[*Terminologie et Traduction*, no.2, 1985]

MT WITH SYSTRAN AT KARLSRUHE NUCLEAR RESEARCH CENTER

1. Introduction

For many years the Fast Breeder. Project (PSB) of the Karlsruhe Nuclear Research Center (KfK) has enjoyed a close cooperation with France concerning the development of fast breeder reactors, and consequently its scientists receive many French reports. To ease the transmission of information, in 1979 PSB started to investigate whether machine translation could be used here. There is now a small development project, with a staff of three people, cooperating with the European Commission. KfK has received the Systran programme for translation from French into English from Luxembourg in exchange for feedback of technical vocabulary and general suggestions for improvement of the programme.

2. Objective

Our aim is to demonstrate that fully automatic translation can be used in the transmission of scientific information. This is a limited goal and we have already received several positive responses from users of Systran translations. Nevertheless, we have a number of difficulties and we do not think that we have reached our goal yet. Our restrictions and difficulties are listed in Figure 1.

. In the first place we have limited the application of MT to informative scientific texts

. Secondly we are concentrating on French-English translations with Systran

. Thirdly we do not offer post-editing of the raw MT output to the users.

Regardless of these restrictions we still have problems enough:

. Input has been done manually so far, but we consider this too slow and want to use an Optical Character Reader to make the printed texts machine readable. .

. We have not yet reached complete reliability of information transfer, although up till now we have entered about 10,000 technical words from the nuclear field in the Systran dictionary and the Commission in Luxembourg has improved the Systran programme appreciably in recent years.

. At present we leave it to the users to cope with the problem of checking raw MT output. Preferably they should seek the assistance of a colleague in their field who knows some French and English.

For our purpose the task of a translator is to improve the computer programme, and not to perform routine translations or post-editing .

3. Chances for MT

If machine translation is to be successful, one of the main requirements is that it should be much faster than conventional human translation. To provide insight into the speed of MT, the times needed for the various successive operations are listed in Figure 2.

. Manual input can be done at a rate of 3-5 pages per hour
. Optical character reading by machine is much faster, but is not free from errors.
. Correction of the input text might cost much more time; 10-20 pages/hour may be optimistic.
. Translation by the Systran programme is at 1,000 pages per hour
. Printing of the raw MT output is also very rapid
. Post-editing, on the contrary, is a tedious and difficult task.

It appears that human operations are by orders of magnitude slower than machine operations. To obtain the full benefit of MT, it is therefore mandatory to limit human work to a minimum.

4. Translation Quality

Evaluation of the quality of a translation is a very personal matter. A true linguist will appreciate a good style and is hurt if his language is maltreated. A scientist can accept all kinds of insuffiencies in style, both in respect of syntax and vocabulary, as long as the content of the original text is reproduced understand-ably in the other language.

This is our aim at Karlsruhe and this standard automatically leads to a first semiquantitative criterion of translation quality: how many sentences in a text sample are translated understandably. This criterion can be differentiated somewhat further by distinguishing between understandable, partly understandable, and incomprehensible sentences. This kind of analysis of a text is quite crude, but it is quantitative. Subsequent analyses of a text sample after updating of the translation program will indicate whether and how much improve-ment was achieved each time.

We have used two typical French nuclear technology texts as samples for statistical analysis. These texts were translated 5 times with different updated versions of the Systran programme over the course of 5 years. The comprehensibility of successive translations is shown in the histogram of figure 3. The number of understandable sentences is indicated by the lower blocks of the histogram. Init-ially 75 to 80% of the translated sentences were understandable. Now, in 1985, more than 95% are understandable. The dotted blocks indicate partly understandable sentences. The remainder are incom-prehensible . The number of incomprehensible sentences has fallen from 6% in 1980 to 1% in 1985. This supports our belief that it is possible to set up an MT service without post-editing.

5. Implementation

Figure 4 shows how we intend to handle machine translation of reports at Karlsruhe.

- The French text will normally be received in the form of a photocopied typewritten document.
- This document will be read page by page by an Optical Character Reader.
- The output of the OCR will be received by a Personal Computer and its format converted into Systran input format. The text will be scanned for spelling errors.
- The information will then be transferred to our main frame IBM, where the words occurring in the text will be compared with the Systran dictionaries. If the text still contains mutilated words due to reading errors of the OCR these will show up in the "Not Found World Last" and can be corrected.
- Then the corrected French text will be translated into English by Systran and printed out, side by side, with the source text. Transmission via data line to the user is possible.

