[From: Internat.conf. *Methodology & Techniques of Machine Translation*, British Computer Society, Cranfield, 13-15 February 1984]

Recent Canadian Experience in Machine Translation

Elliott Macklovitch

Translation Bureau, Government of Canada

The experience to be discussed is that of the Translation Bureau of the Government of Canada, which provides translation services to all federal departments and agencies, from Parliament to the National Film Board. Canada is officially bilingual. Under the Official Languages Act, this means that all services offered to the public by the federal government must be provided in both official languages, and that all federal public servants have the right to work in English or French. The progress of official bilingualism in Canada has resulted in a steadily increasing work load for the Translation Bureau. The Bureau translated 73 million words in 1968-69; in 1981-82, it translated 276 million words, and now employs about 1800 persons.

This enormous work load accounts for the Bureau's long-standing interest in machine and machine-aided translation. The Bureau began to subsidize research and development in MT more than ten years ago. It was the Translation Bureau that funded the development of the METEO system by TAUM, the University of Montreal research group in machine translation between 1974 and 1976; and the Bureau has been responsible for the operation of METEO since then. METEO translates weather bulletins from English to French at a rate of over 11,000 words a day, 365 days a year, and is recognized worldwide as one of the major successes, in the field of machine translation. (For a detailed description of METEO, see Chevalier et al, 1978.)

In 1976, the Translation Bureau commissioned TAUM to develop the prototype of a new system to translate the maintenance manuals of the CP-140, a coastal patrol aircraft that the department of National Defence had just ordered and was to receive in 1980. The technical documentation for that aircraft, according to unofficial sources amounted to some 90 million words. For anyone who has never seen maintenance manuals for equipment as sophisticated as an aircraft, it should be stressed that only highly specialized translators can translate them; indeed, the educated layman can barely understand them. It was estimated in 1976 that it would take four specialists ten years to complete the French translation of the technical documentation for the CP-140, by which time the aircraft would be obsolete. It was for this reason that the Translation Bureau looked to machine translation for help.

TAUM's initial estimate was that it would take about three years to develop the prototype of this new system, but this turned out to be overly optimistic. New high-level programming languages had to be designed and tested. Organizational problems arose in co-ordinating the activities of one team in Montreal and another in Ottawa. To complicate matters, TAUM lost its director early on in the project. Additional personnel had to be recruited and a great deal of time invested in training. Only then could the task of writing the extensive grammars and dictionaries required for the complicated sublanguage of aircraft maintenance manuals be undertaken in earnest. With the approval of the Translation Bureau, TAUM decided to base the construction of the new prototype on a 70,000 word corpus constituted of extracts from hydraulics manuals of several different aircraft.

In March 1979, TAUM gave a public demonstration of the new system, baptised TAUM-AVIATION. The demonstration itself was a success; under television lights, it was shown that AVIATION could translate the hydraulics manuals of the CP-140. However, it had become obvious that without large additional investments, AVIATION would not be able to translate the other manuals of the CP-140 and in any event, not in time for the expected delivery of the plane in 1980. Several months later, the federal Treasury Board approved a new contract between the Translation Bureau and the University of Montreal, on the condition that TAUM-AVIATION be subjected to independent evaluation and that a feasibility study be conducted on the system's extendibility.

The evaluation of TAUM-AVIATION (cf Gervais 1980) was conducted in March 1980 and had two main objectives: 1) to assess the system's linguistic performance, and 2) to analyse its development and operating costs. Samples taken from the hydraulics manuals of the CP-140, the Lockheed 1011 and a tank recovery vehicle were submitted to TAUM-AVIATION for translation and then revised by specialists from the Translation Bureau and an outside translation firm. The revisors were also asked to rate the intelligibility, fidelity and style of each sentence translated by the system, following a procedure used for an evaluation of the CEC's version of SYSTRAN in 1977. The same three texts were translated by qualified technical translators, two from the Bureau and one an outside freelancer, and these translations too were revised and rated. This allowed the evaluator to compare the ratings assigned to the machine translation and to the three human versions. As it turned out, the translations produced by TAUM-AVIATION received a satisfactory overall rating, about 80 percent of the rating assigned to the human versions. However, the system did not produce any translation for about one third of the sentences, titles or table entries that made up the test

corpus.

