PRACTICAL EXPERIENCE OF MACHINE TRANSLATION
V. Lawson (ed.)
North-Holland Publishing Company / © ASLIB, 1982

THE METEO SYSTEM

Benoît Thouin

Computational Linguistics Consultants Ltd.
P.O. Box 339, Sta. DeLorimier
Montreal (Quebec)
CANADA H2H 2N7

METEO is a machine translation system that has been in successful operation since May 1977. It is of particular interest in that translators were intimately involved in its development, and because of its degree of integration within a large communications network. And there are improvements in the works ...

INTRODUCTION

Picture Dorval, a suburb west of Montreal. Beside the Transcanadian Highway a building about five stories high carries the somewhat pretentious name "West Isle Tower Building". Its immediate neighbours are a wine depot, the headquarters of the Boy Scouts of Canada and one of the numerous buildings that belong to a photocopy multinational.

This building houses the Canadian Meteorological Center (C.M.C.), a few dozen people whose activities revolve around one of the most powerful computers in Canada. This computer's principal function is to compile complex meteorological data coming in from all over the country and then draw the weather charts and deduct from them a forecast for the next few hours by the application of a mathematical model. You will, no doubt, forgive me for not detailing the mathematics here.

Like any civil servant, the CMC computer has the right to a rest period to collect itself (about an hour a day for maintenance) and to a coffee break. The computer must enjoy strong drink because it uses its break, that is, about fifteen minutes of CPU time a day, to translate part of Environment Canada's Atmospheric Environment Service's forecast from English to French.

Having defined the METEO system as the CMC computer's cup of tea, now we can tackle the meaty part of the subject - a description of the system's different aspects. After some historical background we'll deal with telecommunications, the computational and linguistic make-up of the system and then it's economic and human aspects, before going on to conclude by suggesting a possible future for this constantly evolving system.

HISTORICAL BACKGROUND

On May the 15th, 1975, the Translation Bureau of the Canadian Government's Secretary of State Department entrusted the TAUM group (the University of Montreal's Automatic Translation Research Team) with the development of an automatic translation system for the weather forecasts. This contract to develop the system came after ten years of research financed by the National Research Council and the Secretary of State.

A year later, the University of Montreal delivered the Secretary of State a first version of the system, already christened METEO, which demonstrated that automatic translation in this area was feasible. It took another year to develop an operational version of the system, set it up at the CMC computer center and install a first terminal at the Translation Bureau's Meteorological Sub-section. The Canadian Government called on Computational Linguistics Consultants Ltd. to finalize the operational version of the system and, subsequently, to document it and keep it up to date [1].

The system began being used on May 25th 1977, initially translating only those messages from the regional offices in Halifax and Toronto. Its use was extended to the whole country in June 1978. Since May 25th 1977, the METEO system has been in regular use twenty-four hours a day, except for necessary interruptions for maintenance work on the computer equipment.

TELECOMMUNICATIONS

The METEO system is fully integrated into the Canadian network for the transmission of meteorological forecasts. The various regional offices (there are eight of them) issue the forecast bulletins that are their responsibility. Each office types these messages on a terminal that is connected to CN/CP's nationwide telecommunications network. All the messages sent over this network are channelled to a central computer in Toronto where they are sorted and sent to their destination (or destinations).

I should clarify, at this point, that there are messages which the METEO system can translate (i.e. regional forecasts, maritime forecasts and forecasts aimed specifically at farmers and boaters) and others it was not intended to process (such as general synopses, technical synopses, ice warnings).

The CN/CP computer in Toronto sends many messages to the CMC. Besides forecasts from the different regional offices, these include data collected at Environment Canada's numerous observation stations. The METEO system selects from this data the messages it can process, translates them and prints the sentences from these messages that it could not analyze on a terminal for the attention of a human translator. The translator finishes up the translation done by the system which then sends the French version to Toronto. From there it is broadcast to the meteorological network's French service subscribers.

All the forecast bulletins are also sent to other terminals, near the first, at the Translation Bureau's Meteorological Sub-section. The translators "let through" bulletins that were submitted to the system and only translate the others - the French versions of which are sent off to Toronto and then, as in the first case, to the various subscribers.

Whether translated mechanically or manually, from the moment the forecast is typed on the terminal at its weather office of origin, it is available in English and later in French, but only in an electronic form. It does not appear on paper until the very end of the process, on the telexes of the network's subscribers.

COMPUTATIONAL ASPECTS

a. The hardware

The CMC's main computer is a Control Data CDC 7600 which is mainly used to carry out simulations of atmospheric conditions. An automatic job submission system interrupts this program every five minutes and loads the METEO system so that it will translate any messages that might have arrived since the last interruption.