Two pages of recently translated texts are added following the figures. The first sample is typical for the quality of raw MT output which we have reached. The second page is a favourable example of which we consider the translation to be good enough to be handed to the user without post-editing.

FWA HABERMANN

Kernforschungszentrum

Karlsruhe

FIG 1: <u>MT WITH SYSTRAN AT KARLSRUHE NUCLEAR CENTER</u>

OBJECTIVE: FULLY AUTOMATIC MACHINE TRANSLATION

LIMITATIONS:

- INPUT: INFORMATIVE SCIENTIFIC TEXTS
  PROBLEM: CONVERSION OF TYPEWRITTEN TEXTS IN MACHINE READABLE FORM
- TRANSLATION: NON-INTERACTIVE FRENCH-ENGLISH MT WITH SYSTRAN
  PROBLEM: RELIABILITY OF TRANSLATED INFORMATION
- OUTPUT: WITHOUT POSTEDITING
  PROBLEM: CHECK IF INFORMATION OF SOURCE TEXT IS TRANSMITTED CORRECTLY

FIG 2: <u>SPEED OF MT IMPLEMENTATION</u>

		PAGES/HOUR
INPUT	. TYPIST	3-5
	. OCR	100
	CORRECTION OF OCR	10-20?
TRANSLATION		
	. SYSTRAN	1000
OUTPUT	. LASER-PRINTER	500
POSTEDITING		
	. TRANSLATOR	2-5