At first glance, the percentage of units for which TAUM-AVIATION produced no translation may appear extremely high. The principal evaluator, however, did not find this alarming:

> "Il faut se rappeler cependant que si le système TAUM-AVIATION ne produit aucune sortie dans certains cas, cela ne signifie pas qu'il en est incapable. Cela découle plutôt d'une décision des concepteurs du système. Ceux-ci ont jugé qu'il valait mieux ne rien produire plutôt que de risquer de produire du texte incompréhensible. Il est fort possible que dans un contexte d'exploitation il s'avère préférable de procéder autrement." (p59)

The risk of incomprehensible output is minimized in a second generation system like TAUM-AVIATION by basing the translation phase on a complete analysis of each source language input unit. Whenever a unit did not receive a complete analysis in TAUM-AVIATION, no translation was produced. As the evaluator points out, this is a perfectly reasonable strategy ... for a system under development. What happened in the evaluation was that many of the errors that prevented units from being analysed were caused by incomplete or incorrectly stated selectional restrictions in the analysis dictionary. In an operational context, it would not have been difficult to modify the system so that this sort of minor, local error did not always block the translation of an entire unit.

The proportion of untranslated units did prove to be significant, however, when it came to establishing the direct operating costs of producing translations using TAUM-AVIATION. Direct operating costs were calculated by adding the cost of putting the test corpus into machine readable form, the cost of the actual machine time required to translate the texts, and the cost of revision time. Revision time accounted for 37 percent of the total cost of a final version of the machine translation; and since the revisors gave a generally favourable rating to the translations that the system did produce, much of this revision cost must be attributed to the time they spent in translating the units for which the machine produced no output. The direct operating cost of producing a revised translation of the 14,000 word test corpus using TAUM-AVIATION turned out to be $0.183 per word; the cost of human translation and revision of the same corpus was $0.145 per word.[1]

Yet it was not this $0.038 a word difference that was most damaging for TAUM-AVlATION, for the evaluator noted that direct operating costs could reasonably be expected to

decrease once the system was implanted in an operational context. What did prove fatal were the system's indirect operating costs, and particularly the cost of adding new dictionary entries. Based on a rough extrapolation of the rhythm at which the dictionary teams were working at the time, the evaluator estimated that a person could index no more than 450 new terms a year, at a cost of about $49 per term. Moreover, to amortize the cost of maintaining an eight—man operating team, the system would have to translate between five and six million words a year. On the possibility of eventually operating TAUM-AVIATION cost effectively, the evaluator was thus led to the following conclusions :

> "Il est impossible d'affirmer, à la lumière de la présente évaluation, que l'utilisation du système TAUM-AVlATION peut, dans un avenir prévisible, devenir rentable, c'est à dire coûter moins cher que la traduction humaine, principalement à cause de ses coûts indirects et des conséquences qui en découlent." (p145)
>
> "La nécessite de trouver annuellement 5 à 6 millions de mots à traduire pour rentabiliser partiellement l'exploitation du système rend inopportune la poursuite du développement sans envisager d'autres applications." (pl49)

One of the objectives of the feasibility study (cf Gobeil 1981) requested by Treasury Board was precisely to determine whether TAUM-AVIATION could be extended to texts other than the hydraulics manuals for which it was designed. To that end, a 5800 word corpus taken from the electronics manuals of the CP-140 was submitted to the system, and the results compared with those from the March 1980 evaluation. This part of the study was not entirely conclusive, however; lack of time and resources prevented the translations produced from being revised and rated in the same rigorous manner as in the evaluation conducted by M. Gervais. Generally speaking, however, the results obtained on this electronics corpus were of comparable quality to those obtained on the hydraulics test the previous year. The performance of the system's grammars improved, but dictionary problems increased, as one would expect when texts in a new domain were being translated using entries conceived for hydraulics manuals.

Another of the objectives of the feasibility study was to inventory the types of texts translated by the federal government, classifying them according to their syntactic complexity and extent of vocabulary in order to identify those most amenable to machine translation. This inventory showed that the Bureau did not regularly translate five to six million words a year of maintenance manuals in

hydraulics or other related domains. Recall that this was the volume that TAUM-AVIATION would have to translate in order to be operated cost effectively.

