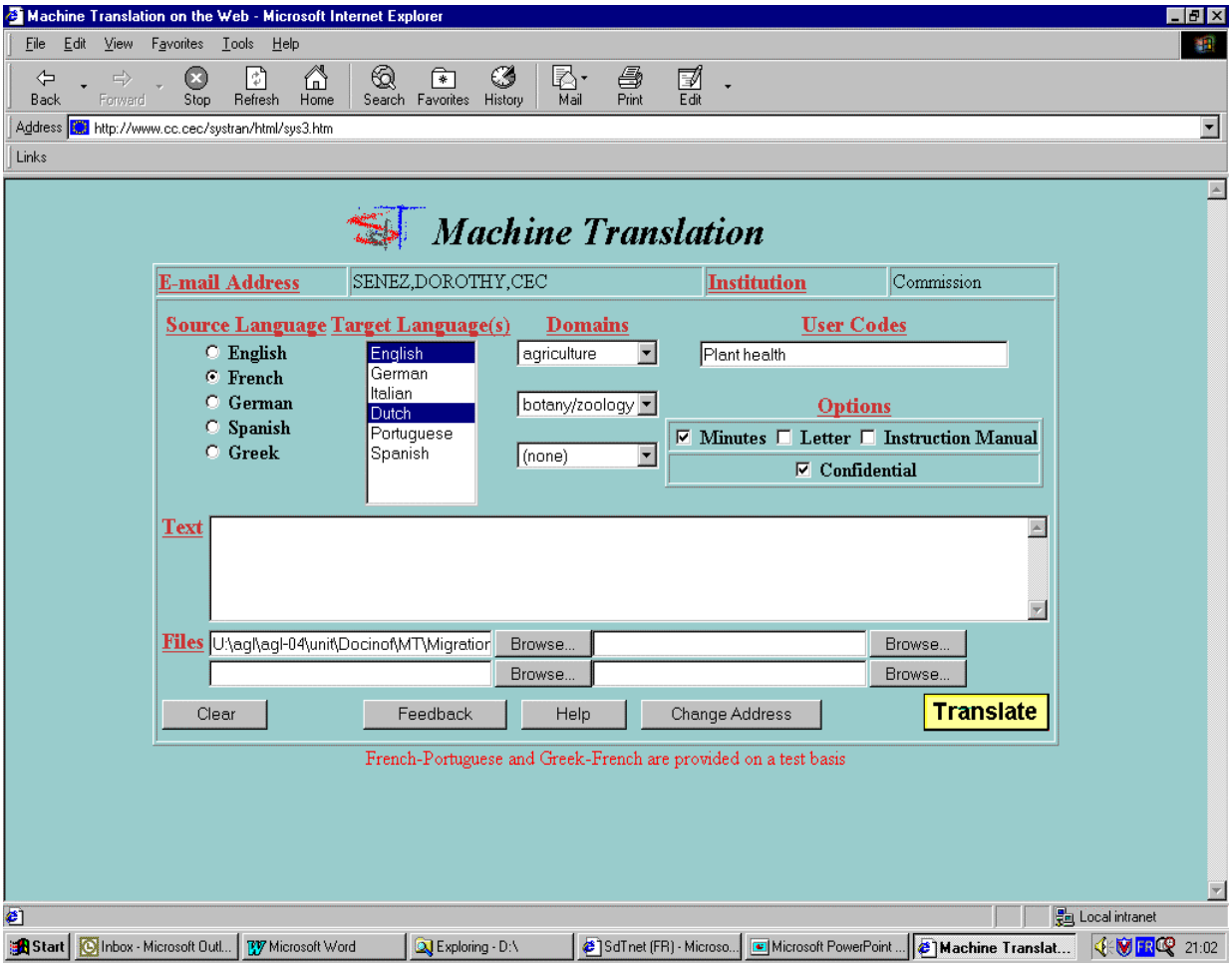

Figure 4 –Web Interface

MT is easy to use. Users simply send their documents to a special mailbox stating their requirements in language combinations and subject fields (*Domains*). Normally, the translation is returned by e-mail within half an hour, depending on how heavy the traffic is. Sometimes end-users receive their translation in only a few minutes. Various file formats are supported, but official policy is now to encourage use of RTF as it is too costly to upgrade the converter every time a new word-processing package is released.

Commission translators can also access MT through a more advanced interface known as *EURAMIS* (European Advanced Multilingual Information System - see *Figure 5*), which offers a wider range of facilities.

