[*Terminologie et Traduction* 1.1998, pp.319-326]

JOHN BEAVEN

Future MT Developments

The next few years will see a number of interesting developments taking place not only in the Machine Translation world at large, but also within the Commission's system, aimed at improving and extending the range of services on offer.

The Commission's version of SYSTRAN (known as EC-SYSTRAN) currently runs on an Amdahl mainframe. For a number of technical reasons that will be discussed below, it will soon be necessary to *migrate* SYSTRAN to a more up-to-date computing platform, such as Unix or Windows NT.

Of the 110 language pair combinations possible with 11 languages, SYSTRAN currently covers 17. In fact, some official languages are not covered at all right now. Efforts will be made in the next few years to extend the number of MT language pair combinations available, either by adding new SYSTRAN pairs or by making new commercial systems available.

These two developments, *technological migration* and *acquisition of new language pairs, are* the object of this chapter.

1 Migration to a new computing platform

EC-SYSTRAN currently runs on an Amdahl mainframe computer housed at the Commission's Computing Centre in Luxembourg, under the OS/390 operating system (a variant of MVS). The software is mostly written in assembly language and, as a result of this, it is difficult (and consequently expensive) to maintain and improve, as specialists with the appropriate skills are becoming increasingly rare.

Furthermore, the mainframe is expensive to run and, as part of a long-term strategy to move to open systems, the Computing Centre has announced that it intends to phase out the Amdahl over the next couple of years (ideally by the end of 1999). It will therefore become not only desirable, but also necessary to migrate SYSTRAN. The Commission is interested in exploring a number of possibilities that would enable it to have the service, together with its associated development tools, available on a new computing platform (Windows NT or Unix).

In order to understand the options available for this migration, it is worth reviewing the background of the Commission's version of SYSTRAN and comparing it with its commercial cousin, sold by SYSTRAN SA and its subsidiaries.

1.1 *A tale of two SYSTRANs*

FRANCINE BRAUN-CHEN's article "La traduction automatique à la Commission européenne: d'hier à aujourd'hui" in this volume gives a brief history of SYSTRAN developments at the Commission since 1976.

In parallel to these developments, WTC (World Translation Centre) and later SYSTRAN SA[1], which own the rights for the commercial version of SYSTRAN, have also made substantial enhancements to the system through SYSTRAN SA's wholly-owned subsidiary SSI (SYSTRAN Software Inc, La Jolla, California). In addition to the creation of new language pairs and the improvement of existing ones, SYSTRAN SA converted the system to the C language, and brought out a number of commercial versions based on it, notably a standalone version for PCs under MS Windows (3.1, 95 and NT), and client-server versions under the Unix and Windows NT operating systems, all of which are widely available through standard commercial channels.

More recently (December 1997), SYSTRAN SA joined efforts with Digital Equipment Corporation (DEC) to produce a version running on DEC's Alpha processors, and linked it to AltaVista, the popular Web search engine. This is essentially the same version as the commercial one, compiled for the very fast Alpha processors, and has been enjoying tremendous success on the Internet, providing on-the-fly translations of Web pages: by early February 1998, it was translating a million Web pages per day, and the figure was increasing at the rate of 15% per week. Predictably enough, EC-SYSTRAN users who have seen the AltaVista system in action have been asking themselves (and those at the Commission responsible for EC-SYSTRAN) why a similar facility offering almost instant translations of HTML documents is not available under our system. The answer is that there are a number of technical difficulties in doing so under the current setup, but the situation will change once the system has been migrated.

---

[1] Contact for SYSTRAN SA: 26 Avenue de Paris, F-95230 Soisy-sous-Montmorency, France. Tel: + 33-11-3989-9011, fax: +33-11-3989-4934.

As a result of a number of years of divergent developments, the linguistic performance of EC-SYSTRAN is more suited to the Commission's requirements than the commercial version. On the other hand, it could be interesting for the Commission to draw on SYSTRAN SA's considerable experience in porting SYSTRAN to C under Unix or Windows NT. If the best of both systems could be put together, it may well be possible to combine EC-SYSTRAN's linguistic performance on Commission documents with the impressive speed offered today by the commercial version seen on AltaVista.

1.2 *Migration feasibility study*

In order to examine the possible routes available for this migration, a feasibility study has been commissioned to a firm of external consultants.

The primary aim of this study, which should be completed by April 1998, is to evaluate and make recommendations on the best strategy to follow in order to migrate EC-SYSTRAN to a new computing platform (Unix or Windows NT). A secondary objective is to assess the feasibility of merging the two main versions of SYSTRAN currently available, the Commission's and SYSTRAN SA's.