Treasury Board had also requested that the feasibility study determine whether there were any other commercial MT systems which could help the Bureau meet its needs. A detailed questionnaire was therefore prepared and sent to twelve suppliers or potential suppliers of MT systems. Those that translated from English to French or from French to English were asked if they would be willing to submit their systems to a practical evaluation. The suppliers of three systems agreed: ALPS, SYSTRAN II and WEIDNER. Each was given the same 6300 word corpus to translate, comprised of extracts from trademark journals, staffing documents and the maintenance manuals of the CP-l40. The raw machine output was submitted to revisors who were asked to rank the different versions and to note the time it took them to produce an acceptable translation. A unit cost for translation and revision was then calculated for each system. The authors of the feasibility study found that the direct operating cost of producing a revised translation using each of the above—mentioned systems was lower than the cost of human translation and revision as determined on the 1980 evaluation, and thus lower than the unit cost of producing revised translations using TAUM-AVIATION. [2] The cost of making new dictionary entries was also found to be significantly lower than the $49 per entry estimated for TAUM-AVIATION. In terms of the quality of the translations produced, however, the results were far less satisfactory. In fact, the revisors refused to rank the translations in terms of technical accuracy, saying that they were all "pénible à reviser", or arduous to revise. In many cases, they did not modify the machine translation but found it easier to retranslate directly from the original. Moreover, none of the systems delivered the increase in translator productivity that their suppliers advertised. The authors of the feasibility study were thus unable to recommend that the Bureau purchase or make use of any of the three systems for its regular operations without further studies being conducted on much larger samples.

The feasibility study was completed in May 1981. In September of that year, TAUM was disbanded for want of funds. Former TAUMists, like myself, are often asked how the Translation Bureau could abandon machine translation in Canada. This is somewhat of a misconception, based on a misunderstanding of the relationship between TAUM and the Trans1ation Bureau. The contract that the Bureau signed with the University of Montreal in 1975 was for the development within three years, of a system that could eventually be used to translate the maintenance manuals of the CP-140. Following the presentation of that system to the Bureau in 1979, TAUM was granted an additional one-year

contract to continue the development and documentation of the system. On the basis of the evaluation conducted in 1980 and the feasibility study conducted in 1981, the Bureau decided, in September 1981, to abandon its objective of using TAUM-AVIATION to translate the manuals of the CP-140. From the point of view of the Canadian taxpayer, this decision was certainly justifiable. Between 1976 and 1980, the Bureau had invested over $2.7 million in MT. In return, it found itself the owner of a system whose cost-effectiveness had not been demonstrated. The Bureau therefore decided that it needed a period of reflection in order to draw the lessons of its recent involvement in MT. As for TAUM, it made the unfortunate error of putting all its eggs in the same basket. When the AVIATION contract with the Bureau ended, it found itself with no other source of funding.

None of this is intended to suggest that TAUM-AVIATION was a failure. On the contrary, from a scientific point of view, the project carried many of the principles of second generation MT to their logical conclusion. The result was an extremely sophisticated system that produced fully automatic, high quality translations of texts in a well-defined sublanguage. However, TAUM-AVIATION was not, in the fall of 1981, a system that was ready for large-scale operational production; nor, given the high cost of extending its dictionaries, was it a system that could easily become economically viable. It is important to ask why this is so. Why was dictionary construction so costly in TAUM-AVIATION? In particular, was this due to some fundamental flaw in TAUM's basic approach?

Under the second generation sublanguage approach employed at TAUM, a MT system is designed for a specific sublanguage, not for arbitrary texts from any domain. Such a system seeks to take advantage of each sublanguage's lexical, syntactic, semantic and textual restrictions in order to achieve maximum disambiguating power.[3] In AVIATION'S analysis dictionary, for example, the entries for predicate words defined co-occurrence restrictions on their arguments; these restrictions were stated in terms of semantic classes that were found to be particularly relevant for texts in hydraulics maintenance. At transfer, each potential context thus defined for a lexical unit could then be used to state the necessary translation tests.[4] Writing dictionary entries under this approach requires an extensive corpus that is representative of texts in the particular sublanguage, and a careful study of that corpus in order to first determine the relevant semantic classes and then establish each lexical unit's co-occurrence restrictions and translation tests. This is a time-consuming and therefore a costly process, but one that I would maintain is necessary if a system is to automatically produce high quality translations in a sublanguage as complex as aircraft

maintenance manuals. To take just one example, consider the following typical maintenance command:

(1) Remove fitting and drain plug.