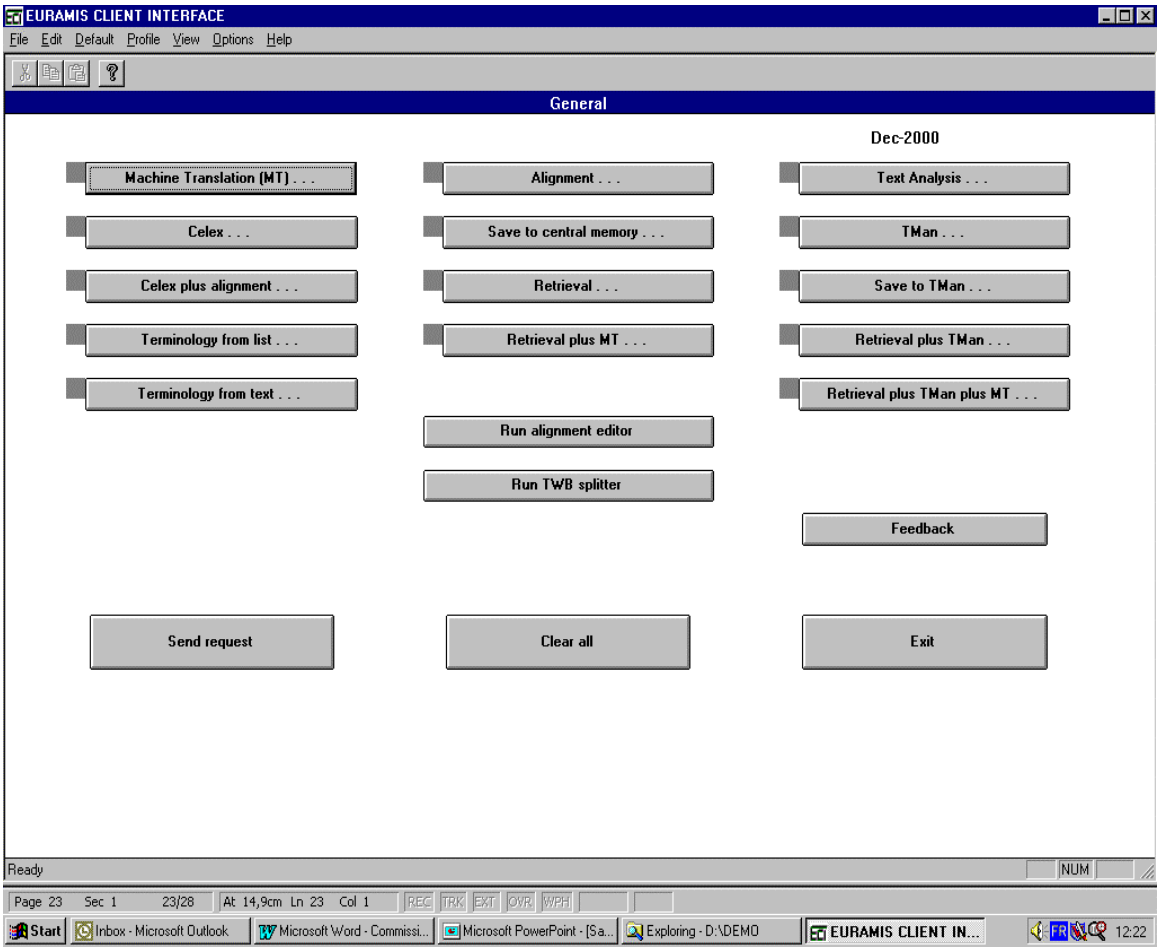

Figure 5 –EURAMIS Interface

EURAMIS provides access to the following:

1) Machine translation (EC SYSTRAN).

2) Extraction of legal references for matching with their target-language counterparts in the CELEX legislative database. This service is already provided by default with machine translation (see 3.10); the difference here is that translators can request an extraction of the legal references alone.

3) Terminology from list: automatic look-up of a pre-established list of terms using the EURODICAUTOM terminology database. MT is not involved, so any combination of the official EU languages may be requested, including those not supported by EC SYSTRAN (for example, Swedish-English).

4) Terminology from text: automatic look-up by EURODICAUTOM (via MT) of the terms contained in a text. MT is used to extract terms from the text; these are then looked up by EURODICAUTOM. As the MT analysis programs are required here, the number of source languages is limited to the 5 supported by EC SYSTRAN. On the target side, however, any of the official languages may be requested since EURODICAUTOM covers them all (in other words, French-Swedish would be possible, but not Swedish-French).

5) Translation memory and text alignment tools. EURAMIS offers a central repository for aligned source documents and their (human) translations. If a source document is of a standard format - for example, an invitation to tender or contract - the secretariat may send it to EURAMIS for pre-processing. The result is a file which contains translations for any parts of the document which have been translated previously. Text for which there is *no* translation can be sent automatically to MT or left untouched for the translator. Once the translation is finished, it too can be aligned (sentence by sentence) with its source document and stored in the EURAMIS memory. The translation memory thus allows translators to benefit from the collected wisdom of their colleagues, saving time and ensuring greater terminological consistency in the process.

6) TMan, an in-house development which generally makes phrase substitutions below sentence level. This too can be combined with MT and/or memory.

