

COMPANY PROFILE: EUROPE'S LARGEST

TECHNICAL DOCUMENTATION COMPANY DIVES INTO LANGTECH

by Andrew Joscelyne

Suppose your company designs, develops, manufactures, and then markets a state-of-the-art flying-machine. One by-product of such a mammoth engineering job will be the large gobs of documentation spewed out in its wake, seemingly at the speed of light. You'll have to provide your invention with user documentation, parts lists, technical data sheets, inventories, pilot and technician training manuals, fourth line inspection data, and plenty more. Before your brainstorm can reach for the skies, you'll already be swamped with loose-leaf folders, file cards, and a storage problem.

Then you'll sell a fleet of your flying-machines to a client who needs all that documentation in another language. The whole lot will have to be translated.

A year later, a new information storage device will appear on the market, allowing your clients to retrieve all this information via a computer terminal. So every document will have to go on a CD-ROM or videodisk, in turn requiring access tools so that engineers and support staff can find what they need.

This entire exploding circus of information — whether printed on paper, scanned into a CAD database, or provided online via a videotext service — will then have to be updated, as modifications, upgrades, and new models appear.

Now wait a minute! You make airplanes, not documents. You've got a major technical communications problem, and you need help. Either you buy a port-a-cabinet and install a tech-coms department in it — hoping your budget will stretch that far — or you farm the job out to experts.

But which experts? A technical writing company or an information management outfit? A translation bureau or a computerized maintenance management service? A software house specializing in hypertext interfaces or a print shop?

One outfit that attempts to offer the full range of these services to major industrial clients is the mammoth French services company SITE, whose head office is just a diskette's throw from the Villacoublay military airfield at Valizy, southwest of Paris.

NO COMPETITION

"SITE's biggest disadvantage," quips Erik Lebrat, head of SITE's translation department, "is that we have no competitors."

Maintaining standards and assuring quality constitute a major preoccupation of all high-tech firms these days — and that's the league SITE is in. But there's no doubt that it's difficult to keep an objective sense of your own worth when you're the biggest technical communications company in Europe.

Of course, it hasn't always been so. In 1984, a company called Sorovision was set up to meet urgent documentation needs as French industry awoke from the ruins of war. Fifteen years later, two of Sorovision's employees split off from that company to set up their own competing outfit, ITEP.

Then, in 1985, the French banking, perfume, and publishing group Cora Révillon decided to develop a technology arm, buying up Sorovision. And three years later, Sorovision bought up ITEP. So nearly thirty years after they'd split up, the two market leaders merged again to become SITE.

SITE's latest takeover has been of Sédéc, France's third largest technical documentation firm and formerly a wing of the major nationalized electronics firm Thomson. Sédéc is now a full SITE subsidiary.

Today, SITE's massed ranks comprise some 17 centers in France, employing 1700 technical writers, translators, computer scientists, information managers, and technical staff.

In the run-up to the single European market, SITE's ambitions, like those of many other enterprises, are bursting beyond national boundaries.

The company has bought up technical communications businesses in Italy and Spain, and is currently negotiating a cooperative agreement with the UK company Omnitech, supplier of technical documentation for the Channel Tunnel. Acquisitions in Belgium and West Germany are rumored to be the next step.

On Site

France's Site is big (1700 technical writers, translators, computer scientists, information managers and technical staff), it thinks big (it's into 20-year commitments, not one byte stands), and now it's bitten off a big challenge.

The EW company profile.

VIVE LA FRANCE
**Strategy**

The underlying strategy behind SITE's healthy appetite for documentation companies is founded on the company's policy of locating technical centers close to major clients' industrial facilities.

In France, SITE technical centers can be found in Toulouse, nice and close to major client aeronautics company Aerospatiale, and in Toulon, convenient for the French navy. SITE has two mainframe computing centers, at Maisons-Alfort and Toulouse, which are linked to all the other centers.