This sentence is syntactically ambiguous, ie it could be parsed as a conjunction of imperatives, in which case drain is taken to be a verb, or as a single imperative with a conjoined object, in which case drain is taken to be a noun. The only way of blocking the former incorrect analysis in a second generation MT system is to specify in the source language dictionary entry for the verb drain that plugs are not drainable, although such objects as tanks and reservoirs are. In other words, a syntactic enumeration of permissible structures is often insufficient; the system must be provided with semantic features that distinguish between such objects as plugs and reservoirs, as well as with a specification of each predicate's complementation. This is a fairly fine semantic distinction, but one that would appear to be necessary for the automatic translation of hydraulics maintenance manuals.

A system like WEIDNER does not provide for selectional restrictions on predicate arguments. The only source language information given in its dictionary entries is the lexical unit's syntactic category. WEIDNER can distinguish homographs like drain in certain syntactic configurations, eg when the word is immediately preceded by an article; but not in a configuration like that in (1), where both a verb and a noun may occur after the conjunction (cf: Remove fitting and drain tank. Obviously, new dictionary entries will be relatively inexpensive in such a system. What will be expensive is revision. An interactive system like ALPS may, for sentences such as (1), interrupt the analysis process and ask a human operator to help it resolve the ambiguity. This too takes time, but one would normally expect to be compensated by less revision effort. Unfortunately, in the feasibility study and in a subsequent operational trial of ALPS at the Translation Bureau, this did not prove to be the case.[5] Moreover, the human operators tended to find it frustrating to be asked the same sorts of questions over and over again.

Now it may be objected that revision costs for TAUM-AVIATION were also found to be high in the 1980 evaluation. This is true, but not for the same reasons as the other systems tested in the feasibility study. The translations produced by TAUM-AVIATION were generally of good quality, and certainly revisable. The main problem was that the system failed to produce translations for too high a proportion of units, and these had to be translated by the revisor. The other systems tested in the feasibility study nearly always produced translations, but these were too often agrammatical and hence unrevisable. The ideal

solution, of course, would be not to have to sacrifice quality, but to increase the proportion of units translated by TAUM-AVIATION, by making the system more fail-safe, ie more resistant to minor errors in its dictionaries or grammars. Ways in which this could be done are discussed in Isabelle 1981. One of the suggestions made there is to build a sort of monitor into the analysis component which would be activated when no analysis was produced and which would retake the analysis of the current unit after temporarily neutralizing a series of semantic or syntactic restrictions. In this way, a translation would be produced for a much higher proportion of units.[6] Something as straightforward as linking TAUM-AVIATION to *a* word processor would also facilitate revision and lower overall translation costs.

Dictionary construction in such a system, however, will always be relatively costly, or at least costlier than in systems like ALPS or WEIDNER. This is not due to a flaw in TAUM's basic approach, but simply because TAUM aimed for a higher level of comprehension of the texts it automatically translated than did ALPS or WEIDNER. We saw this in the example discussed above. In TAUM-AVIATION, the pivotal representations that were the output of analysis and the input to transfer sought to identify the basic predicate-argument structure of each sentence and distinguish between the various meanings of words.[7] This was thought to be a minimum without which the system would not be able to consistently produce revisable translations. We also saw the kind of dictionary effort that was required to attain this level of comprehension. Another factor which influences the cost of dictionary construction in a second generation system like TAUM-AVIATION is the complexity of the sublanguage being translated. The more varied the range of structures in the sublanguage, the longer it takes to describe lexical co-occurrence restrictions. The larger the vocabulary of the sublanguage, the more homography it tends to display; these homographs must be distinguished in the system's dictionaries if they are not to reappear as problems at revision. Indeed, one of the principal lessons to be drawn from the experience of TAUM-AVIATION is that certain sublanguages are so complex that it is extremely difficult to attain the level of comprehension necessary for their automatic translation by means of second generation technology.