5.2 MT Demand

5.2.1 Requests and Pages

MT statistics over the last 10 years reveal that the number of users is steadily increasing, not only in the Commission, but in other EU institutions and the Member States, thus paralleling the growing *public* use of MT technology on the Web.

The generalisation of PCs together with easier access, several awareness campaigns and an increasing workload, has caused demand to rise by more than 500% since 1993. A total of 97,199 requests were made in 2000, up more than 23% on the previous year (see *Figure 6*).

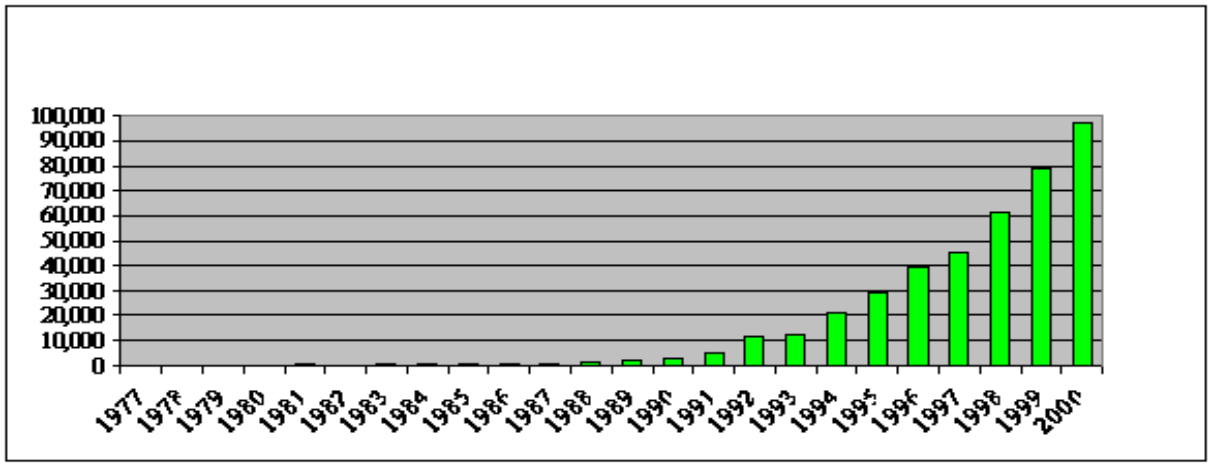

Figure 6 - No of MT Requests per Year, 1977 - 2000

Demand now ranges from 45,000 to 50,000 pages monthly, whereas at the beginning of the 1990s it stood at 2,000 pages per month. A total of 546,248 pages were machine-translated in 2000 (see *Figure 7*). The Commission accounted for 77% of this figure, almost half of which was in turn requested by its translators. The remaining 23% was shared evenly between other EU institutions and Member State authorities.

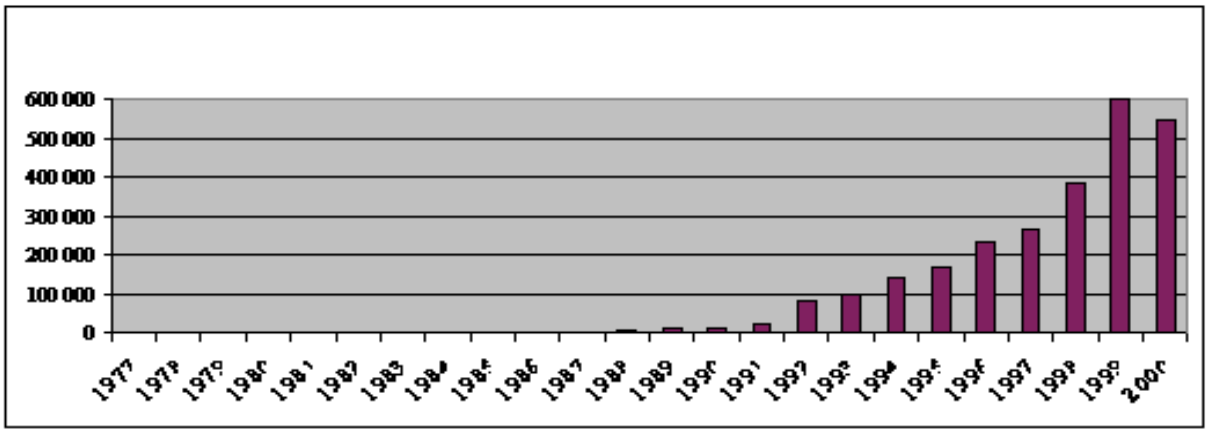

Figure 7 - No of MT Pages per Year, 1977 - 2000

There was a slight drop in pages relative to 1999, attributable mainly to a reduced page count in the Translation Service itself - in spite of a 12% increase in translator *requests*. Perhaps the generalisation and expansion of translation memories within the Service plays a role here, with memories being favoured for longer documents.