An extensive network of technical skills, computing resources, and project-management experience, backed up by a commitment to offering leading-edge expertise, is fuelled by a grandiose yet quite realistic vision of how the company's work fits into VLSI - Very Large Scale Documentation Integration.

You can compare SITE to a company selling helicopters to the government. It offers its product at "ownership price," after the Japanese model of selling a complete package, and is willing to make itself responsible for that product for a 20-year period. This means that not just the documentation, but also the parts lists, nomenclature, and other sub-documentation systems all have to be jigsawed into the wrapping service.

"SITE's approach," as Lebreton puts it, "is to place ourselves at the epicenter of our clients' needs. They don't necessarily know how to frame their ideas for the information system to accompany a fleet of aircraft, oil platform, or factory production line."

Lebreton explains the integrative nature of the SITE service by describing the client's product - say an aircraft - as not so much a one-off machine or vehicle, but a "vector." A whole collection of electronic, electromechanical, and engineering systems and subsystems move in unison along the vector, each with its manufacturing origins, documentation, and spare parts lists.

"Technical communication is not just desktop publishing or translation, but the technologizing of a wide range of management systems into a whole. And translation is not a separate intellectual task. It is a natural export support for our clients. This is how our services fit together in a systematic way."

The total approach SITE offers includes technical writing, translating, digitizing large scale documents - the company claims the largest concentration of large and small format processing equipment in Europe - and even printing. In addition, with its staff of 350 computer scientists, the company also develops standalone computer-based products inhouse.

**Do the Doc**

At Vélizy, some 40 technical writers bear away on specialties from mechanics through computing to medical documentation and aeronautics.

The writers charged with handling a given documentation job for, say, a Mirage 2000 jet - the pride of planemaker Dassault's military stable - will kick off by plowing through an unsorted collection of information provided by the client. On this basis, they then plan the structure of the documentation. Most of the technical specifications upon which writers base their descriptions of complex pieces of equipment are laid out in diagrams or charts received from the technical developer's office.
Next, the writers find out about all the international documentation standards—what is needed for maintenance, pilot, ground crew, and weapons manuals—as well as the actual technical standards on security, electrical, mechanical, and other issues. Once all this has been prepared, the writers can go to work—on computer.

Writers currently input text using IBM PCs. SITE—like everyone else—anticipates a future increase in the use of Macs as technical writing tools, but doubts about their total adaptability to a network configuration. SITE's technical writers came rather tardily to word processing software. Believe it or not, but today they use Word, as do their bilingual colleagues on one floor below.

In SITE's eyes, a good technical writer should also be a good project manager. One of the crucial considerations here is the ability to work to external demands such as deadlines.

The writer's relationship with the client involves everything from drawing up a schedule to providing product follow-up in the wake of client feedback. The technical writer in some ways "trains" the client, offering a window on the interface between product and user.

In drawing up a schedule, a technical writer will make an estimate of the time to be spent on each chapter of a document. Often, clients have no knowledge of the information sources available for the file at hand and are concerned more with the cost-effectiveness of various logistical problems.

If the documentation is destined ultimately for translation, the schedule will differ from when only a first version of English documentation is needed. SITE tries to find out what the client's budget is and offers a range of prices accordingly. These can range from the full documentation works, with complete storage solutions, printing, and updating services, all the way down to rapid work to satisfy one-off needs via the fax machine.

**UPSTAIRS DOWNSTAIRS**

As for translators, every major contract that Vélizy's 40-strong translation team tackles is assigned to a project leader, who manages the incoming work—including selecting the translators to work on it—client relations, and copyediting.

SITE's translation department has contributed enormously to the company's success with major industrial clients. And with one such client, plane maker Dassault, the relationship is particularly close. While translators are putting the maintenance manuals for, say, a Falcon 900 aircraft into English, SITE's technical writers may still be working on the original documentation in the same building, one floor up. Translators can just climb the stairs and consult the writers when they want ambiguities clarified.