A CYBER 71 serves as a front-end processor and assures, in particular, the interface between the system and the translators' terminal, a DecScope VT52 from

Digital Equipment Corporation. Besides carrying out the user-machine interaction programs and managing the data base of translated messages, it is the CYBER 71 that submits to the CDC 7600 the control instructions to load the translation system nucleus. A Data General NOVA computer manages the reception and expedition of messages over the CN/CP telecommunications network. The terminals used for human translation of the messages not submitted to METEO are two HP-2621P from Hewlett-Packard.

b. The software

The METEO system's software is comprised of four sub-systems.

The translation system properly speaking is the one released every ten minutes on the CDC 7600 and this sub-system is made up of the following stages:

- Decoding and selection of those messages that the system *is* able to translate. Such messages are identified by a code from a predetermined list.

- Recognition of the sections that normally comprise a bulletin: heading code, title of forecast, list of localities or regions and the forecasts themselves. Then conversion of the messages into a series of graphs compatible with the translation software described below.

- The actual translation. This is effected by a specialized software we call "Q-System". This software, developed at the University of Montreal in 1969-70, provides a general framework for the transformation of graphs whose vertices carry trees [2]. A graph corresponds to one unit of translation i.e. one title or sentence in METEO's case.

  This software comes with a very high level language for describing transformations to be carried out. This meta-language allows the author of the grammar to describe the transformations to be carried out on the graph that represents the source sentence in order to effect the analysis, lexical and structural transfer and generation which will produce a graph of the sentence in the target language.

- The re-assembling, in French, of the messages and of their different sections from the graphs resulting from the translation program. These messages are then directed to the translator's terminal and placed in a queue.

The second sub-system permits the translator to complete the translations done by the first sub-system and to send them over the telecommunications network. The translator's role is not to postedit the translations produced by the system but merely to translate sentences rejected by the system on account of spelling or typing errors, faulty transmission, failure to respect presentation guidelines or weaknesses in the machine translation dictionary or grammar.

The third sub-system manages a data base of translated weather bulletins so that if a translation is lost because of faulty transmission the translator can, on request, re-issue it without having to do it again.

The last sub-system is somewhat independent. Its purpose is to check and apply the grammar and dictionary updating transactions, and it is never used by the translators themselves.

LINGUISTIC ASPECTS

The METEO system demonstrates the two principal characteristics of a second generation machine translation system. First, it is based on a linguistic model

of the special purpose language (or sub-language) *of* weather forecasting. Secondly, its software allows for the separation of the algorithm that does the transformation (which has been programmed once and for all) and the linguistic data (dictionary and grammar).

Two other important characteristics of the METEO system from a linguistic standpoint deserve to be pointed out:

- As the texts submitted to the system are formulated in telegraphic style, the parser cannot count on all the syntactic signals of normal English. As a result, while the surface analysis is syntax oriented, the deep structure analysis is actually semantics oriented. Analysis of most of the units results in a standardized structure which consists of a weather condition, or a weather condition followed by an expression indicating change, followed by another weather condition.

- As soon as a word is missing from the dictionary (which may be due to a spelling mistake or to faulty transmission) or whenever a construction is ambiguous or surpasses the linguistic model foreseen in the grammar, the system rejects the unit (sentence) in question and submits it for human translation. It makes no attempt to propose partial translation of the unit or to submit several possible translations to the human translator.

The linguistic processing is done in the following stages:

- the rejection of units containing errors due to transmission problems

- translation of headings (the CN/CP codes that identify a message)

- expansion of abbreviations

- identification of idiomatic expressions

- processing of place names

- general dictionary: this dictionary gives the syntactic and semantic characteristics and translations of morphemes

- processing of dates

- processing of times (conversion of AM/PM to twenty-four hour system)

- processing of temperatures

- analysis of time references

- analysis of locatives

- analysis of noun phrases: adverbs, conjunctions of adjectives, comparatives, determinative objects

- building of structures representing a weather condition

- building of the standard structures of the units

- rejection of partially analyzed units

- syntactic generation: breakdown of noun phrases

- morphological generation: agreement of adjectives, elision, contraction

- stylistic rules

MAN'S PLACE IN THE SYSTEM

The METEO system is (regrettably) unique in that translators played a crucial role in all stages of the design, development, use and refinement of the system [3].

Everything began when a translator from the Meteorological Sub-section suggested that forecasts might lend themselves to machine translation on account of their repetitive character and limited scope.

Then, translators participated in the choice of corpus, contributed to the elaboration of a dictionary and followed closely each stage in the development of the system, including those stages in which they were not directly involved.