FIG. 3: COMPREHENSIBILITY OF SYSTRAN TRANSLATED TEXTS

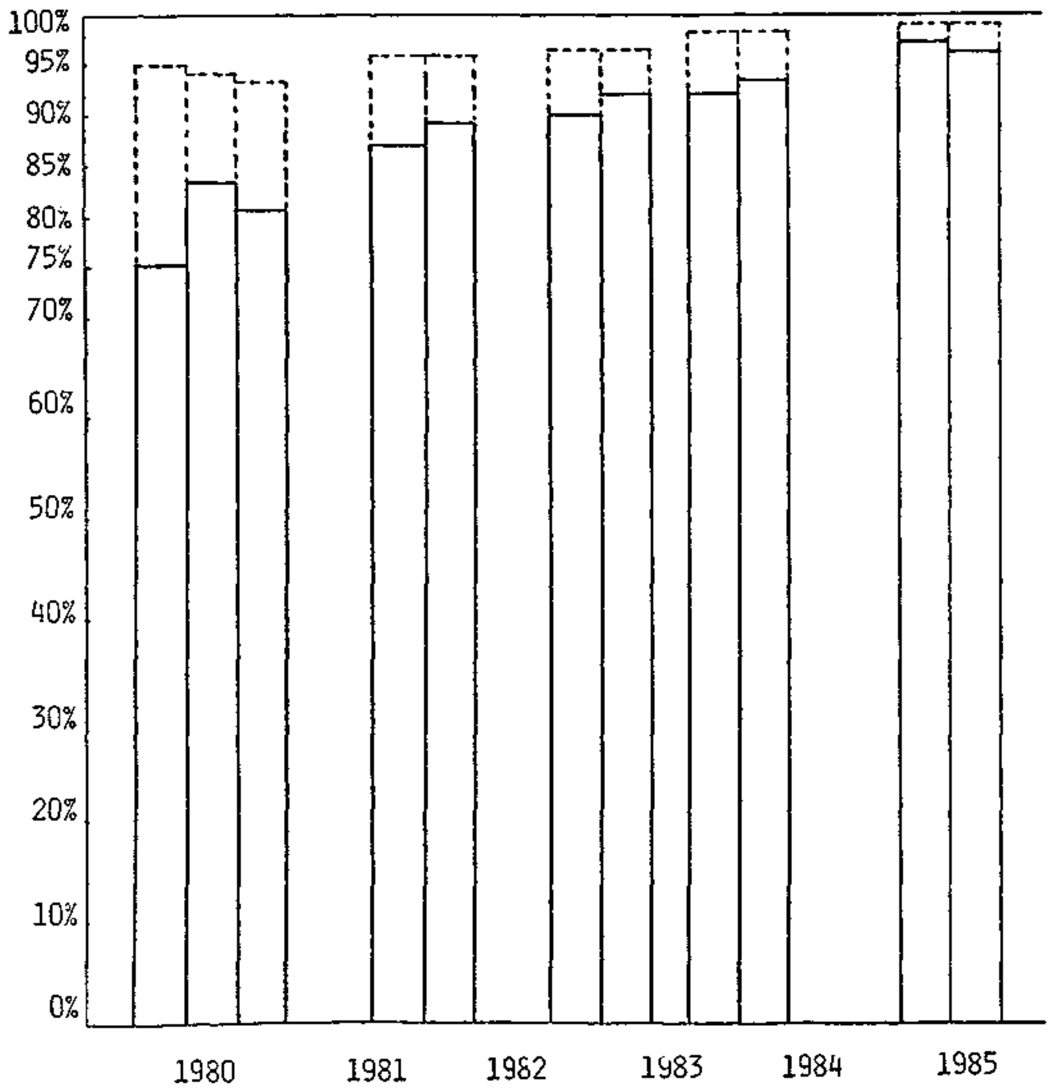
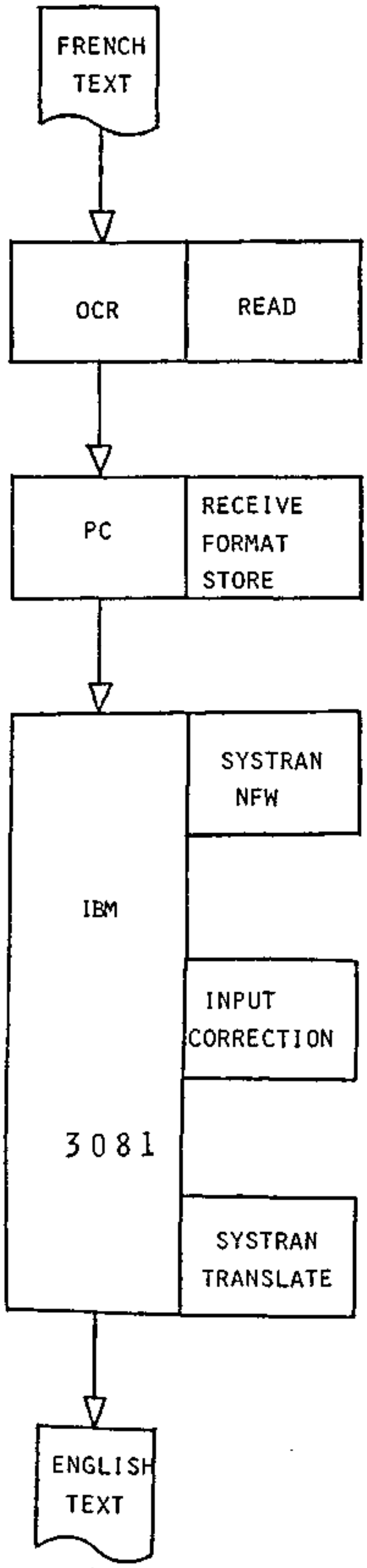


FIG 4: **FLOW DIAGRAM OF MACHINE TRANSLATION**

- INPUT WITH OCR
- TRANSLATION BY SYSTRAN



- TYPEWRITTEN SCIENTIFIC REPORT
- OCR = OPTICAL CHARACTER READER
- RECEIVE OCR OUTPUT
- CONVERT INTO SYSTRAN INPUT FORMAT
- STORE
- SYSTRAN DICTIONARY LOOK UP AND LISTING OF NOT FOUND WORDS
- MANUAL INPUT CORRECTION AT TERMINAL WITH HELP OF NOT FOUND WORD LIST
- FULL RUN OF SYSTRAN TRANSLATION PROGRAM
- SIDE BY SIDE PRINT OF FRENCH SOURCE TEXT AND ENGLISH TRANSLATED TEXT
- TRANSMISSION VIA DATA LINE POSSIBLE

