Invited Talk

"SPECIALIZED" MACHINE TRANSLATION

John Chandioux and Annette Grimaila

John Chandioux Consultants Inc.
1253 McGill College Avenue, Suite 450
Montreal, QC H3B 2Y5 Canada

Abstract

The Centennial Olympic Games in Atlanta gave our company the opportunity to demonstrate our expertise and to provide a made-to-measure machine translation solution to the U.S. National Weather Service whose challenge was to ensure real-time access to French and English weather information for the duration of the Games. We firmly believe this experience has validated the dedicated approach we have been promoting for twenty years.

1. Definitions

For us at John Chandioux Consultants Inc., the term "specialized" best describes our tools, principally our programming language GramR®, which we use for all of our linguistic developments and which we license to other developers (like Claude Bédard of Traductix Inc., who developed the ATAO® system with GramR) as well as to certain clients who use it to update the systems we developed for them.

In our marketing efforts, we speak of "made-to-measure" solutions ("made in accordance with measurements taken", Oxford Dictionary, Third Edition) but the term also describes the methods we have perfected over the years to arrive at the results our clients require.

Finally, we refer to the different systems we have developed as "dedicated" MT systems ("having a single-minded loyalty", idem.) Indeed, our systems are each dedicated to their users in a special way : METEO® is used for Canadian weather forecasting, METEO 96® was designed for weather forecasting in the southeastern United States during the summer, LEXIUM® deals only with Canadian trademark descriptions, the General TAO System ensures one insurance company's documents are available in both official Canadian languages, and so on.

2. From METEO to METEO 96

METEO started in production in March 1977 and has been continually updated, improved, transported to new computing environments since then. Today, it is practically the only equipment and software used by the employees of the Translation Bureau who are responsible for the post-editing of METEO's output and for the translation of non-standard weather reports for Environment Canada.

2.1 The need

According to the World Meteorological Organization, "National Meteorological and Hydrological Services have been providing increasing levels of support to the modern Olympic Games since 1896" (Weather and Sports, publication no. 835, WMO, Geneva, 1996). In fact, "for the Centennial Summer Olympic Games in Atlanta in 1996, 30 US National Weather Service (NWS) forecasters" were "joined by four meteorologists from the Atmospheric Environment Service, Canada, and two meteorologists from the Bureau of Meteorology, Australia" (idem).

The Olympic Charter in 1896 clearly identified French and English as the two official languages of the modern Olympics and, perhaps in honour of the Olympics' first century, all of the information to be disseminated by the Atlanta Committee for the Olympic Games, including weather information, had to be bilingual. This was the NWS's challenge: how to provide real-time English and French forecasts, warnings, watches and statements for the duration of the Centennial Olympic Games.

2.2 The recommendation

It was due to one of the four Canadian meteorologist, Mario Gaudette, working in the Quebec Meteorological Centre located in Ville Saint-Laurent and well aware of METEO and the services we have been rendering indirectly to Environment Canada for close to twenty years, that our company was asked to look at the problem and propose a solution.

2.3 The analysis

Our first contacts with Lans P. Rothfusz, Meteorologist-In-Charge of the NWS Olympic Weather Support Office in Peachtree City, Georgia took place in October 1995, a little over eight months before the scheduled start of the Games.

A preliminary analysis of a few sample texts received on October 31st looked promising but there were major differences between the American and the Canadian styles of writing and, since time was of the essence, we (JC and AG) travelled to Georgia in early

December 1995 to pursue our discussions with the NWS and the Translation Services of ACOG, to analyze their needs in greater detail, and to determine the basic technical set-up upon which we could base our proposal.

2.4 The proposal

After the receipt in mid-December of more sample texts, we were confident that we were the right people to solve the NWS's Olympic challenge. Our proposal was despatched on January 2, 1996 and approval was confirmed on February 16th, less than five months before the start of the pre-Game period on July 1st which was the official deadline.