Tentative figures for the first half of 2001, however, suggest that both the SDT's requests *and* pages are again increasing, and that overall demand for the system could reach a new high.

5.2.2 Language Pairs

The language pairs most often requested (see *Figure 8*) are English-French and French-English, as these are the languages most used in the Commission, and they can provide an acceptable quality. In third and fourth place come French-Spanish and English-Spanish respectively. These language pairs are largely used within the Translation Service (as opposed to the administrative departments) and are best suited to translators' needs owing to the plethora of feedback contributed by Spanish translators.

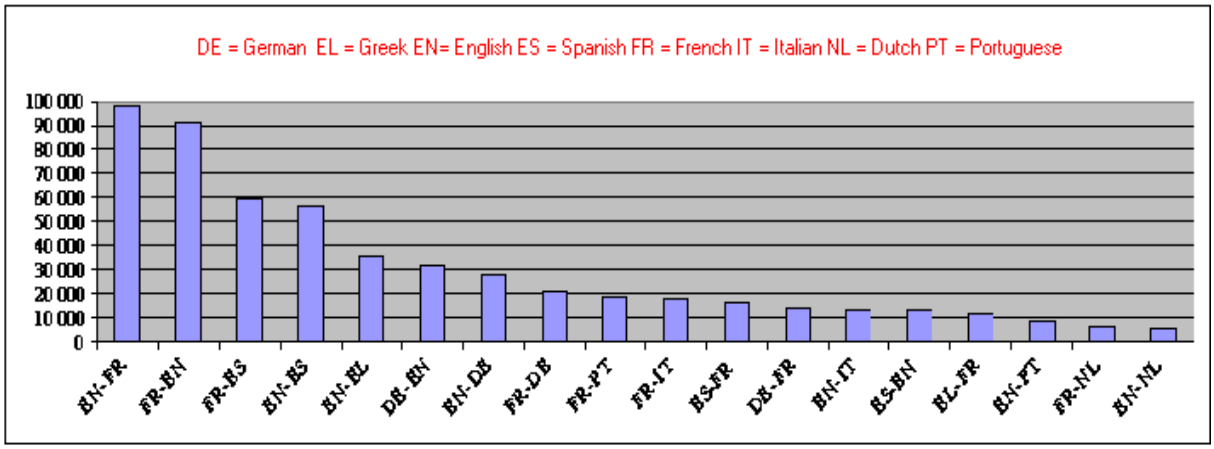

Figure 8 - No of Pages Requested by Language Pair in 2000

5.3 Why MT is Requested

Users can be divided into two groups: administrators and translators. Extensive surveys carried out by the MT Help Desk and management team have shown that Commission administrators request machine translation for three reasons:

1) For **browsing** texts written in a language they do not know. The quality of the translation may not be high, but the speed is remarkable: the computer can translate 2 000 pages per hour. Users can then decide if they wish to submit their texts (or part

of them) for human translation or whether the information provided in the raw translation is sufficient.

2) For the **fast translation** of urgently needed texts which often have a standardised structure and terminology (minutes of meetings, reports, etc.). A reasonable translation quality can be obtained after correction by someone who has the target language as his/her mother tongue. The texts can then be distributed for internal use: MT should not be used for legislation or documents intended for publication, and is usually less suitable for long texts (because of the degree of revision involved).

   For those who wish to avoid the task of correction, an editing service is on hand (see 5.4 below).

3) For **drafting** in a language other than their mother tongue or main language. Some officials prefer to write a text in their own language first, request a machine translation and then correct the output.

Translators, on the other hand, use MT almost exclusively as a basis for providing a more polished translation. In 2000, translators accounted for 44% of pages submitted to MT at the Commission, with 56% coming from the administrative departments.

Machine translation is well received amongst administrators – they dislike the amount of correcting involved, but find MT a useful stopgap. In a survey in 1996, 95% of administrator respondents stated that machine translation was a useful tool to have at their disposal.

Reaction in the SDT is more mixed. Spanish is the language which has been most adapted to translators' needs, so not surprisingly French-Spanish and English-Spanish are the most popular language pairs here (see *Figure 9*). Adaptations have been made on the basis of feedback from Spanish translators themselves, and in particular from one inveterate "MTer" whose own section sends virtually all French texts to EC SYSTRAN and then edits them. Other Spanish sections have taken an interest too, thus creating a virtuous cycle of feedback for developers. A similar trend is now developing for French-Portuguese, and English-French also has a reasonable demand.

