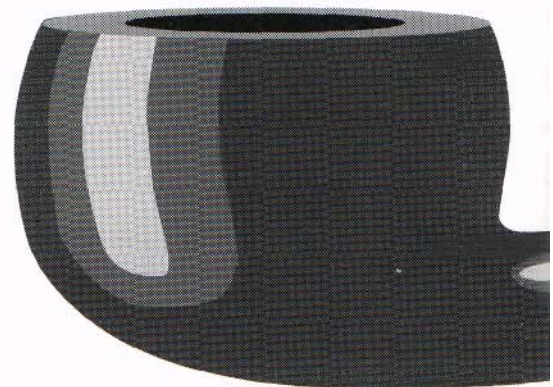


Calliope and other



*Ceci n'est
une tra
automat*

France began a machine translation project in 1983 to commercialize the Ariane translation algorithms. Five years and 65 million francs later, Calliope still hasn't played a tune. Here's why.

By Andrew Joscelyne

National machine translation programs hardly lie thick on the ground. And with most MT folk riveted to the Japanese MITI's fifth generation megaproject, it's all too easy to forget that as early as 1983 the French government sank some FF32.5m (US\$5m) into a National Computer-Assisted Translation Project (NCATP) aimed at making the already existing Ariane algorithms pay for themselves. A group of industrial operators put up the other half of the FF65m startup fund, and a long list of acronyms standing for various dusty-corridor government agencies constituted the project coordination committee. Since in those days France was cruising along in socialist gear, the whole operation was managed by the state-run Agence de l'Informatique (ADI), set up to find industrial uses for government infotech research and expertise.

Someone with a nose for political survival programmed the project's first three phases — specification, development and the launch of a prototype — to be ready by early 1986, just as the government changed hands and the economic liberals took over. ADI was finally dismantled in February 1987. Now uncertain of its future, NCATP still muddles through on borrowed time. Asked about what they expect next, staff shrug their shoulders and don't give much away: "The Ministry of Industry will coordinate any future investment in the product." The industrial partners concerned appear to be waiting to see what next year's presidential elections will bring.

In the meantime, the venture's stated aim has been adjusted to "the creation of a brand-new translation environment for technical documents." NCATP staff recently rechristened the project Calliope after the lush-voiced Muse of Eloquence — or perhaps after the fairground steam-whistle organ. A not wholly operational prototype was presented at Expolangues in 1986. Early in 1987, Odile Vaissade, chief MT engineer at SG2 — a major computer products company and the project's "general foreman" — gave a general update at the Paris-held European Translation Days, but concrete evidence of the system's precise current capabilities were not forthcoming.

Calliope consists of three related products: a translation program proper (Calliope-Système), a "linguistic module" handling the chosen translation domain (of which one has been fully developed: Calliope-Aéro, for aeronautics), and a translator's workstation allowing dialogue with translation software and access to various terminology banks (Calliope-Révision).

The central translation module — Calliope-Système — controls the three-phase translation process, using an expert system backed by a language-independent inference engine to produce target language equivalents of the source text. The program has been designed and written by the Automatic Translation Study Group (GETA) of Grenoble

University, led initially by the late Bernard Vauquois and latterly by Prof. Christian Boitet. The group is at present producing a transportable version of the same software, converting its 100,000 odd lines of machine code into LISP. This will make Ariane compatible with market hardware — an essential move if Calliope is ever to pay its way. Owing to staff shortages, this time-consuming chore was not finished by the end of 1986 as originally planned. Completion is now promised within 1987.

In November 1984, Grenoble language engineers founded a company called B'VITAL, now an NCATP private-sector backer. The company is responsible for developing the machine-readable dictionaries and "static" grammars used by the central translation algorithm in its trial-and-error analysis of source text. As Ariane's layout (see sidebar) shows, the rule component has been elaborated at the expense of mere look-up word-equivalence. An interesting result is the need for Artificial Intelligence tools to handle the large number of relational possibilities caused by the augmented body of rules. In fact, GETA's translation motor foreshadows many of the fifth generation methods of Japanese CAT programs.

Since the project's outset the favored translation domain has been aeronautics, not surprising in view of the vast quantity of instructions and technical specifications accompanying exports like Mirage warplanes. Calliope-Aéro is a "linguistic module" designed to represent "Aeronautspeak" style 'n' vocabularies in a knowledgebase before text is dispatched to the French-to-English translation motor. Early forecasts suggested this module would be available by late 1987. But it's still being played around with by

translators at Sonavision, another private-sector backer and itself producer of the aeronautics terminology base. A second module, Calliope-Informatique, first planned to come out this year but still on the chocks like most Calliope products, will handle style features in an English-to-French unit for bulk translation of dataprocessing manuals.

Calliope-Révision, billed as an integrated workstation, was developed by engineers at SG2, a company with an abundance of contracts for government and large banks to its name. Organized round a Bull Questar 400 micro, accessing an IBM with a VM/CMS operating system where the translating takes place, the workstation should eventually in-

"As it stands, Calliope can't be put into production," says Sonavision's head of translation. "What's been achieved so far is a laboratory prototype with promising results." Promises, promises. Potential customers can't even get a look at a functioning workstation.

er pipedreams

st pas
duction
tique

clude access to image and term databases as well as input/output peripherals – a scanner, an OCR and a printer – for desktop publishing. Bull hardware was forced on the project because it was French, though Calliope staff would have preferred a more widely used terminal. In fact, there seems little doubt that the proposed workstation is the project's Achilles heel and will never actually be industrialized with

current hardware.