3. The project

3.1 What needed to be done

The METEO 96 system would need to be integrated to both the NWS systems, mainly HP workstations, and to Info 96, the system developed by IBM to collate all data for use by administrators, athletes, trainers and the media. It would have to translate the three standard forecasts (today's, tomorrow's and the 3-5 day forecasts) issued every day for each of the eighteen Olympic Venues as well as any warning, watch or statement for potentially dangerous weather conditions. It would need to be as accurate as possible since it was doubtful that a specialized translator could be delegated to look after post-editing. It would have to be fast: a maximum delay of 10 minutes between English and French texts.

The linguistic model would have to recognize and translate place names where required, e.g. the "Olympic Ring" became "le cercle olympique" but "Columbus Golden Park" would remain the same. Repetitive messages, known as Calls To Action, would have to be pre-translated and approved since they gave people specific advice in case of dangerous weather.

In addition to the forecasts prepared by the Peachtree City, Georgia office, five other NWS offices were involved: Savannah, Georgia; Sterling, Virginia; Birmingham, Alabama; Orlando and Miami, Florida. These offices would provide information for the competitions held in their areas.

Because of the international origins of the participants, all weather data had to be provided in both US and metric measurements. Peachtree City generated this information for their own forecasts, but some of the satellite offices were not equipped to do so, nor to provide their texts in upper and lower case. We therefore proposed and provided a conversion module, through which all the satellite products were channelled, which did the case conversion as well as the metric conversions. This conversion module served two purposes since it allowed us to pre-edit and to standardize some portions of the texts which simplified the translation process.

3.2 How it would be used

METEO 96 would be installed, at our recommendation, on the same PC running under OS/2 that would be used as a weather data repository for Info 96. The Canadian meteorologists would work both as forecasters and as translation revisers. The system would have to be as foolproof as possible since there was no time to be lost with system crashes or other failures. One series of forecasts would be translated around 8:00 p.m. but would be picked up by ACOG only at midnight for display on the Olympic systems for the next day's activities.

Because of the time constraints now upon us, we decided to develop the user interface with Visual Basic 3.0 under Windows 3.1 since it is one of the standard programming tools with which we are most familiar thus allowing us a degree of flexibility and the required speed of development in this particular instance.

3.3 What we asked for, what we got and why

We requested sample texts from each of the forecast offices and especially of the non-standard forecasts. The standard forecasts were produced with the help of a text-generation system, known as ICWF, and were of such good and predictable quality that little if any revision would need to be done.

Since the satellites would only come on line on July 1st, we asked for copies of the forecasts produced in the summer of 1995 during a sporting competition which served as a trial period for the Olympic Weather Support Office.

We asked for texts written by the same people who would be assigned to the job during the Games, since each person has a different writing style. Unfortunately, the Olympic Weather Support Team would only report for duty on or about June 27th.

We did receive warnings, watches and statements from the previous year, but their format was not that required for Info 96. They were of some use, but not sufficiently so. During the entire period of our development, the NWS Olympic Systems Administrator, L. Clark Safford, was of great assistance to us, programming special data captures to allow us to view the satellite products and generally following all of our developments and providing advice and information on their UNIX-based systems.

3.4 How we managed

The first version of METEO 96 was sent to the NWS on March 20, 1996. This served to test the basic user interface and its compatibility with OS/2 and the UNIX systems. From then through July 19th, we worked on improving the overall design, completing the linguistic models and getting everything to work together.

By early June, we knew we had to go back to Georgia to test the system ourselves and to train the Canadian meteorologists in its use. We arrived on June 29th and, as far as we could tell since Info 96 was not ready for us, we were operational on July 1st as required by our contract.

4. The results

METEO 96's results were 16 days of near perfect (93.2%) weather translations with real-time access to the target language by Info 96. A few hours before the Opening Ceremonies, we downloaded the first of nine versions of the linguistic model which we were constantly updating on the basis of the previous days' weather reports.