On some occasions, translators have even been known to leave their airy open-plan offices for the Le Bourget airshow grounds north of Paris to discuss a point with a technician. Dassault designs and buys both civil and military planes. Also, it has its own Ground Support Equipment Division, producing mountains of additional documentation. So SITE translators have their work cut out.

"Qualified translation is a manpower problem in France. With product time cycles shrinking yearly, we'll never be able to find and train the people to translate the next generation of documentation."

—Erik Lebreton

Since the military work is for the French state, however, non-French citizens—who include many SITE translators—do not get security clearance to work on the texts. Consequently, they can't officially use the dedicated terminal lurking in a corner of their department. This terminal is hooked up for online access to a database storing the translations of every part of the total range of French military equipment, with a complete set of validated terms.

The current text production technology in the translation department consists of a Token Ring Network with a server offering support for Word 4.2—"Word 5 has too many bugs," says one translator—on PCs. Printing is on laser printers. In addition, the department uses SITE's own terminology management tool Phénix, plus a large range of conversion features to allow interfacing with files from other word processing software.

The translation department's technical support staff will produce pathways between any desktop publishing format or text editor on hardware from PCs to mainframes. It sometimes trains translators on new tools, such as Interleaf, for client-specific projects.

**DISCRETE PRODUCTS**

On the software front, although language technology is the field in which SITE's software profile is now highest, it is still a fairly recent innovation. The company—preceded by its parent companies—has clocked up more than a decade's worth of experience in the development of documentation and information management tools.

The documentation tools developed by SITE's 350 computer scientists include CD-ROM-supported database-accessing tools such as Hyperview and Genview—both originally developed by Sèdec as Thomson products—and a standalone document-retrieval system called Archivision, used with DDD and DBMS-generated databases.

In addition, there are the military aeronautics logistics management systems, Gimi and Gistoock, which handle all kinds of maintenance, inventory, and operational decisions for the Mirage 2000 aircraft. And there's an integrated logistics tool, CILS, designed under UNIX in C, to run with the Oracle DBMS and portable to any computing environment.

However, those highend information management systems, developed for both military and commercial datacrunchers, are not for sale on the open market. This is partly due to security considerations but also those of corporate image—"SITE has no ambitions to set up shop as an all-round software manufacturer," says Erik Lebreton.

By contrast, the company has no such reservations when it comes to the marketing of its linguistic software. It has run avaricious advertising campaign for its high-end termware products Phénix and Aqua. And if Ariane—SITE's first foray into the precarious world of machine-trans
lalion — is successful, it will amount to something of a company flagship.

FROM PAPER ASHES, PHÉNIX

"Back at the beginning of the 80s," recalls Erik Lebreton, "we realized that as the export drive and growth in documentation in general were beginning to put translation quality and productivity to the test, it was time to systematize our terminology, especially since we calculated that translators were spending up to 40% of their time looking up terms.

"So over the next seven years, we developed a tool that would respond to the whole range of SITE's needs. From the ashes of filecards piled up in a shoebox arose Phénix."

SITE looked at the competition in terminology-handling available at the time and concluded there was nothing to fit the bill.

On the one hand, mega-terminbases like the Canadian government's English-French Termium were woefully unadapted to translator workstation conditions: call up a term, and the mainframe running the database would just discharge everything held under it. The only people to gain from a hotline to Ottawa were the phone company.

On the other hand, the PC-based "electronic dictionary" terminology tools were conceptually limited, the main criticism being that the computer was not used intelligently; dictionaries came as separate lists rather than relational databases.

Dividing up the whole range of terms into isolated "little box" glossaries (based on language-pair, language-pair plus domain, or language-pair plus domain and client-use) created redundancy while wasting the computer's cross-referencing abilities (trace was lost). Using more than two languages could create anarchy for the user.

