MACHINE TRANSLATION'S BIG TEN YEARS

1930 - 1990

Loll ROLLING

The Commission of the European Communities

LUXEMBOURG

May 1982

Machine translation is here to stay, imposed by the pressure of the market : 150 million pages are actually translated per year, and 50 million remaining untranslated for lack of translators. The market volume is increasing by 10 per cent every year. Three parallel developments are taking place today :

. Eurotra, the research project jointly undertaken by the European Commission and the member countries of the E.C., is expected to yield a prototype system by mid-1987, for the translation from any of seven E.C. languages into any other E.C. language.

. Systran—type operational systems are slowly taking a large part of the industrial market in a limited number of the world's trade languages.

. Pocket and desk-top translators based on man-machine interaction are expanding rapidly into a new market of low quality products.

The likely developments of these types of systems are assessed, and a number of predictions made on the state of the translation profession in 1980.

Fully automatic translation *is* impossible. In other words, it is impossible for a machine to do exactly what a human translator does.

One may also say that it is impossible for human beings to fly - nevertheless, if necessary we can be in New York this evening without having to spread our wings, thanks to Boeing who have designed fine aircraft and a pilot who has been properly trained. But since the 747 cannot land on the roof of our hotel, we still have to take a taxi to get us to our destination, supplementing the "automatic flight" by a more traditional activity.

Machine translation, then, which is not entirely automatic, has nevertheless made its breakthrough. While its initial appearance, during the Fifties, may have been due to the curiosity of some researchers who wanted to see what else one could make the computer do, this new toy, nowadays it is imposed on us by the pressures of the market. We have shown that it is economically viable at the present level of quality; it will be increasingly used in trade and industry, thus justifying the additional expenditure of further improving its quality.

The translation market

The study of the translation market, completed in 1981 at the European Commission's request, shows that in one year 150 million pages were translated by 175 000 translators, professional or amateur, that their turnover was of the order of 18 thousand million FF, and that these figures *are* increasing by 9 to 10% per annum.

And that 50 million pages, translation of which would have been desirable, could not be translated owing to a lack of qualified translators.

How is this market made up? Who are the users ? This is shown in the following table (fig. 1):

If the international institutions, from the European Community to the United
Nations, have the largest translation services and the most qualified translators, the major market is nevertheless the industrial one. Hundreds of thousands of pages of publicity brochures and maintenance manuals have to be translated before an industrial firm can hope to sell a pharmaceutical product, a domestic iron, or a jet fighter across the linguistic borders.

By comparison, the translation efforts of the multilingual countries like Canada, Switzerland or Belgium appears less important, and literary translation, on which some translator schools still concentrate, represents no more than a negligible percentage of the total effort.

The cost of "human" translation varies considerably as a function of the countries concerned, the languages involved and the type of text to be translated (fig. 2).

While a rate of 10 to 20 FF per 100 words is applied in the United States and in Israel, the price is between 35 and 50 FF in France and between 70 and 80 FF in Germany or Sweden. Higher rates are applied for translation into the rarer languages, and may exceed 100 FF per 100 words. The international institutions also have fairly high translation costs, both because their translators, having been selected for their competence, are particularly highly paid, and also because it is necessary for these institutions to have a reserve translation capacity.

The cost of machine translation varies from one system to another, but is usually below that of human translation; it is of the order of 4 FF/100 words for Systran. But to make a meaningful assessment, it is necessary to compare translations of equivalent quality, i.e. to the cost of the translation as such must be added the costs of the input and of the checking and correction by a human reviser; in this case one arrives at approximately 40 FF for Systran.

It thus appears that a saving is made each time an entirely human translation is replaced by a "post-edited" machine translation.

As however the significant investment put into the development of the systems and the dictionaries has to be amortized, machine translation will be profitable only if one can make it process a considerable number of pages.

The cost of the development of Russian-English Systran was recouped very quickly in the United States, and the annual production of the Commission alone would be enough to amortize the French-English and French-German systems. It would not be profitable, on the other hand, to develop a system to translate Danish into Dutch or Greek.