Calliope's final raw translation printout comes in the usual MT form: parallel columns of source and target text, including messages showing syntactic bafflement and semantic disgust – but the text processor promised for editing it isn't ready yet. SG2's Odile Vaissade admits the present impracticability of integrating Calliope into an industrial documentation system, given both its nonavailability on compatible hardware and its inability to handle varying inhouse publication styles.

To offset these gaps, SG2 hopes to offer a product to operate independently of the workstation: a terminology aid for translators in small companies. Billed as Calliope-Manuel, its split-screen WP will allow window access to Calliope term banks as well as user entry updates and additions. Meanwhile, the big Calliope package will have to wait patiently until the LISP rewrite is out of the workshop.

"As it stands, Calliope can't be put into production," says Sonavision's head of translation. "What's been achieved so far is a laboratory prototype with promising results." Promises, promises. Potential customers can't even get a look at a functioning workstation. Jean Paul, ex-Technical Advisor to NCATP at ADI and now language industries advisor to the junior minister responsible for promoting the French language, goes further, suggesting a better way to sell the research done so far would be to develop intermediate Computer-Aided-Writing (CAW) products like spelling checkers and style correctors for the office and education markets – rather than spend still more time and money trying to integrate current modules into the elusive workstation.

Unless NCATP decides soon what its priorities are to be, the display shelf will be too crowded for Calliope. The competition's already there of course – Systran, Weidner and Logos, each with their benefits and drawbacks. Another future rival in the French marketplace is Cap Gemini Sogeti, the leading European dataprocessing services company, whose Cap Sogeti Innovations subsidiary (CSI) carries out industrial language research. CSI were at one time bidders in the NCATP, having previously worked on multilingual projects with GETA, only to be turned down. According to Roland Varenne, CSI's deputy general manager, the market for CAT products in industry is weak: for most users they are too expensive. And they are distrusted and dis-

dained by human translators. (Luddism, or a conflict of industrial language philosophies?)

Varenne believes that the language industries' central problem is "how to develop tools to help users conceptualize what they expect from natural language interfaces with machines." CSI's rival product is a "language engineering workshop" – a box of intelligent language tools both for large-scale document producers and local translators, providing knowledge-based approaches to specific needs and a dedicated WP that can help but not replace the human translator.

The kit's central powerhouse is what Varenne calls a "meaning generator," which analyzes text for semantic chunks, transforms them into an abstract representation and then offers them to a battery of dedicated programmes: a multilingual translation system, technical summary writer and a natural language KB for an expert system. The core system sounds suspiciously like the Ariane configuration, but Varenne claims that the CSI product is based on a "radically different" approach to computing. GETA's ideas, he claims, are based on "old dataprocessing techniques."

Built-in obsolescence, chronic delays, uncertainty about future state and private sector backing. Is Calliope bound for the scrapheap before she's even made the audition? Project engineers claim that once plugged in, the translation motor does generate satisfactory output at a sufficient rate, given prespecified constraints on text type. "What keeps us optimistic," says one project member, "is that the basic translation concept is good, even though the superstructure is weak." NCATP staff were amused yet frustrated earlier this year by a journalist who tested a number of CAT products including Calliope on a passage from Victor Hugo, blatantly ignoring both MT design and purpose. Talking about some vague n% performance coefficient for translation, they argued, is pointless unless a range of strict and relevant parameters is laid down first.

So, if you need it, Calliope can in theory deliver rapid raw Aeronautspeak. The question that remains is whether the large industrial concerns that can afford it will eventually feel like switching to Calliope, or whether they will prefer to invest in the next generation of intelligent language tools. Airbus Industries, an obvious potential customer, are still not automating their translation service in Toulouse and make no promises. And Aerospatiale, Europe's biggest word cruncher, is giving Systran a trial, mainly because of its EEC compatibility. Oh well, per ardua ad astra, eh Calliope!

Andrew Joscelyne is *LT's* contributing editor in France.

How GETA's interlingua approach works.

Fellow travellers will be familiar with GETA's long experience in the MT field and especially with its pioneering work on the "pivot language" or "interlingua" method of moving from source to target language. In this approach, labelled tree structures produced by the initial source analysis are transferred to an intermediate structural representation – the "pivot" – independent of any given language, where they are assigned special lexical entries based on "universal" semantics. Finally, the output of this transfer stage is re-interpreted into the target language, in a workdown from lexicalized nodes through syntactic structuring to morphological details. In practice, the ideal pivot language has so far proved an impossible construction, and the transfer stage in Ariane 78 is performed at a level closer to that of bilingual equivalences, where an abstract labelled syntactic structure is given a word-for-word translation in the target language before the output sentence structure is generated. The three stages, written in special linguistic programming languages, will allow significant cost benefits when the system is expanded to cover multiple language pairs, since each language is rendered machine-readable once and for all, regardless of whether it is a target or source.