SITE's approach was to map out a set of dependencies in the conceptual structure of the lexibase, rejecting "little box" architecture in favor of an elaborate tree-branching model. In so doing, Phénix's designers first made a crucial distinction between a term's applications domain and its industrial sector. A term is associated with a domain, but that domain may occur in any number of sectors.

For example, the term oil pump is from the hydraulics domain, which can be found in such sectors as aeronautics, automobile engineering or oilfield technology; a mother card is from the electronics domain and is often found in the computer hardware sector.

The decision as to whether domain or sector is more important at a given moment in the translation process is a function of the document in question. And this is where a further complication in major project terminology handling comes in: the client.

"What we wanted to provide was not the totality of alternative terms that the terminbase could throw back at the unwitting translator, but the right translation in the sector in which they were working at the time," explains Lebreton.

And this type of focus had to be available in a system that would maintain the multiple links between different terms in three languages at a time, so as to offer immediate access. By preselecting the domain, the sector, and eventually the client, the translator gets the only possible translation displayed for the term in question.

UPDATING VIA THE NETWORK

The eight million francs' worth of development work on Phénix began in 1985, initially on bottom-end XT's, but with workstation ambitions. "And we eventually wanted to be able to support distributed processing and integrate the system in local and wide area networks," says Lebreton.

Another desirable feature then raised its head: user participation. If scattered users could have immediate access to a developing terminbase, then why shouldn't they also be able to add new terms to the base?

To provide such a facility, SITE added a bulletin board, allowing newly encountered terms to be held until final validation. "Phénix is a democratized system," says Lebreton, "allowing not just the totality of terms to circulate between various centers, but also giving users some say in its development."

With Vélizy generating around 200 terms a week, and SITE's 17 centers a total of 500, about 80 terms are always being validated at any one time.

S I T E ' S W O R D W O R K I N G

P R O F E S S I O N A L S

SITE translators are among the best paid in Europe, but then they produce twice as much as many of their counterparts elsewhere. Morale is so high that some SITE translators are planning a pressure group to promote their profession's interests. This group's concern will not end with translation. It will extend to technical writing, for which demand is booming but training and professional status — at least in France — are almost nonexistent.

F E E F I E F O H F U M

Every spring for the past three years, Erik Lebreton and SITE translation manager Mike Hunt have flown to the UK. They go there in search of young blood among recent graduates from various UK translation courses, such as those at the City of London Polytechnic and Bradford University.

To find just seven possibilities with the right SITE caliber, they may interview 80 or more candidates. "We're not interested in wordprocessing ability, but more in natural, human qualities such as the ability to think for yourself, adapt to teamwork, and react intelligently to the translation process. 50% of the profile is sheer personality."

Once translators are engaged for SITE's initial six-month probationary period, they receive 12 two-hour training sessions in the wordprocessing, terminology, and other support tools that are part of the SITE translator's workbench. Training is also given in preparing and using client-specific stylesheets and other areas of document-formatting.

A translator is expected to produce quality rather than quantity. But as a standard benchmark, the department breaks even if translators produce a minimum 2000 words a day (or eight 250-word pages).

"This is nearly double the expectations in some privileged international organizations," says Hunt. "But we believe there's more of a challenge in a dynamic company like ours than in the lower productivity expectations found in most in-company translating departments."

Translators who stay on at SITE after the probationary period find themselves working in what is reputed to be the best paid outfit in Europe. If the rumors of princely rewards are true, credit must be mostly due to Erik Lebreton's tough bargaining with management over the needs of the translation department.

"If you want high quality, then you're going to pay high salaries," he says flatly. And he's boosted SITE translator salaries by 23% over the past two years.

Further good news for SITE recruits is the lack of an obvious hierarchy in the translation department. SITE appears to avoid the administrative-type grading used in large international or
Ariane-78: SITE'S CAT HOPES

Ariane-78, a basic computer-assisted translation software package created by the Grenoble-based GETA research team, is SITE's bid in the machine translation stakes.