5055/85

SYSTRAN II — SIDE-BY-SIDE

- 23 -

l' analyse de l' essai CABRI B1 était *difficile* en raison des effets bidimensionnels du mouvement et de la perturbation apportée par la cage de centrage. Le code a retrouvé l' instant et l' endroit où s' amorce le mouvement et le fait *que celui-ci est d' abord* ascendant. La propagation du front de fusion et la hauteur de la zone fondue ont été calculées à 7 % près. De plus comme dans l' expérience on constate que la relocalisation se fait à l' intérieur de la *zone fondue, c' est-à-dire* dans la partie fissile (faible pénétration de l' acier liquide sur la gaine intacte) et que les centreurs constituent une zone de resolidification préférentielle. Par contre le calcul n' a pas retrouvé la fusion des centreurs qui a pourtant été observée. Par ailleurs on peut vérifier sur la figure 32 que si l' emplacement de certains bouchons est correct, d' autres *n'* ont pas été *retrouvés*; mais ceci est lié à des particularités de l' expérience qui ne se retrouveront pas en réacteur. On peut donc considérer que ces essais ont permis de qualifier le code ALFA; des essais complémentaires permettront de préciser l' effet des centreurs (essais MONOBRI) et de l' irradiation (essais CABRI). Par ailleurs ils ont permis d' établir plusieurs points dont nous verrons l' importance pour l' application au réacteur:

396 - 23 -
397 The analysis of the CABRI test B1 was *difficult owing* to the
two-dimensional effects of the movement and of the
perturbation brought by the cage of centering. The code
398 found the moment and the place when starts the movement and
the *fact that this one is initially ascending.* The
399 propagation of the front of fusion and the height of the
melted zone were calculated with a margin of 7 %. Moreover
400 as in the experience it is noted that the relocalisation is
done inside the melted zone, i.e. in the fissile part (small
penetration of molten steel on the intact clad) and that the
centring devices constitute a preferential resolidification
zone. On the other hand calculation did not find the fusion
401 of the centring devices who however was observed. In
402 addition one can check on figure 32 only if the site of some
plugs is correct, others were not found; but this is
403 connected with characteristics of the experience who will
not be found in reactor. One thus can consider that these
404 tests allowed to qualify the ALFA code; complementary tests
405 will make it possible to specify the effect of the centring
devices (MONOBRI tests) and of the irradiation (CABRI
tests). In addition they allowed to draw up several points
406 of which we will see the importance for the application to
the reactor:

SYSTRAN II		SIDE-BY-SIDE
- 29 -	502	- 29 -
Dans B3 au contraire par suite de la déformation de la gaine, la pression interne a été réduite à une valeur très proche de la pression régnant dans le canal.	503	In B3 on the contrary in consequence of the deformation of the clad, the internal pressure was reduced to a value very close to the pressure reigning in the channel.
La masse éjectée est alors inférieure à 1 g, valeur qui ne peut être vérifiée expérimentalement puisqu' elle se situe au-dessous de la limite de sensibilité de l' hodoscope.	504	The ejected mass is then lower than 1 g, value who cannot be checked in experiments since it is located below the limit of sensitivity of the hodoscope.
Néanmoins, en l' attente de résultats de confirmation, nous pouvons considérer le calcul d' éjection comme correct.	505	Nevertheless, in waiting of results of confirmation, we can consider the ejection calculation like correct.
En ce qui concerne l' interaction on admet que la masse de combustible éjectée se fragmente instantanément et que la zone d' interaction est homogène en température, pression et taux de vide.	506	With regard to the interaction it is admitted that the ejected fuel mass splits up instantaneously and that the interaction zone is homogeneous in temperature, pressure and rates of vacuum.
Les deux paramètres du calcul sont le rayon des grains (actuellement pris égal à 100) et la hauteur de la zone d' interaction ZI.	507	Both parameters of calculation are the radius of the grains (currently taken equal to 100) and height of the zone of interaction ZI.
Dans le cas de A3 la valeur ZI = 10 cm permet de retrouver correctement l' évolution du débit supérieur et le début de l' évolution du débit inférieur (figure 35).	508	In the case of A3 the value ZI = 10 cm makes it possible to find correctly the evolution of the higher flow and the beginning of the evolution of the lower flow (figure 35).
Par contre c' est la valeur ZI = 1 qu' il faut utiliser pour obtenir le même type de résultats dans B3 (figure 36).	509	On the other hand it is the value ZI = 1 that it is necessary to use to obtain the same type of results in B3 (figure 36).
La même tendance se retrouve en ce qui concerne l' évolution	510	The same tendency is found with regard to the evolution of