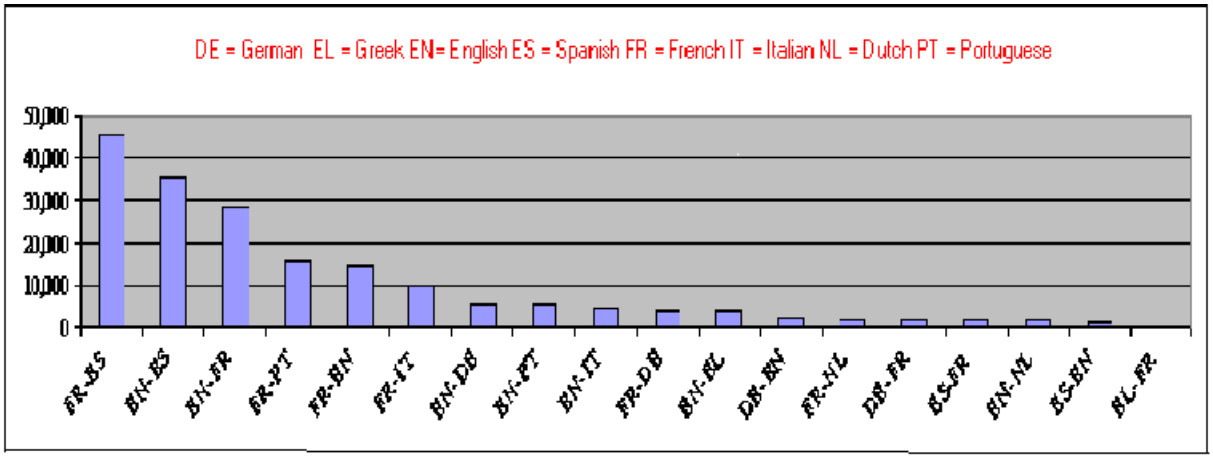

Figure 9 - No of Pages Requested by Language Pair in the Translation Service 2000

There are "cells" of MT users in other languages, but on the whole, requesters are more concentrated (i.e. less evenly distributed) than in the administrative departments. This is natural since, unlike in the rest of the Commission, the Translation Service's staff are "segregated" by mother tongue, and the quality of each language pair varies considerably (so some language sections will use MT more than others). For that matter, the Nordic languages are not supported at all. In general, a language pair's quality will depend on length of development and the syntactic and lexical affinity of the languages concerned. The quality

of a specific machine-translated text will also be determined by variables such as text type, subject, and clarity of the source document. The most satisfactory results are offered by pairs involving English, French, Spanish and Italian, while greater effort is still necessary to improve Dutch and German: as a result, Dutch-speaking translators are not served in the same way as their Spanish counterparts!

A more important distinction is that administrators and translators do not have the same requirements of MT: generally the former are just looking for the gist of a text, or want a quick translation of a short-lived document - a minutes text "for the record", a working paper which is going to be modified a dozen times before the final version is produced.... In contrast, translators are expected to produce a top-notch translation on every occasion: they can use MT if they like, but the final result should be seamless. This means that the standard of editing is high, and if the MT output is poor, correcting it might take as long as starting from scratch. An administrator's MT tolerance threshold is therefore greater than that of a translator, and what is useful for the one may not always be of help to the other.

That said, a series of practical experiments conducted with translators who regularly use MT have shown that that savings of up to a third could be achieved in translation time *in the right circumstances* (language pair, text type, domain, style). This proviso is very important, but even so, it is fair to say that MT can serve a purpose, and the Spanish experience suggests that more can be done.

5.4 Rapid Post-Editing Service

The Translation Service offers an external Rapid Post-Editing Service for Commission administrators with tight translation deadlines which (given its workload) the SDT might find difficult to respect. In such a case, officials can send their texts to the MT Help Desk in Brussels, which will in turn translate them with EC SYSTRAN and pass the results on to freelance translators for correction. Emphasis is on speed and accuracy rather than style or in-house jargon. As a result, this service is available for internal documents only; in the case of documents intended for external distribution, administrators must ask the SDT for a fully polished "human" product.

5.5 The MT Correspondents

In order to enhance the system and target development more precisely, the SDT has set up a network of *MT Correspondents*. The twenty or so Correspondents are mostly translators who use MT in their daily work as a basis for a final polished translation (internally, the act of revising MT is known as *post-editing*). They are expected to contribute to discussion on general strategy and to provide feedback to the outside contractor responsible for maintaining and upgrading the system.

There are Correspondents for all 11 official languages of the EU: if Swedish, Finnish and Danish colleagues cannot work on EC SYSTRAN, they can still play a useful information-gathering role (by keeping abreast of MT developments in their countries) and may be called upon to assess any new systems which come on the market.

6. Migration Project

For 25 years, EC SYSTRAN has operated with old *IBM Assembler*. In spite of the age of its computer language, the system has remained one of the most robust tools available: in times of limited resources yet increasing demand, it therefore seemed more useful to invest in the MT dictionaries and enhance the quality of output rather than rewrite the programs.