As has been mentioned above, the Commission has made a large investment in the development of terminological and linguistic resources. As a result of these investments, the linguistic quality of raw EC-SYSTRAN output is better (at least for the sort of text that are of interest to the Commission) than any alternative available in the market.

Consequently, a key criterion to be borne in mind when envisaging the various technical solutions available is that this investment should not go to waste: no deterioration in the quality of MT output can be tolerated.

This migration will not only facilitate the maintenance of the system, thus guaranteeing its future for the decades to come: it should also make it easier to offer the integration needed to get on-the-fly translations directly from within your Web browser or word processor interface, in the way a number of popular commercial MT products already do.

2 Acquisition of new language pairs

One may wonder, if all official languages are supposed to be equal, why the Commission has chosen to develop the 17 particular language pairs

available out of the 110 that are possible with 11 official and working languages.

The criteria for developing one language pair instead of another have been based on five main issues:

1 On the internal needs of the Commission. For the SdT, this means prioritising those language pairs with English and French as source language, as these two languages account for over 85% of our originals. On the other hand, users in the DGs who may want SYSTRAN for browsing purposes are more likely to want those pairs with English and French as target, since Commission staff understand at least one of these two languages.
2 On the translation quality expected from related languages (combinations of two Romance languages or two Germanic languages tend to give better results).
3 On the budgetary restrictions imposed on us, which make it impossible to develop 110 language pairs. This point is further elaborated below.
4 On political reasons which do not concern us here.
5 On the willingness of Member States to co-finance the development of a particular language pair. Two new language pairs, English-Greek and Greek-French, are currently being developed with the co-operation of the Greek government, and plans are under way to have a similar arrangement with Portugal.

2.1 *MLIS*

Taking the Greek example as a model, the SdT has expressed its interest in co-operating on MT projects with other administrations in the European public sector, within the framework of Action Line 3 of the Multilingual Information Society (MLIS) programme concerning *Advanced language tools in the public sector,* set up by DG XIII (Telecommunications, Information Market and Exploitation of Research).

Following a call for expressions of interest launched in 1997, there are plans to set up with the Member States concerned at least three projects covering the following languages:

1 Greek: continued development of English-Greek and Greek-French. In addition, introduction of two new language pairs, French-Greek and Greek-English.

2 Portuguese: continued development of English-Portuguese and introduction of new French-Portuguese, Portuguese-English and Portuguese-French.
3 Dutch: development of existing pairs (with Dutch as target) and introduction of new pairs with Dutch as source (into English, French and German).

The conditions under which new language pairs will be introduced are still under discussion, but it is expected that consortia will be set up involving public administrations and research centres in the participating Member States, and these will issue calls for tender for the acquisition of MT language pairs currently available in the market or for the development of those that do not yet exist.

2.2 *Calls for tender for new language pairs*

Finally, the SdT intends to launch calls for tender to acquire MT provision in those language pairs which are not covered by EC-SYSTRAN or any MLIS project, but for which the market may have something to offer.

It is important to have realistic expectations on this front: commercial developers have tended to favour the "big" languages and similar language pairs as the Commission, often for analogous reasons, so while there are several MT systems between English and French for instance, there is nothing between Greek and Danish, a language pair that is also of marginal interest to the Commission (see below). However, there are a number of products available covering language pairs that could be of interest to the Commission. We should also bear in mind that EC-SYSTRAN has been improved over the years to make it particularly adapted to the administrative language used in-house. In terms of the linguistic quality of the raw translations produced, it is much more suitable than any alternatives in the market.

2.2.a Which language pairs

As has been pointed out above, it is impossible for budgetary reasons to develop all 110 possible language pairs. The 17 pairs currently available account for about 62% of the volume of translations carried out by the SdT (which, it could be argued, reflects the multilingual needs of the institution). Given the fact that all officials know at least English or French (and usually can understand both), one way of reducing the new language pairs to be developed would be to concentrate on the

38 pairs involving those two languages, as shown in Figure 1. The 1 7 existing pairs are shown in bold.

