

"THE SYSTRAN RENAISSANCE"  
By Denis A. Gachot

1) Introduction

The purpose of this paper is to describe the improvements and changes within the SYSTRAN system and its organization, over the last 3 years. Our description will focus in the following areas: Organizational Changes, SYSTRAN system Enhancements, and Commercial Activities.

2) Organizational Changes

As you may know, the ownership of all SYSTRAN systems, with the exception of the Japanese system, and its corresponding development companies, was acquired by the Gachot Group in 1985.

The Machine Translation activities within the Gachot Group are divided amongst the following 3 development centers:

- LA JOLLA, CA, USA: LATSEC, INC. (strictly a U.S. Government contractor) and SYSTRAN TRANSLATION SYSTEMS, INC. (formally WORLD TRANSLATION CENTER, the commercial company). The La Jolla group presently employs a staff of approximately 30. This group is responsible for administrating all development work from U.S. customers, operating the U.S. network, and the overall technical responsibility for all SYSTRAN systems.
- PARIS, FRANCE: SYSTRAN S.A. has the overall development responsibility for the English into Arabic system, the operation of the Minitel Network, the European PC network, as well as the coordination of the linguistic and dictionary improvements within the 3 groups.
- LUXEMBOURG: SYSTRAN LUX, linguistically and technically responsible for the Arabic system, handles linguistic development for our European customers.

In order to avoid effort duplications, specifically in the area of coding, a direct line has been installed connecting the La Jolla and Paris mainframes. Since all coders on both sides of the Atlantic physically code into the same dictionary, this duplication concern is eliminated. Another advantage is the fact that effort coordination, as well as technical support, is extremely fast and convenient as most answers are sent via electronic mailbox within a 24 hour period.

We are also pleased to announce that we are actively pursuing cooperation with the CEC to avoid this effort duplication problem.

### 3) SYSTRAN System Enhancements

The changes and enhancements to the SYSTRAN system can be classified in the following 3 major topics.

- Basic Software
- Language Pair Enhancements
- Customer Specific Dictionary and other Utilities

#### 3.1 Basic Software

As per the request of the U.S. Government, the SYSTRAN basic software underwent major modifications in order to achieve compatibility with the IBM VM/CMS operating system. As a result of these modifications, the system is now a one step procedure. This means that translations are no longer carried out using a batch process system. Consequently, SYSTRAN literally translates sentence by sentence and no longer is required to scan the entire submitted text prior to translation processing. These modifications allow us to cut the processing time in half for small and medium length texts.

These transformations also enable us to operate SYSTRAN on a standalone IBM PS2 model 60, 70 or 80 equipped with a IBM 7437 processor.

#### 3.2 Language Pair Enhancements

One of the first tasks our group undertook was a translation quality evaluation of the various language pairs. To accomplish this task we devised a method to judge the accuracy of our language pairs. A data base of various texts (anonymous to the developers), is periodically tested against newer versions of our system. Depending on the type and gravity of the mistakes, points are deducted from an ideal translation. Partial results of this quality analysis are summarized in figure 1. It should also be noted that the possible origin of the mistakes are categorized (analysis, synthesis, dictionary and others) so that we are able to quantify the effect of various enhancement projects on our system.

Since 1985, our group has developed additional language pairs, namely English into Arabic, English into Dutch, French into Dutch, Spanish into English, Italian into English, Portuguese into English as well as English into Korean and Korean into English.

During the development of these various language pairs, numerous efforts have been made to maintain the modularity of SYSTRAN. Consequently, now all SYSTRAN system linguistic components, as shown in figure 2, are totally independent. In other words, the English analysis of the English into French and English into German systems are identical. Similarly, the syntheses are also modular, and all dictionaries are multitarget, with one common source language coding, regardless of the target language.

We also have further refined this modularity by introducing the concept of a romance language analysis "Trunk", designed and implemented for all romance language analysis. All this effort into modularity, greatly enhanced our productivity and capabilities.

### 3.3 Customer Specific-Dictionary and other Utilities

As a result of the constant coding efforts done in La Jolla, in conjunction with the Paris and Luxembourg groups, and our cooperation with the European Commission, the SYSTRAN dictionaries of various language pairs have been dramatically increased. The result of all this effort and cooperation can be seen in the relatively large dictionary sizes as described in figure 3. In addition, a change occurred in the SYSTRAN philosophy as far as dictionary handling is concerned. Through discussions with our customers, it has come to our attention that it was necessary to devise a way in which they could have an impact on the output. After numerous studies, it was decided that the best way to handle this situation was to create the SYSTRAN Customer Specific Dictionary (CSD). This buffer dictionary, limited in size, will accept entries made directly by the customer without any intervention of SYSTRAN personnel, and will act as an override of the main dictionary during the translation process. Once this dictionary has reached the authorized limit, the entries are then verified and completed by SYSTRAN personnel prior to being integrated into the main dictionary.