In 1997, however, the Commission's Data Centre in Luxembourg announced that the mainframe computer which supported Assembler would be phased out within 5 years. This left the SDT the choice of finding a modern emulator for the Amdahl or rewriting EC SYSTRAN's programs in a more contemporary computer language. One feasibility study later, it was decided to place EC SYSTRAN on a new, open systems platform (UNIX). This entailed converting, or *migrating*, the basic programs from *Assembler* to *C* and was justified as follows:

- the future lay in open, non-proprietary systems;
- it was becoming increasingly difficult to find technicians with expertise in IBM Assembler;
- the commercial version of SYSTRAN had already been migrated, and benefit could surely be gained from the experience there;
- it opened up the possibility of new facilities, such as translation on the fly of Web pages, and perhaps a personal coding tool to allow users to create their own dictionaries.

In late 1998, the SDT therefore launched a 2-year migration project. In spite of the knowledge gained from the commercial version, the work proved complex and time-consuming (EC SYSTRAN has its own special characteristics). Indeed, linguistic development was largely shelved in the second year of the project. At the time of writing, there are still a few months of running-in ahead before the new system can fully enter production mode, but thereafter users should begin to enjoy the benefits of a modern computer platform.

7. Other Projects

As indicated above, the Commission will be considering plans to allow MT users to enter their own terms in a private dictionary by means of a personal coding interface. The trick will be to ensure that those terms take precedence over the translation provided by the main MT dictionaries. The possibility of translating EC Web pages dynamically also features among future projects, as does the (semi-)automatic coding of new entries in MT dictionaries on the basis of corpora and glossaries.

Moreover, as part of the IDA programme (Interchange of Data between Administrations), the Commission will be conducting a feasibility study on potential MT needs in the Member States, with a view to reinforcing the system. The study consists of three parts:

1) a survey concerning the principal needs of European public administrations in the field of MT;
2) a definition of the infrastructure necessary: a) to coordinate access to the Commission's MT system and b) to carry out the technical, linguistic and terminological developments requested; questions include means of access (Web, batch, e-mail, multiple sites, etc.), confidentiality, and efficient integration of linguistic and terminological resources;

3) assessment of the financial and human resources needed to complete the work.

8. Conclusion

After twenty-five years of development, machine translation has now become a helpful option for some of the everyday translation needs in the Commission's administrative departments. It can also be used by translators as an effective support tool, although the picture varies according to target language.

Machine translation is not aimed at replacing human translators. It cannot compete with them, since the computer does not have the experience and the knowledge of the world that only humans can acquire. It is simply a complementary, surprisingly fast tool, that can rescue translators or administrators from some dull work. The majority of MT customers are not translators. They have urgent translation needs which the Translation Service cannot always satisfy by conventional means. And that's where EC SYSTRAN comes in.

9. References

Leick, J-M (1995) *Machine Translation and Multilingual Tools in the European Commission*, EC, Luxembourg, internal document.

Oakley, B. et al (1991), *Evaluation of the Commission's Multilingual Action Plan 1976-1991*, Luxembourg, internal document.

Paesmans, H. (1994), *The Translator's tools*, European Commission, Translation Service, internal document.

Petrits, A. (1994), *The current state of the Commission's SYSTRAN MT system*, in Proceedings of '**Machine Translation Ten Years On'**, Cranfield, UK.

Pigott, I. (1992), *SYSTRAN development at the EC Commission, 1976 to 1992,* European Commission, internal document.

Senez, D. (1994), *Developments in SYSTRAN,* ASLIB Proceedings.

Toma, P. (1986), *SYSTRAN's contribution to mankind*, in **Terminologie et Traduction** 1, ISSN 0256 7873.