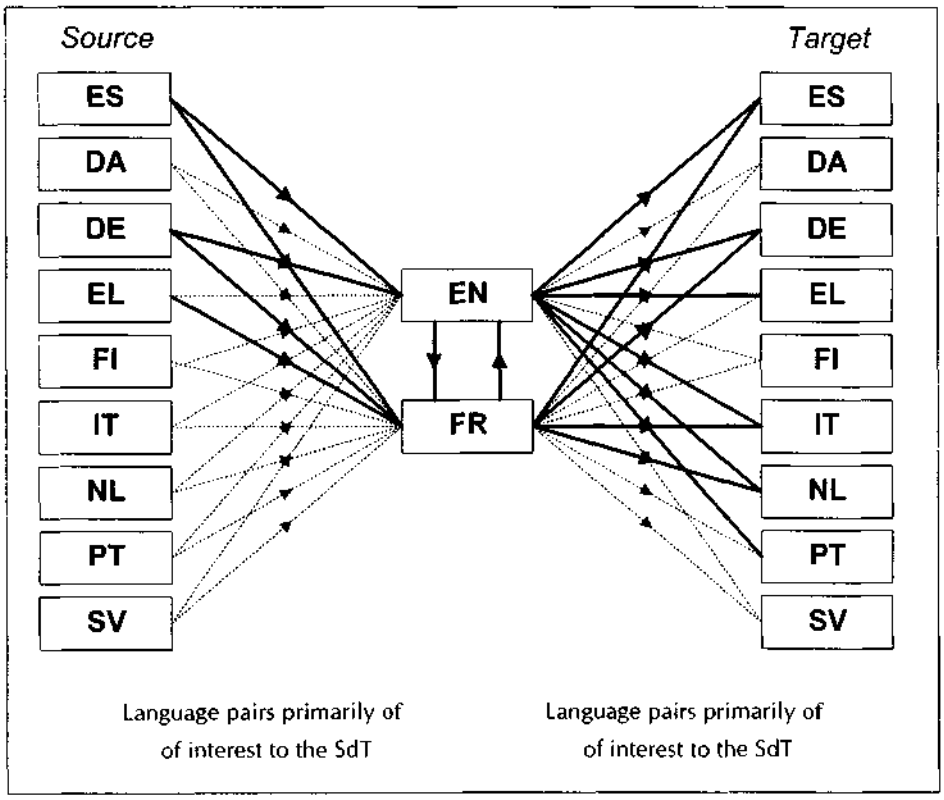

Figure 1: Key language pairs

The Machine Translation Working Croup, representing the interests of MT users, will be consulted in order to draw a list of priorities for new language pairs and for assistance in the evaluation of competing offers when the case arises. This is the reason why the working group includes translators (MT correspondents) in all eleven official and working languages, rather than just the seven target languages covered by EC-SYSTRAN.

2.2.b Licences or teleservices?

There are two possible ways in which the use of MT facilities could be made available to staff. The standard procedure is to acquire a licence to run the software on the Commission's

computers. This is the way in which most software at the Commission (such as Word or Translator's Workbench) is licenced. Licencing fees depend on the number of people using the software, and not on how much use they make of it: you pay for "all you can eat". The alternative is to "pay as you go" by using teleservices: the software then runs on the vendor's computers, which can be anywhere in the world, the source documents and the translations are transmitted electronically between the customer and the company, and payments are dependent on the volume of texts translated or based on a subscription scheme. A number of companies already offer such facilities through the Internet or through dedicated service providers. For instance, subscribers to CompuServe have had access to Machine Translation services for a couple of years: users send the text that they want to translate to a designated address, and get their raw machine translation shortly after, with the fee being automatically added to their monthly CompuServe bill. Logos has been offering MT through the Internet for some time now, and SYSTRAN SA has also recently started doing this, though the company has also been offering MT to French subscribers of the Minitel service for several years. These are but a few examples of what is available. In principle, the SdT is open to both approaches to making the use of new MT pairs available, and it is conceivable that for language combinations without much expected demand, using teleservices rather than purchasing licences may be more advantageous.

3 Conclusions

It is hoped that these new developments, the technological migration and the acquisition of new language pairs, will see the consolidation of Machine Translation as a major tool at the disposal of EU officials for decades to come.

The first revolution in popularising its use took place when DG XIII made it available to all staff through the one interface immediately accessible to everybody who had a computer on their desk: electronic mail. Its use has rocketed since. 1997/8 brought us another major development with the introduction of a more user-friendly Web interface on EuropaPlus and EuropaTeam, the Commission's and inter-institutional

intranet servers. The effect that this new interface will have on demand will only be known over the next few months.

Perhaps we can now look forward to the day when every major application on our desktop PC will feature a "Translate" button offering immediate Machine Translation of word processor documents, Web pages, e-mail message, spreadsheets, and database consultation results.

JOHN BEAVEN