Ariane is currently under development as a commercial product by Ingetech company B'VITAL (Bernard Vauquouis informatique et Traitement Automatique des Langues). The unlikely name of this Grenoble company, now 65% SITE-owned, is a homage to the late emeritus grise behind a long series of GETA translation systems which began in the 1950s.

As an automatic translation system, Ariane is a second-generation rig. It offers a tree-structural analysis of the sentences of the input text before transferring them into the structural model of the target language.

Ariane's programs, which compile data, grammars, and dictionaries, are written in specialized languages for linguistic programming developed by the GETA team.

For morphological analysis, the language is ATEF, for syntactic analysis, the powerful tree-translation language ROBRA. Lexical transfer is carried out by a bilingual dictionary consultation tool written in TRAN, while the target morphological specifications are generated by a program written in SYGMOR.

The complexity of interaction throughout this labyrinth of language, as current GETA boss Beitel sees it, is such that the translation of one word can require vast quantities of computing power - 5 Mips at the latest count when running on a mainframe.

Operating system, it can handle up to 16 languages, including local variants, and is designed to support Cyrillic and Arabic scripts.

Since this kind of investment only makes sense for large corporations, SITE has developed a standalone PC version of Phélix, called Aquila, adapted to the personal workstation. By maintaining the basic organizational principles and dropping the network management software, SITE claims Aquila is selling well at around FF8,400 (US$1400) apiece.

Ariane-78, unlike translator's terms, is a monolingual French lexibase and will include phonetic, morphological, and lexical data, as well as syntactic and semantic coding.

SITE's job will be to develop a dictionary making workstation to handle this task, so that once the process of entering a given term with its full panoply of relevant coding is automated, the rig can be commercialized as a specialized tool.

Quite what ultimate application CNET has in mind has not been revealed, but in line with the organization's extensive research into speech recognition and synthesis technology for remote access to databases, no doubt the material to be managed on the lexicon-station will eventually find its place inside a "smart" talking information point.

WP TO CAT

As reported in EW/17, SITE recently bought a 65% share in the small Grenoble-based machine translation company B'VITAL.

B'VITAL had long been trying, unsuccessfully, to market its Ariane translation engine, a legacy of long years of research devoted to the French National Computer Assisted Translation Project, which was later axed (see EW #9).

Coincidentally, the former Sonovision component of SITE was actually an alpha-test site for the original Ariane translation engine in 1986, when the whole project suddenly ground to a halt.

In returning to Ariane now, SITE's strategy is clear. Having developed an expertise in termbase
building—a prerequisite to any useful computer-aided translation activity—SITE is preparing itself for its role as the major technical translator on the French industrial scene.

"Qualified translation is a manpower problem in France," explains Lebreton. "Six years ago, Dassault would give us a couple of years to translate the documentation for a Falcon 200 aircraft. A year ago, when the Falcon 900 came out, we only got 12 months. With product life cycles shrinking yearly, we'll never be able to find and train the people to translate the next generation of documentation."

Rather than hang around waiting for a Japanese MT product, SITE decided to test the unused Ariane translation system on an industrial testbed. The French Ministry of Industry will subsidize the operation, but SITE still needs to find an industrial partner—preferably one of its clients—to put up the development capital necessary to maintain and beef up the system as a fully functioning workstation tool.

"We're not interested in developing a server-based system. Ariane must be a tool that can be fully integrated into the in-company documentation system—which means developing fully integrated man-machine interfaces."

As Christian Boitet, head of the GETA research team which originally developed the programs underlying the Ariane engine, said recently on the future of current commercial MT systems: "There's no future for MT systems whose developers want to hang on to coding secrets at the heart of the program. They are going to have to sell the complete system together with maintenance and control over the code or gradually go out of business."

MT ON TRIAL

Since the Ariane system has never had a chance to prove its worth with serious translation loads, SITE will evaluate the real industrial potential of Ariane over 18 months, translating 5000 pages of documentation for Dassault's Falcon 900 aircraft.