In total, between July 19th and August 4th, METEO 96 processed 1,708 reports (an average of 106.75 reports per day) representing 306,000 words (an average of 19,125 words per day). The post-editing was done by the three bilingual Canadian meteorologists working on two eight-hour shifts per day, the first from 5:00 a.m. to 1:00 p.m., the second from 1:00 p.m. to 9:00 p.m. Each day, these two meteorologists were as productive as nine professional translators who each normally produce approximately 2,000 words per day.

There were no system crashes caused by METEO 96. Only one of two service calls during the Games directly concerned METEO 96 (a problem with the header of the French files), but this was due to delayed testing by the Info 96 people and it was corrected and downloaded within 6 hours.

5. The secrets of our success

5.1 We know the limits

There is an immense diversity and a constant evolution in our human or natural languages. We all know there are semantic differences between languages and we can but observe and wonder at their

origins. For example, in French we differentiate between a river/*une rivière* and a river that runs to the sea/*un fleuve.* Why is the difference important in one language and not in the other?

Within a single language, there are different connotations to the same words which we think we can easily decipher as human beings but which are impossible to program, such as 'Time flies like an arrow." (Are we timing flies or is time flying?)

Add to that, the innumerable changes a language is subjected to which are impossible to predict, not to mention the simple spelling errors that creep in when texts are written under stress. At one point, during the Olympic project, we realized one of the English-speaking meteorologists in one of the satellites had taken to writing "tommorrow", not to mention the gradual reduction of Georgia Horse Park to Horse Park to GHP over a matter of a few days.

The computers we use today are vastly more powerful than what we knew just a few years ago. We talk of super-computers with capacities of millions of instructions per second, or parallel processors for animation and modelling. Even these, however, will not be able to deal with all of the variables of our natural languages. Some of the limits are in the hardware, some in the software, but most of them in the very ambiguous nature of our languages. Star Trek's universal translator is not for tomorrow.

5.2 Advantages of dedicated MT

There are two basic advantages of a dedicated MT system: the "fit" is good and the "quality" is very high, even if it is not perfect. Although expensive at first glance, its cost effectiveness improves with use and productivity can increase as much as 800%.

5.3 Disadvantages of dedicated MT

A dedicated MT system, especially if a made-to-measure approach is used to develop it, is not "re-usable" or "transportable". It is not a high volume product and therefore relatively expensive both for the developer and for the client. Each problem must be analyzed in depth and a specific strategy designed to resolve it.

5.4 The constraints of a dedicated approach

For the developer, the constraints of such a dedicated approach can be used to define a structured approach.

> Always start with an in-depth analysis of the problem. Get sample texts, more sample texts, and even more sample texts. Get real texts written by the people who will be providing the source texts for the system.
>
> Look for every method of accelerating development whether it be the user-interface, the dictionaries, the grammars.
>
> Always use a modular approach. If a system cannot be re-used itself, perhaps parts of it can.

Although a dedicated approach is extremely efficient if the problem meets the basic criteria for such a made-to-measure solution (see *Machine Translation, Dream or Reality?* Association des conseils en gestion linguistique, Montréal, 1991), it is not appropriate to most translation work

where a much more interactive approach is required. For such a user-directed approach, we suggest a look at Claude Bédard's strategy in ATAO, the Atelier de traduction assistée par ordinateur, which is now available as two separate Windows modules: *Logiterm* and *Logitrans.*

6. Conclusion

We are proud of the results we achieved for the Atlanta Olympics. Although there were similarities between the Canadian forecasts we are used to and the American ones, there were sufficient differences to preclude a simple transfer of the linguistic model. Indeed, we know for a fact that a machine-translation system must be "organic" : it must grow and evolve with the needs of its users. Moreover, this growth and evolution requires continual observation and analysis to ensure it maintains its high level of quality.

The Olympic experience was a perfect opportunity for the U.S. National Weather Service to test new technologies and to improve their work methods and the essential services they and other National Meteorological Services render to the general public. We are extremely happy that we made the team - the Olympic Weather Support Team of 1996.