ANNEX -
Sample Translations from English into French and German
Sample 1

ORIGINAL - English	EC SYSTRAN – German	EC SYSTRAN - French
Telematics Applications Programme.	Telematikanwendungsprogramm.	Programme d'applications télématiques.
Second Call for Proposals for the Specific Programme for Research and Technological Development, including Demonstration, in the field of Telematics Applications of Common Interest (1994-1998) (95/ C xxx/xx) (Text with EEA relevance)	Zweite Ausschreibung für das spezifische Programm für Forschung und technologische Entwicklung einschließlich der Demonstration im Bereich der Telematikanwendungen von gemeinsamem Interesse (1994-1998) (95/C xxx/xx) (Text von Bedeutung für den EWR)	Deuxième appel de propositions pour le programme spécifique pour la recherche et le développement technologique, y compris la démonstration, dans le champ des applications télématiques d'intérêt commun (1994-1998) (95/C xxx/xx) (texte présentant de l'intérêt pour l'EEE)
The new Telematics Applications Programme was launched on 23.11.1994 as a specific programme of the Fourth Framework Programme.	Das neue Telematikanwendungsprogramm wurde am 23.11.1994 als ein spezifisches Programm des vierten Rahmenprogramms gestartet.	Le nouveau programme d'applications télématiques a été lancé le 23.11.1994 comme programme spécifique du quatrième programme-cadre.
The Programme runs from 1994 until 1998, and will fund research and technological development and demonstrations of applications using information and communications technologies in four areas covering 13 different sectors of common interest.	Das Programm läuft von 1994 bis 1998 und finanzieren Forschung und technologische Entwicklung und Demonstrationen von Anwendungen Informationen verwendend und Kommunikationstechnologien auf vier Gebieten, die 13 verschiedene Sektoren von gemeinsamem Interesse erfassen.	Le programme fonctionne à partir de 1994 jusqu'en 1998, et financera la recherche et le développement technologique et des démonstrations des applications utilisant des technologies de l'information et des communications dans quatre secteurs couvrant 13 secteurs différents d'intérêt commun.
In addition, programme support actions addressing issues common to all the sectors will be funded.	Darüber hinaus werden Programmunterstützungsmaßnahmen, gemeinsame die Fragen an alle Sektoren richten, finanziert werden.	En outre, les actions de soutien de programme abordant des questions communes à tous les secteurs seront financées.
In the different sectors of the Programme, the efficiency and quality of the services traditionally provided to the public could potentially be improved through the introduction or development of telematics-based systems and services.	Auf den verschiedenen Sektoren des Programms könnten potentiell die Effizienz und Qualität der Dienstleistungen, die traditionell der Öffentlichkeit erbracht wurden, durch die Einführung oder die Entwicklung telematikbasierter Systeme und die Dienste verbessert werden.	Dans les différents secteurs du programme, l'efficacité et la qualité des services traditionnellement fournis au public pourraient être potentiellement améliorées par l'introduction ou le développement des systèmes basés sur télématique et les services.
This new Programme aims to build on the activities already carried out in sectors such as transport, health, education, libraries, telematics engineering, linguistic research and engineering, and electronic publishing.	Dieses neue Programm zielt darauf ab, auf den Aktivitäten zu bauen, die schon auf Sektoren wie zum Beispiel Verkehr, Gesundheit, Ausbildung, Bibliotheken, Telematikingenieurwesen, linguistischer Forschung und Datenverarbeitung und elektronischer Publikation durchgeführt werden.	Ce nouveau programme vise à se fonder sur les activités déjà effectuées dans les secteurs tels que le transport, la santé, l'enseignement, les bibliothèques, l'ingénierie télématique, la recherche et l'ingénierie linguistique, et l'édition électronique.

Sample 2

ORIGINAL - English	EC SYSTRAN - German	EC SYSTRAN - French
Article 1	Artikel 1	Article premier
The Community and its Member States, of the one part, and the ACP States, of the other part (hereinafter referred to as the 'Contracting Parties'), hereby conclude this cooperation Convention in order to promote and expedite the economic, cultural and social development of the ACP States and to consolidate and diversify their relations in a spirit of solidarity and mutual interest.	Die Gemeinschaft und seine Mitgliedstaaten des eines Teils und die AKP-Staaten des anderen Teils (der im folgenden als die 'Kontrahenten' bezeichnet wird,) schließen hierdurch diese Zusammenarbeitskonventi on, um die wirtschaftliche, kulturelle und soziale Entwicklung der AKP-Staaten zu fördern und zu beschleunigen und um ihre Beziehungen in einem Geist von Solidarität und gegenseitigem Interesse zu konsolidieren und zu diversifizieren.	La Communauté et ses États membres, de la seule partie, et les États ACP, de l'autre partie (ci-après considérée comme les 'parties contractantes'), concluent cette convention de coopération afin de promouvoir et accélérer le développement économique, culturel et social des États ACP et consolider et diversifier leurs relations dans un esprit de solidarité et d'un intérêt commun.
The Contracting Parties thereby affirm their undertaking to continue, strengthen and render more effective the system of cooperation established under the first, second and third ACP-CEE Conventions and confirm the special character of their relations, based on their reciprocal interest, and the specific nature of their cooperation.	Die Kontrahenten bestätigen dadurch ihr Unternehmen, um das System der Zusammenarbeit fortzusetzen, zu verstärken und effektiver zu machen, das unter den ersten, zweiten und dritten AKP-CEWG-Konventionen festgelegt wird, und den speziellen Charakter ihrer Beziehungen zu bestätigen, der auf ihrem gegenseitigen Interesse und der spezifischen Art ihrer Zusammenarbeit basiert.	Les parties contractantes affirment ainsi leur entreprise pour poursuivre, renforcer et rendre plus efficace le système de coopération établi en vertu des premières, deuxièmes et troisième conventions ACP-CEE et pour confirmer le caractère spécial de leurs relations, basé sur leur intérêt réciproque, et la nature spécifique de leur coopération.
The Contracting Parties hereby express their resolve to intensify their effort to create, with a view to a more just and balanced international economic order, a model for relations between developed and developing states and to work together to affirm in the international context the principles underlying their cooperation.	Die Kontrahenten drücken hierdurch ihren Entschluß aus, ihre Bemühung zu verstärken, mit Blick auf eine richtigere und ausgewogene internationale wirtschaftliche Ordnung ein Modell für Beziehungen zwischen entwickelten und sich entwickelnden Staaten zu schaffen und zusammenzuarbeiten, um im internationalen Zusammenhang die Prinzipien zu bestätigen, die ihrer Zusammenarbeit zugrundeliegen.	Les parties contractantes expriment leur résolution d'intensifier leur effort pour créer, en vue d'un ordre économique international plus juste et équilibré, un modèle pour les relations entre les états développés et en développement et pour collaborer pour affirmer dans le contexte international les principes à la base de leur coopération.

