

PROVISION OF SYSTRAN MAINFRAME PRODUCTION SERVICE

James Houston

Cara Data Processing, Dublin, Ireland

Since April 1984, Cara Data Processing Limited have been providing a Systran production service to the European Commission in Luxembourg.

Based on our experience of providing the service over the last two years, I will discuss the subject of Systran production service under the following headings:

Language pairs available in the current production service

The section will merely detail source and target language pairs available.

Operation of Systran production service

The section will discuss how users access the mainframe system, submit text for translation, and retrieve the translated output from the system.

Installation of new releases of the Systran software

The Systran development centre in Luxembourg has proved to be very productive and releases a new version of the software every two to three months. This section will discuss how the new versions are introduced into the production environment without disruption of service to the user.

Reliability and availability in the Systran production service

This section will discuss some aspects of the reliability and availability of the service, obviously items of major importance in a production environment.

Computer resource requirements

This section will detail the computer resources required to support the production service.

Manpower requirement

Based on our experience in Dublin, this section will detail the manpower resources required to support the production service.

Language pairs available in current production system

The production service currently supports the following source and target language pairs.

English	-	French
English	-	German
English	-	Italian
French	-	English
French	-	German

The French-German language pair is the most recent.

Operation of Systran production service

Text to be translated is prepared on the standalone workstation in the Commission's premises in Luxembourg. During this process the workstation is not connected to the mainframe in Dublin. In the Commission's premises both a Wang computer and NCR Tower computer are used for workstation support, and I believe that optical character recognition equipment is increasingly being used to read source text directly into the workstation rather than typing the text into the workstation.

In order to connect to the Cara mainframe the workstation must appear to the mainframe to be an IBM 3780 batch terminal. This is a very common protocol and most minicomputers and business microcomputers can emulate a 3780 terminal.

When the text is ready for transmission to the mainframe, the workstation dials up the Cara computer, identifies itself as an authorised user of the system, and transmits the text, preceded by the appropriate job control (JCL) cards. The workstation logs off from the system at this point and disconnects the telephone line, owing to the fact that transmission handled over a trunk telephone line from Luxembourg to Dublin.

The Cara computer selects the input translation job automatically for processing and executes the job in the next available partition. Most translation jobs are completed within 15 minutes of receipt of the text. The output from the translation is stored on disk at the mainframe under the appropriate authorised user identification.

Subsequently the workstation in Luxembourg redials the Cara computer and retrieves its now translated output.

The dial-up link between Luxembourg and Dublin operates at a line speed of 2400 baud giving a transmission rate of, very approximately, 240 characters per second. We have therefore been concerned to keep the amount of data transmitted on this link to a minimum.

- 3 -

There are two types of output from the translation job. One is the translated text itself, and the other is the associated diagnostic listings produced as a by-product of the translation process. The user wishes to receive the translated text as soon as possible, but the diagnostic listings are of less immediate concern and in fact I understand they are mainly of interest to the Systran development team.

We have therefore changed the Systran production procedures to allow the diagnostic listings to be held under a separate user identifier, so that in the normal course of events a user accessing the system to retrieve output will receive only the translated text. At a time more convenient to him, the user can access the mainframe and supply the appropriate identifier to receive the diagnostic listings for the translation jobs previously processed.

We have further examined the diagnostic outputs from the system and suppressed all outputs which will be required in a development phase but are not needed in a production environment.

Installation of new releases of the Systran software

Every two or three months a new version of the translation software dictionaries is released from the Systran development team in Luxembourg. The new version is distributed to each user in the form of a magnetic tape containing about 60 files, including both program libraries and dictionary files.

It is of course essential that the new release be implemented with minimum disruption to the production service.

The system is set up in such a way that the user does not make any change to his job control cards when moving from one version to a later version; he simply continues to refer to the production version.

On receipt of a new release of the software, Cara loads the files under a test name and runs sample jobs against each language pair supplied. We are not testing the quality of the translations, but merely ensuring that each execution of the translation software terminates normally.

When the new release has successfully completed the sample translation runs, it becomes the production version.

During the installation and testing of the new release, the user has the current production version available to him, and can continue to run translations. The process of making the new tested version into the new production version is accomplished using a 'command procedure' to rename and recatalogue the appropriate files. During the execution of this command procedure, translation jobs cannot be run. However, since the execution time for the command procedure is less than one minute, it is a trivial matter to ensure that the users do not experience any interruption in service.

When a new production version is introduced, the previous release is not deleted from the system but retained as the 'last' version.

This 'last' version is retained for reasons derived from 'Murphy's Law' which of course states that what can go wrong, will go wrong, and at the worst possible time.

If a production version with serious defects is introduced, the user can continue to run translations using the 'last' version while the current system defects are resolved.

Our experience has been that the new releases of the software are very reliable and in the past two years I can recall only one occasion when we have had to use the 'last' release of a particular language pair for a short time.

Reliability and availability of service

In any one execution of the Systran system, the computer is executing 17 program steps to which are supplied at least 40 separate parameters. If the user had to supply these parameters on each run there would clearly be scope for him to encounter a frustrating series of failures as one or more of the parameters proved incorrect.

Systran shields the user from such complexity and in the average translation run the user is required to supply only three parameters, the language of the source text, the target language required, and the name of the text file to be translated.

The Systran program will always attempt to recover from errors encountered during the translation process. Even when the error is serious enough for the program to incur abnormal termination, Systran will intercept the termination request, write appropriate records to the diagnostic output for later analysis and continue the translation process.

As mentioned previously, the production version of the software is effectively available at all times including the period during which a new version is being installed.

Again, if there is a defect in the production version for any language pair, the fact that the previous version is always available to the user, allows him to continue to run translations.

Computer resources required by Systran system

Operating System	IBM MVS
Programming Languages	IMB Assembler and PL/1
Partition Size	1 Megabyte
Disk Space	50 Megabytes per Version
Translation Rate	1,500 Words per CPU Minute Based on IBM 158

- 5 -

Manpower requirements

Filing of JCL Listings	2 hours per month
Installation of New Release	0.5 man days after procedure is streamlined.
User support	0.5 man days per month

Conclusions

Cara has found that the Systran system is a very robust system in operation. Upgrades to the system are provided on a regular basis and have proved to be very reliable. Taken together, these factors mean that a continuous, reliable, and increasingly effective machine translation service can be offered to users.