

Développer

Pas facile de parler l'« européen »

MULTILINGUISME ■ Tous les textes produits par le Conseil de l'Union européenne le sont dans des outils d'indexation et d'interrogation multilingues facilitent la tâche du Secrétariat général.

La mission du Conseil de l'Union européenne est d'adopter les textes législatifs. Les sujets sont proposés par les commissions puis traités d'abord par des groupes de travail composés d'experts, ensuite par le Coreper (Comité de représentants permanents), s'ils sont retenus, par le Conseil, qui bénéficie pour cela du support administratif et logistique du Secrétariat général. Autant dire que chaque texte donne lieu à de nombreuses versions successives. « Et ce

sont des textes légaux, ils doivent donc être traduits dans les douze langues officielles de l'Union européenne et être aussi précis dans chacune des langues », insiste Philippe Vleminckx, administrateur principal à la division technologie de l'information du Conseil. La grande majorité des 2.400 personnes qui composent l'effectif du Secrétariat général travaillent sur des documents, les traduisent, les copient, les diffusent... Au total, le Secrétariat général produit quelque 80.000 documents par an, qui représentent un volume global de 500.000

pages. Pour retrouver un texte sans perdre trop de temps, ne pas travailler sur une version obsolète d'un texte ou savoir comment les locuteurs d'une autre langue de l'Union traduisent une expression, la division technologie de l'information a développé deux projets, baptisés Terminological Information System (TIS) et Archive Information System (AIS). « Les deux projets sont vraiment multilingues, cela signifie qu'ils peuvent traiter des douze langues simultanément et ne se limitent pas seulement à afficher les écrans ou les menus en plusieurs langues », précise Philippe Vleminckx, qui est aussi chef du projet AIS.

Le projet TIS, lancé à la fin des années 80, visait à homogénéiser la terminologie utilisée dans les différentes langues communautaires afin de garantir la cohérence des textes légaux. Il consiste en une base de données, qui contient des « concepts » traduits dans les douze langues officielles. TIS est ainsi le dictionnaire multilingue de référence de l'Union, interrogeable par les traducteurs depuis leur poste de travail. Son vocabulaire est relativement stable ; la base de données s'enrichit de 5.000 à 6.000 concepts par an. Aujourd'hui, TIS est disponible sur CD-ROM, ce qui permet aux traducteurs de s'en servir sans se connecter, lorsqu'ils accompagnent les ministres du Conseil en déplacement, par exemple.

AIS, pour sa part, a été lancé en 1993. Son but est de faciliter l'accès à tous les documents qui ont déjà été produits, « car, dans le processus d'élaboration des documents légaux, les textes se ressemblent beaucoup », explique Philippe Vleminckx, on ajoute un paragraphe ou l'on modifie un mot, mais on n'écrit pas un texte nouveau à chaque fois ». AIS est constitué d'une base de données qui contient 340.000 documents, soit 16 Go (gigaoctets) d'informations. Un utilisateur peut consulter ces archives, mais aussi exploiter les documents, qui sont accessibles au format traitement de texte, autrement dit directement réutilisables sur un poste de travail.

Mots clés. Pour mener à bien le projet AIS et enrichir les fonctionnalités de TIS, la division technologie de l'information a retenu SearchServer, le moteur de recherche documentaire plein texte et multilingue de l'éditeur canadien Fulcrum. Ce produit répondait au cahier des charges : possibilité d'accès simultané aux douze langues de l'Union, à la fois en recherche et en visualisation des documents ; produit disponible en architecture client-serveur sous Unix et Windows et capable de supporter jusqu'à 750 utilisateurs – ils sont déjà 650 à l'heure actuelle – et d'accéder à des dizaines de milliers de

Le multilinguisme primé par le CNRS et l'Anvie

Chaque année, le département des sciences de l'homme et de la société du CNRS, auquel s'associe l'Anvie (Association nationale pour la valorisation interdisciplinaire de la recherche en sciences de l'homme et de la société auprès des entreprises), prime une action de valorisation exemplaire associant sciences humaines et sociales et secteur économique. Fin 1996, ce prix a été remis à deux chercheurs du laboratoire Parole et langage (CNRS-Université de Provence, Aix-en-Provence), Malek Boualem et Stéphane Harie, pour l'éditeur de documents multilingues qu'ils ont mis au point : MrScript.

MrScript permet aux utilisateurs de saisir et de mettre en page des documents dans les langues européennes, mais aussi dans des langues d'écritures différentes comme l'arabe, l'hébreu, le grec ou le coréen, entre autres. MrScript est d'ores et déjà utilisé dans trente-deux pays. Développé dans le cadre de projets européens, MrScript a associé des partenaires scientifiques ou universitaires et des industriels dans de nombreux pays européens.

La traduction, dernier bastion ?

AUTOMATISATION ■ Les grands projets de systèmes de traduction automatique n'ont jusqu'ici donné que de piètres résultats.

Il s'agissait du service proposé par Compuserve, utilisant le logiciel Transcend. La médiocrité du résultat l'avait incité à nous communiquer le texte restitué par le système de traduction, texte dont voici les premières phrases : « L'information fiable est devenue le lifeblood pour la survie dans today's le monde qui rapide-changeant. Tous les jours, plus de compagnies créent plus de leurs opérations

Voici la traduction produite par Transcend : « Ce livre explore une classe nouvelle et prometteuse de systèmes de direction de base de données, le DBMS objet-relationnel. Le spectre de régions d'application couvertes par le DBMS objet-relationnel étend du video et de la direction de bien graphique dans l'industrie de divertissement pour chronométrier problèmes d'analyse de série dans le marché de services financiers. Les bases de données scientifiques

commercialisé en France environ 7.000 francs. « Ce livre explore une nouvelle et prometteuse classe des systèmes de la gestion de la base de données, le DBMS objet-relationnel. Le spectre des régions de la candidature couvert par les rangées DBMS objet-relationnelles de video et gestion de l'atout graphique dans l'industrie de l'amusement pour fixer des problèmes de l'analyse de la série dans le marché de services financiers. Les bases de données

A quoi un utilisateur moyen consacre-t-il son temps ?

... à lire des informations

5 % seulement du temps est consacré à la lecture d'informations vraiment utiles et correspondant à la demande initiale