Sample 3 (Test Suite)[1]

Original	EC SYSTRAN - German	EC SYSTRAN -French
The bandage was wound around the wound.	Der Verband wurde um die Wunde umgewickelt.	Le bandage a été enroulé autour de la blessure.
The farm was used to produce produce.	Der Bauernhof wurde verwendet, um Produkt zu produzieren.	L'exploitation agricole a été utilisée pour produire des produits.
The dump was so full that it had to refuse more refuse.	Die Mülldeponie war so voll, daß sie mehr Abfall ablehnen mußte.	La décharge était si complète qu'elle a dû refuser davantage d'ordures.
We must polish the Polish furniture.	Wir müssen die polnischen Möbel polieren.	Nous devons polir les meubles polonais.
He could lead if he would get the lead out.	Er könnte führen, wenn er heraus das Blei erhalten würde.	Il pourrait conduire s'il obtiendrait le plomb.
The soldier decided to desert his dessert in the desert.	Der Soldat beschloß, seinen Nachtisch in der Wüste zu verlassen.	Le soldat a décidé d'abandonner son dessert dans le désert.
Since there is no time like the present, he thought it was time to present the present.	Da es keine Zeit wie die Gegenwart gibt, dachte er, daß es die Zeit war, die Gegenwart darzustellen.	Comme il n'y a aucun temps comme le présent, il a pensé qu'il était temps de présenter le présent.
A bass was painted on the head of the bass drum.	Ein Barsch wurde auf dem Kopf der tiefen Trommel gemalt.	Une perche a été peinte sur la tête du tambour bas.
When shot at, the dove dove into the bushes.	Wenn geschossen auf die Taubetaube in die Büsche.	Lorsque germé, la colombe de colombe dans les buissons.
I did not object to the object.	Ich erhob nicht gegen den Gegenstand Einspruch.	Je ne me suis pas opposé à l'objet.
The insurance was invalid for the invalid.	Die Versicherung war ungültig für ungültig.	L'assurance était invalide pour l'invalide.
There was a row among the oarsmen about how to row.	Es gab eine Reihe unter oarsmen darüber, wie man rudert.	Il y avait une rangée parmi oarsmen sur comment ramer.
They were too close to the door to close it.	Sie waren ebenfalls nahe der Tür, um sie zu schließen.	Ils étaient également près de la porte pour la clôturer.
The buck does funny things when the does are present.	Buck tut lustige Dinge, wenn tut vorliegen.	Le buck fait des choses amusantes quand fait sont présente.
A seamstress and a sewer fell down into a sewer line.	Eine Näherin und ein Abwasserkanal fielen herunter in eine Abwasserkanallinie.	Une ouvrière couturière et un égout sont tombés vers le bas dans une ligne d'égout.
To help with planting, the farmer taught his sow to sow.	Um beim Pflanzen zu helfen, lehrte der Landwirt seine Sau zur Sau.	Pour aider à la plantation, l'agriculteur a enseigné sa truie à la truie.
The wind was too strong to wind the sail.	Der Wind war zu stark, um das Segel umzuwickeln.	Le vent était trop fort pour enrouler la voile.
After a number of injections my jaw got number.	Nach mehreren Einspritzungen erhielt mein Kiefer die Zahl.	Après un certain nombre d'injections ma mâchoire a obtenu le nombre.
Upon seeing the tear in the painting I shed a tear.	Über das Sehen des Risses im Malen verschüttete ich einen Riß.	Lors de voir la déchirure dans la peinture j'ai versé une larme.
I had to subject the subject to a series of tests.	Ich mußte das Thema einer Reihe von Tests unterwerfen.	J'ai dû soumettre le sujet à une série d'essais.
How can I intimate this to my most intimate friend?	Wie kann ich dies zu meinem vertrautesten Freund andeuten ?	Comment puis-je annoncer cela à mon ami le plus intime?

[1] Courtesy of Mr Malek Boualem of France Telecom, who submitted the suite to the EAMT (*European Association for Machine Translation*) newsgroup.