Since the close of the AVIATION project in 1981, the Translation Bureau has not lost interest in machine translation. With a workload approaching 300 million words a year, it cannot afford to do so. Moreover, the Bureau still believes in the utility of the second generation sublanguage approach. The problem, as the Bureau now sees it, is to judiciously select the sublanguage to which a second generation system is to be applied. Weather

bulletins combine the ideal characteristics for machine translation: high volume coupled with restricted syntax and vocabulary. Aircraft maintenance manuals are on the opposite end of the complexity scale; in fact, they may be too complex for second generation systems. But what about the many sublanguages in between? Which of these might be amenable to second generation technology and what other factors, aside from a suitably defined notion of linguistic complexity, need to be considered in order to guarantee success? The Bureau has been subsidizing research into these questions, conducted by Professor Richard Kittredge of the University of Montreal. Professor Kittredge examined samples of 17 varieties of texts which seemed likely or desirable candidates for machine-aided translation within the Bureau, and identified a number of sublanguages that could be handled using current technology, (cf Kittredge 1983) The Bureau will be conducting a feasibility study of one such application this year, and if the results are satisfactory, the development of the first of a series of small-scale MT systems could begin in 1984—85. Other MT-related projects for the current year include the transfer of METEO onto a micro-computer and the introduction of TERMIUM III, the new version of the government's computerized terminology bank.

The Bureau now has a better understanding of the limits of second generation MT systems, and recognizes that fundamental research will be required before development can begin on the next generation of systems. To help orient the direction of such research, the Translation Bureau and the Department of Communications recently commissioned a 1arge-scale study into the current state of natural language processing and artificial intelligence, with special emphasis on applications to machine translation and other related fields. This study is expected to contain recommandations on the manner in which the government can best co-operate with universities and private enterprise in order to reactivate MT in Canada. All those interested in machine translation in the country are anxiously awaiting the publication of the final report.

Notes

1- It should noted, however, that overall, the revised human version took two and a half times longer to produce than the revised machine version. The final revised versions were judged comparable in quality by a number of potential users, including Air Canada and the Department of National Defence.

2- Over the three types of texts, WEIDNER was found to be the least costly at $0.089 per word; next was ALPS at $0.113; and finally, SYSTRAN at $0.143 per word.

3- For a detailed description of the sublanguage of aircraft maintenance manuals and a discussion of the relevance of sublanguages for automatic translation, see Lehrberger 1978.

4- For a discussion of the work of the translator in a second generation system like TAUM-AVIATION, see Chevalier et al 1981.

5- A report on the six-month operational trial of ALPS at the Bureau should be available shortly.

6- The units thus produced would of course be flagged for special attention by the revisor.

7- See Lehrberger 1981 for a discussion of TAUM's linguistic model.

References

Chevalier, M. et al (1973) TAUM-METEO - Description du Système, Groupe TAUM» Université de Montréal.

Chevalier, M et al (1981) "La traductologie appliqué à la traduction automatique", META vol 26, 1. Les Presses de l'Université de Montréal.

Gervais, A (1980) Evaluation du système pilote de traduction automatique TAUM-AVIATION, Direction générale de la planification, de l'évaluation et de la vérification, Bureau des Traductions, Secretariat d'Etat, Ottawa.

Gobeil, F (1981) Etude de faisabilité de la traduction automatique, Direction générale de la planification, de la gestion et de la technologie, Bureau des Traductions, Secrétariat d'État, Ottawa.

Isabelle, P. (1981) "Perspectives d'avenir du groupe TAUM et du Système TAUM-AVIATION", Groupe TAUM, Université de Montréal.

Kitteridge, R. (1983) "Sublanguage-Specific Computer Aids to Translation : A Survey of the Most Promising Application Areas", Planning, Management and Technology Branch, Translation Bureau, Secretary of State, Ottawa.

Lehrberger, J (1978) "Automatic Translation and the Concept of Sublanguage", in R Kitteridge and J Lehrberger (eds) Sublanguage: Studies of Language in Restricted Semantic Domains, de Gruyter.

Lehrberger, J (1981) "The Linguistic Model: General Aspects", in L Bourbeau(ed) Linguistic Documentation of the Computerized Translation Chain of the TAUM-AVIATION System, Groupe TAUM, Université de Montréal.