Survey of translation systems (fig. 3)

Contrary to what one might expect, the most sophisticated systems were invented before the more rudimentary ones, and my classification will therefore not be chronological but typological. At one end of the range, one finds the systems based mainly on data processing. The smallest are the "pocket translators" launched on the market by Craig and Lexicon in 1978, which sold very well, even if very few of them were actually used to translate. The activity of IBM in this field, on the other hand, has always been based on its traditional computers, nowadays the 370 series, which are made to carry out relatively elementary terminological research functions, interactively with a translator.

The most promising approach in this range was adopted by the firms WEIDNER, ALPS and KIRK & WHITNEY, whose systems, being based on a standalone microcomputer, are relatively cheap. As they have only limited dictionaries and rather rudimentary analysis programmes, they do not give satisfactory results for the majority of the users.

At the other end of the range will be found very sophisticated systems, based on linguistic models which are elaborate but, being too complex, not yet operational. Here I may list the GETA system of the University of Grenoble, or the METALS system of the University of Texas.

Making such systems operational requires investment too heavy for any but a group of countries or institutions. This is why the European Commission has launched the EUROTRA project, relying on the competence of the computational linguists in several countries (and not on the embryonic systems already in existence in these countries). It is a system which can be justified only on the basis of the combined needs of research, teaching and production.

In the middle of the range are the systems which combine an empirical approach to the linguistic and terminological problems with an economic solution of the problems raised by the data-processing infrastructure.

This is the approach adopted by the restricted syntax systems, such as TITUS, TAUM-METEO and CULT, and also by the systems designed to translate natural language such as LOGOS and SYSTRAN.

It may be noted that the former have found only limited fields of application (textiles, tropical agriculture, meteorology, mathematics) whereas the latter are applicable in a large number of fields.

Development prospects

The basic problems of machine translation systems are linguistic in nature. But as the creation of linguistic analysis models progresses only laboriously, paradoxically this aspect influences systems' future competitiveness to only a minor degree. It is in fact the technical, financial and psychological aspects which will have the greater effect on the future of machine translation.

The analysis and synthesis softwares will be improved, by refinement of the morphological analysis, homograph resolution and stylistic generation modules. The hardware, becoming less and less expensive, will reduce the cost of the translation and will improve the competitiveness of the systems.

But the most valuable assistance to machine translation will be provided by the new techniques of office automation and telecommunications, simultaneously increasing both productivity and acceptability to the user.

The legal and financial aspects are many and complex. The developer of a system may sell exclusive rights of ownership, he may grant user licences restricted by geographical area or by type of application, or alternatively he may utilize it himself by means of a commercial service, providing raw or post-edited translations either at piece-rates or under inclusive contracts. In certain cases, an erroneous evaluation of the real market has led promoters to adopt methods of exploitation which proved to be unworkable, while in others, astute promoters succeeded in selling systems which were not yet ready for operation.

Nor must the psychological aspects be neglected. All new techniques have given rise to suspicion and irrational defence reactions in the people already on the job. Just as translators successively opposed the typewriter, the dictaphone, the terminology data bank and the "translating machine".

Only the abundant provision of information and retraining of the people concerned to make use of the new resources can help to overcome this real psychosis. It should be noted that freelance translators get over this more easily than translators who are civil servants or members of unions.

In the case of Systran, the gradual involvement of a number of experienced translators has been beneficial for the further improvement of the system. Having covered in turn the current level of development of the various systems, the utilisation strategy chosen by their promoters and the technical, legal, financial and psychological obstacles to each of them, one may go so far as to predict that:

- the systems with a heavy emphasis on the data-processing component, while selling well for a limited period, will not corner a significant part of the market;

- provided nothing of a political nature intervenes, the Eurotra system will emerge about 1987 in a prototype state and will have cornered a number of important applications by 1990; its modularity will make it an adequate basis for the introduction of artificial intelligence factures;

- the systems having a good balance between the data-processing and the linguistic components will dominate the translation market at least until 1990; among these systems, circumstances have given Systran a start, and if the firms and institutions developing and using it make due allowance for its technical, financial and psychological constraints, then it will barely be possible for it to be overtaken.