It is also worth mentioning that many utilities have been developed around the SYSTRAN system. For example concordances generators, translation comparators and sentence structure extractors. The latest enables our linguists to extract from any given text all sentences with a specified linguistic type or form. This utility greatly enhanced our testing capability.

#### 4) Commercial Activities

Our group, in this Machine Translation field is primarily involved in the development of Machine Translation system software. Our commercial activity focuses on providing the best available Machine Translation services by listening to our customers and constantly enhancing our SYSTRAN system.

The marketing and sales activities of our group in regards to the SYSTRAN system, have been clearly defined as follows:

##### 4.1 SYSTRAN LEASING

Installation and usage of one or several SYSTRAN language pair systems at customer's location. This option is for the customer with large volumes of translations that cannot be sent via a modem to SYSTRAN'S Main Frame for reasons of confidentiality.

Presently the two following options are available:

- MAINFRAME OPTION;

To be installed on a IBM 370 series Mainframe (or equivalent) running under MVS or VM/CMS operating system.

- STANDALONE WORKSTATION/PC:

To be installed on a IBM PS2 Model 60,70 or 80 with an IBM 7437 processor or an IBM PC AT/370, running IBM VM/PC Software. VM/PC will enable you to run SYSTRAN translations in the background while running DOS applications. Storage requirements are approximately 20 meg bytes per language pair. Translation processing speed will vary depending on hardware.

#### 4.2 SYSTRAN EXPRESS

Access via electronic network, all SYSTRAN operational language pairs installed on our La Jolla based IBM Mainframe.

This option will allow the user to translate their source documents from anywhere in North America, with just a PC and a dedicated phone line, by calling the local access to the TELENET Network.

##### SYSTRAN EXPRESS REQUIREMENTS:

##### HARDWARE:

IBM PC, XT, AT, PS/2 (or compatible) with a minimum of 640K RAM.

Asynchronous Modem (Hayes Compatible) with either 1200 or 2400 Baud Rate.

A dedicated phone line. (Connections through switch boards may not work for data transmission).

##### SOFTWARE:

SYSTRAN EXPRESS access software allows you to select the desired language pair as well as other SYSTRAN parameters.

SYSTRAN software will provide the customer the capability to communicate with SYSTRAN'S Mainframe.

Any Word Processor that allows conversion into ASCII characters.

#### 4.3 DEVELOPMENT

Because of our highly specialized personnel and vast experience in the field of Machine Translation, our Company is capable of undertaking the following in any possible language pair combination:

- Development of a new system language pair.
- Enhancements to an existing SYSTRAN language pair.

5)

Conclusion

This rapid overview of our company's achievements in the SYSTRAN System during the last 3 years is an indication that we are fully committed to maintaining SYSTRAN at the top of the Machine Translation field. We are also looking into the future with the development of multilingual electronic mailboxes and the possible integration of the latest Artificial Intelligence Techniques.

However, we do have the firm belief that our commitment to providing Machine Translation services rather than just a system, to our customers is definitely the right approach. It is essential that a solid relationship exist between the MT developer and user so that the implementation of a MT system within any given organization becomes successful. Based upon the recent increase in our commercial activities, we feel that we are finally on the right track.