The text of this documentation exists in digital form. The cost/time parameters of its already human-translated version are known. The client quality validation has been made. The original technical writer is on record. And SITE has a major aeronautics termbase in Phénix.

By being able to compare the full parameters of machine and human translation versions—how many times has this been done on a large scale corpus?—SITE will be able to check closely on output quality and costs. And the results will provide facts of the hardest kind for all those potential MT users waiting in the wings.

Three types of text will be tried: maintenance work cards, an operating manual, and the flight manual. And a stringent statistical analysis will be made of errors.

"If we manage to reach our target of reducing translation costs from around FF300 (US$55) to FF150 (US$25) a page—this is a 50-page-a-day output, representing 15 translator days—then we'll think about going on to phase two of our general plan: industrializing a two-way English-French system."

The US$25-per-page target for MT output, incidentally, seems to be fairly standard, with MT production companies like Tovna Ltd. citing the same figure.

If Ariane doesn't come up with the goods, Lebreton is categorical: SITE will drop it—though he is quick to point out that natural language processing, in some form or other, will continue. Indeed, the company has a project to develop a technical center for NLP at Vélizy and to pursue terminology work and upgrading on Phénix.

Lebreton is optimistic: "We wouldn't bother to start on this project if we didn't feel the basic material was worth investing in. B'VITAL's own testing has left us feeling very positive."

The successful testing of Ariane, on the other hand, will mean that SITE will have added one more technological answer to the constellations of problems posed by translation.

Perhaps the next step will be for SITE computer scientists to dream up an automatic documentation writing system: where you feed in the data sheets and tech spaces, and out comes the user manual. ■

Andrew Josselyne is Electric Word's 1989 Reporter of the Year.

organizations where translators get promoted to "rereader" or "revisor."

"Revision of someone else's work is sheer slavery!" says Lebreton. "Our approach is to develop tools such as Phénix, which will obviate the need to see revision as a totally separate job for a given individual." In a move to raise the status and promote the "knowledge-worker" image of translators, Lebreton and his colleagues are planning a new professional organization to correspond to the needs and rights of translators in France, similar to the Dynamic Institute of Translators and Interpreters in the UK. It will be like a new professional organization to correspond to the needs and rights of translators in France, similar to the Dynamic Institute of Translators and Interpreters in the UK. It will be the\n
GROWING DEMAND FOR ENGLISH

As for technical writing, training is almost nonexistent in French higher education. According to SITE technical writer Hélène Ressiguier, most of her colleagues pick up their skills on the job. However, given the current growth in demand, this lack of training now poses a "dangerous problem" for companies such as SITE.

The technical writing profession needs to get itself organized in order both to transmit acquired skills via training courses and serve the interests of working technical writers.

The problem is compounded by a growing industrial demand in France for technical documentation in English. A glance through any good professional journal such as The Manual in the UK or even the recruitment ads in Le Monde, reveals numerous vacancies in French companies for technical writers offering English.

SITE has even been thinking about developing an English-language technical documentation group as a special service for clients. Ressiguier: "Certain major clients do their cost-effectiveness sums before deciding whether to do their documentation in English and then translate it into French, or vice versa."

"In France, a product such as a Mirage jet will automatically have its documentation written in French—it's part of the tradition of national prestige. However, a second-rate or simply secondary product, where the national investment is less marked, will probably be a candidate for direct English documentation."

This situation is leading to a reconsideration of career possibilities for translators. One path now open to the experienced English-speaking translator is to become an English technical writer.

Here again, the question of production cycles weighs heavily in the balance. As schedules get shorter, there is less time for developing French documentation, which will have to be translated into English anyway.

For SITE, in the particular context of French industry, the gradual technologizing of the translator-training process is converging with that of the technical writer, offering new exciting options for specialized translators. ■