During the first three years of operation, these translators played an essential role in the refinement of the system, examining printouts of translated forecasts, detecting the system's weaknesses and requesting or suggesting improvements to be made. They also acquired a good understanding of the system and made felt their points of view on the practical aspects of its use.

In particular, they were at the origin of a change in the type of terminal and instigated two types of interaction with the system: the "chatty" mode where the system is very explicit, recommended for new users and, for the initiated, the "silent" mode where messages are reduced to a strict minimum. (Between old friends a nod of the head is enough.)

The METEO project was also, and perhaps this is one of the reasons for its success, the fruit of collaboration between Government, University and the private sector.

Today we can make two interesting observations on the human side of things:

- The fastidious nature of the texts put off translators to the extent that few of them would remain more than six months with the Meteorological Sub-section. Perhaps the fact of working in front of a silent screen rather than with the murmur of typewriters had something to do with it? Or the fact that only the more difficult texts, or those that the system had not been able to analyze, were left to translate? Now, all the translators except one in this unit have been there at least two or three years and all are happy in their working environment.

- A significant proportion of the texts rejected by the system are so treated because the originators do not respect writing and format standards for weather bulletins. Incidentally, none of these rules were established for machine translation, on the contrary, the system was built to already existing guidelines. This suggests that the formulation of yet stricter rules tailored to machine translation would have been a mistake. In any case, although the present rules impose the usage of properly meteorological terms, they leave certain areas of freedom for the writers, for example, in the choice of adjectives... perhaps to liberate their frustrated creativity.

ECONOMIC CONSIDERATIONS

Let's begin with a few statistics. Out of 18,000 words that the Meteorological Sub-section has to translate every day, about 11,000 are submitted to the METEO system. It translates about eighty percent without human intervention i.e. half of the total output of the Meteorological Sub-section. METEO translates between three, and three and a half million words a year.

The existence of the system has allowed the French translation service for the weather bulletins to be extended across Canada without any increase in staff. In any case, personnel were quite hard to find - given how little interest translators showed in this kind of repetitive work. The new working framework created by this system contributed to improving the working lives of the translators, which represents a sure gain even though it is difficult to measure.

An estimate of the computing costs of the system established them at about $0.03 (or $0.04 at the most) Canadian dollars per word. It is difficult to evaluate the cost more exactly. Should one, for example, use the commercial rate for fifteen CPU minutes a day (the coffee break!) of the CDC 7600 computer - which would have gone unused in any case? Or should one consider the METEO system alone responsible for the increase in front-end processing capacity when other users profit from it now?

CONCLUSION: THE SYSTEM'S FUTURE

It is these economic factors which will propel the most likely evolution of the system. In fact, even while the computing costs now only represent a fraction of the cost of human translation, a recent feasibility study concluded that it would be even more economic to install a micro-computer dedicated to the METEO system. The Canadian Government has not yet taken *a* firm stand on this but it remains an interesting possibility [4].

One hears talk, too, of integrating those forecasts from the Meteorological Services of the North-Eastern United States that might be relevant to the major French speaking areas of Canada. A little matter of two hundred and thirty thousand words a day - or about thirteen times the present load.

The linguistic model has reached some kind of ceiling. It does not seem economic to invest a lot of extra effort in dealing with rarer cases or errors in the original text. The next step in this area would be to realize a qualitative leap and treat messages that have not yet been submitted to the system.

One of the most desirable improvements would be to allow translators to reread sentences "successfully" translated by the system, and correct them where needed. It seems unwise, to say the least, that man relinquish the control he should (and hopefully still does) have over the machine.

The ideal?... well, it would be the "summit" of artificial intelligence. At any rate it is a fine research subject: given that we have a weather service that delivers forecasts that are right eighty percent of the time, and a translation system that correctly translates eighty percent of the forecasts, maybe one could arrange that the "weaknesses" of the translation system would correct wrong forecasts (and only those) ?!!! That would certainly be a valuable contribution machine translation might make to our society.

REFERENCES

[1] Gobeil, F., La traduction automatique au Canada, L'Actualité terminologique, vol. 14, no 5 (May 1981).

[2] Colmerauer, A., Les Systèmes-Q, ou un formalisme pour analyser et synthétiser des phrases sur ordinateur, TAUM, Université de Montréal (Sept. 1970).

[3] Thouin, B., Pour les traducteurs, par des traducteurs, in: Proceedings, IX^{e} Congrès mondial de la F.I.T. (Warsaw, 1981).

[4] Chandioux, J. and Guéraud, M.F., METEO, un système a l'épreuve du temps, META, vol. 26, no 1 (March 1981).