FIGURE 1

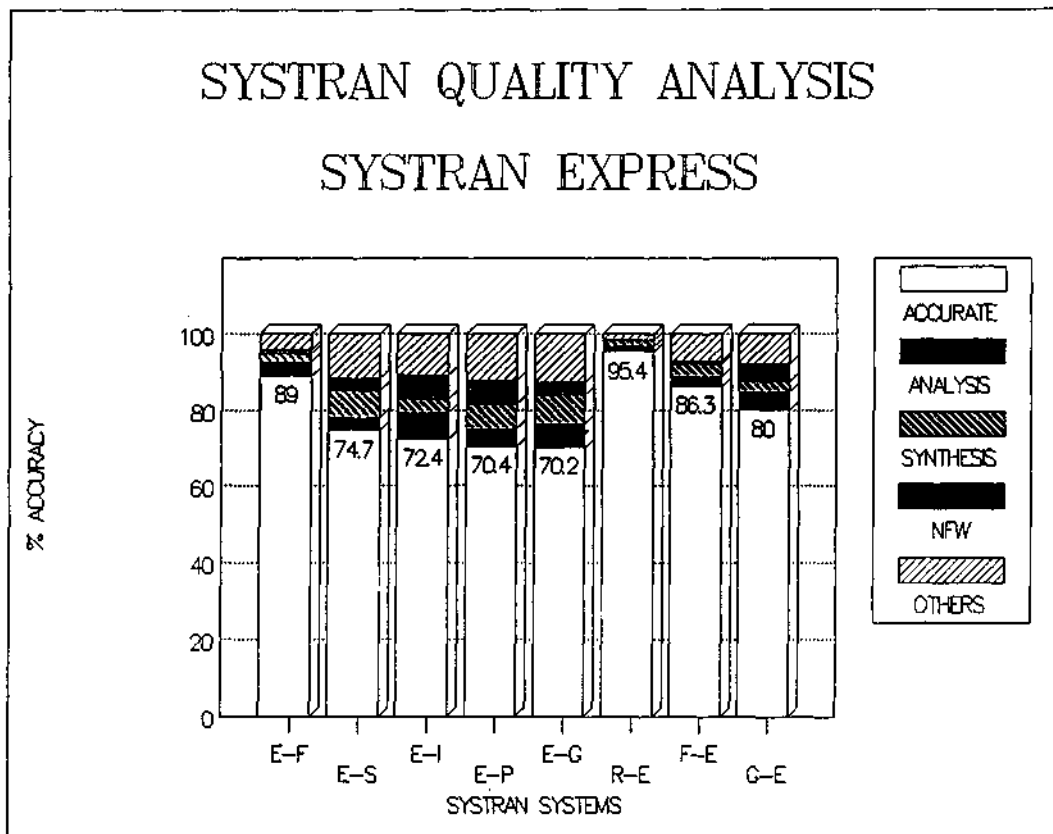
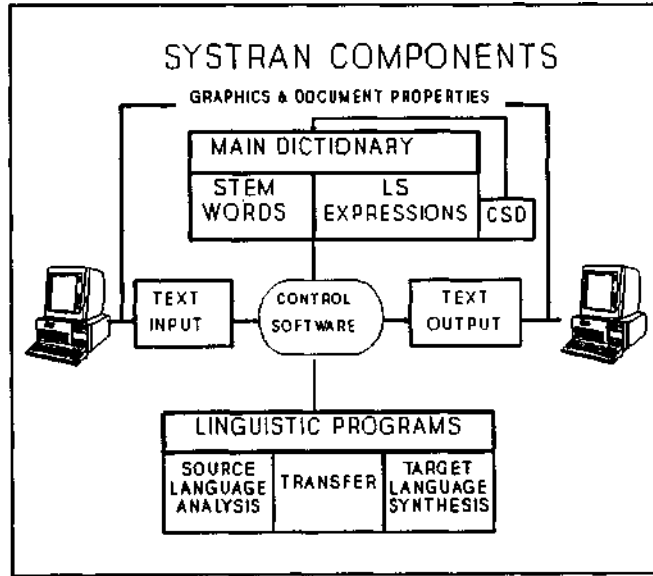


FIGURE 2



DICTIONARY SIZES:

SYSTRAN DICTIONARY ENTRY TOTALS (as of 1/3/89)				
	STEM ENTRIES (MEANINGS)		IDLS ENTRIES (MEANINGS)	
ENGLISH-MULTI-TARGET	95467		62088	
ENGLISH-FRENCH		132874		43046
ENGLISH-GERMAN		96651		31745
ENGLISH-ITALIAN		103009		41591
ENGLISH-PORTUGUESE		31542		1561
ENGLISH-SPANISH		44872		1604
ENGLISH-KOREAN		6270		121
ENGLISH-ARABIC	*70000	*70000	*10000	*10000
ENGLISH-DUTCH	*30400	*30400	*3400	*3400
ENGLISH-JAPANESE		61802		4681
ENGLISH-RUSSIAN	13360	25366	5969	9407
FRENCH-ENGLISH	67848	69631	50741	68413
FRENCH-DUTCH	*28000	*28000	*6200	*6200
FRENCH-GERMAN	62994	47398	0	0
GERMAN MULTI-TARGET	125171			
GERMAN-ENGLISH		154133	4553	5429
GERMAN-FRENCH		26787	1308	1790
GERMAN-ITALIAN		17649	7	5
GERMAN-SPANISH		46153	1198	1472
JAPANESE-ENGLISH	44693	55129	3185	4352
RUSSIAN-ENGLISH	203589	317462	142549	170354
SPANISH-ENGLISH (MT)	32330	36380	339	59
ITALIAN-ENGLISH (MT)	1972	2124	69	3
PORTUGUESE-ENGLISH (MT)	221	264	95	0
KOREAN-ENGLISH	1994	2759	11	18

FIGURE 3

Note: (MT) Multitarget  
 \* Indicates approximate number