Figure 1; THE TRANSLATION MARKET

1. INTERNATIONAL INSTITUTIONS

United Nations
UNESCO, IAEA, UNIDO, WHO, etc.

European Community
Commission, Council, Parliament, etc.

2. NATIONAL INSTITUTIONS

Multilingual states
Canada, Belgium, Switzerland

Foreign business, external trade

3. INDUSTRY AND COMMERCE

Promotion and publicity

Maintenance manuals, user instructions
Aviation, vehicles, data processing, electronics, pharmacy, etc.

4. INFORMATION INDUSTRY

Data bases accessible by networks

Multilingual information systems
TITUS, GERDAT, INSPEC, PASCALINE

5. PUBLISHING

Scientific publications

Literary publishing

Figure 2

TRANSLATORS' FEES

(1979)

Countries	FF / 100 words	
	Own language	Foreign lang.
U. S. A. A. T. A.	8½	
Free-lance	15-25	
United Kingdom Tech.	15-30	
Belgium	20-28	36-80
Canada	21-64	
Switzerland	33	
France	35-50	65-100
Sweden	70	
Germany	65-100	

Figure 3

SYSTEM CHARACTERISTICS

NAME AND ADDRESS	START R & D	START Oper.	LANGUAGES COVERED	DICTIONARIES VOLUME	COMPUTER TYPE
CULT University Hongkong	68	75	CH - E	10 000	ICL
PAHO Georgetown	75	80	SP - E	50 000	IBM
GETA University Grenoble	61	76	R - F, PT - F	16 000	IBM
LOGOS New Hamps. USA	64	71	E - VT, E - D	100 000	IBM
SUSY University Saarbrücken	67	74	R - D, D - ...	200 000	Telef.
SYSTRAN US Air Force	67	70	R - E	600 000	IBM
SYSTRAN WTC Canada	75	78	E - F, E - SP	100 000	IBM
SYSTRAN C.C.E.	75	81	E-F,I,D; F-E,D	100 000	IBM
SYSTRAN Germany	78	81	D - E	10 000	IBM
TAUM - METEO Canada	75	76	E - F	2 000	CYBER-76
TITUS Paris	72	73	E, F, D, SP	14 000	IBM
WEIDNER USA	78	79	E-F, D, SP, PT, F-E, D-E, SP-E	15 000	DIGITAL
Kyoto University		68	E - J	200 000	NEAC

BIBLIOGRAPHY

1. BOITET Christian: Tendances futures en traduction automatisée. COLING Congress, 1980.

2. BRUDERER Herbert: Handbook of machine translation and machine-aided translation. North Holland Publishing Co, 1977.

3. HUTCHINS John: Linguistic models in machine translation. University of East Anglia Papers in Linguistics no.9, 1979.

4. PERSCHKE Sergei and KING Margaret: EUROTRA and its objectives. Multilingua, vol.1, no.l, 1982.

5. PIGOTT Ian: Theoretical options and practical limitations of using semantics to solve problems of natural language analysis and machine translation. Informatics-5 Conference, ASLIB, 1978.

6. ROLLING Loll: La résurrection de la traduction automatique, une aubaine pour les fournisseurs et utilisateurs de l'information documentaire. Documentaliste, vol.18, *no.* 3, 1981.

7. ROLLING Loll: Automatische Uebersetzung heute und morgen. Umschau 1980 no. 11.

8. VAN SLYPE Georges: Revue critique des méthodes d'évaluation de la qualité de la traduction automatique. CEC report, 1979.

9. VAUQUOIS Bernard: L'évolution des logiciels et des modèles linguistiques pour la traduction automatique. T.A. Informations, vol.19, no.l, 1978.

10. WILKS Yorick: Four generations of machine translation research and prospects for the future.

11. Mieux traduire pour mieux communiquer. Etude prospective du marché de la traduction. Infotecture, Paris, 1981; Pergamon Press, London, 1982.

12. Practical experience of machine translation. ASLIB Conference proceedings, North Holland, 1982.