

Une banque de troisième génération

1 Le CTI

La mission du CTI, ou Centre du traitement de l'information du Ministère des Affaires économiques du gouvernement

belge, est triple:

- Il fait office de bureau de service informatique (il développe des applications pour le Ministère).
- Il coordonne les applications de bureautique du ministère.
- Par le serveur *Belindis*, il exploite une série de banques de données documentaires et factuelles.

1.1 Données techniques

Parmi les banques de *Belindis*, figure une banque de terminologie multilingue (comportant notamment le français, le néerlandais, l'anglais et l'allemand), accessible par DCS ou par ligne directe. La conception de *Belgoterm* a été faite par M. Dupont, traducteur-directeur au Ministère des Affaires économiques. Elle a été programmée par la société Orda-B, en collaboration avec Informabel.

Programmée (en *PL1* et en *Assembleur*) il y a quatre années, *Belgoterm* est une banque de terminologie technique. Elle est encore en phase d'alimentation massive. Elle comporte 15 000 termes et 15 000 abréviations d'informatique.

La banque tourne sous *CICS* et supporte les terminaux 3270 (qui ne permettent pas la représentation des caractères spéciaux), 3179, 3279, PC graphiques 3270 AT/G/PS (répondant aux critères de typographie riche). Une commercialisation de la banque par sous-domaines, sur disquettes, est à l'étude.

1.2 Pourquoi une autre banque de terminologie ?

Les banques de terminologie opérationnelles, c'est-à-dire celles qui comportent un nombre suffisant de termes pour que leur taux de réponse atteigne au moins 70% dans les domaines couverts, ont été conçues il y a une quinzaine d'années au moins.

Or les nouveaux utilisateurs de l'informatique, dont font partie la plupart des traducteurs, ont été habitués à des logiciels récents, dont la convivialité est très grande. De plus, des terminaux supportant le graphisme sont à présent commercialisés.

Enfin, la terminologie s'est fortement développée au cours de toutes ces années. Des bilans, des comparaisons des banques de terminologie ont été menés à plusieurs reprises et certaines lacunes ont été pointées pour toutes.

Il n'était donc pas étonnant qu'une banque de nouvelle génération, tirant profit des expériences précédentes, fasse son

apparition sur le marché. Pas plus surprenant que ne l'est, par exemple, la succession des standards en traitement de texte.

2 Contenu de *Belgoterm*

Si les langues incluses actuellement sont les langues «belges» plus l'anglais, d'autres langues comme l'espagnol, l'italien, le portugais, le danois, le norvégien, le suédois, le grec, le russe et le polonais pourront être ajoutées sans modification des programmes.

Les dessins techniques, les formules, les schémas peuvent être incorporés à titre d'illustration des notions.

Une fiche *Belgoterm* comporte les informations suivantes:

- Vedette;
- Domaine (rubrique obligatoire);
- Définition (rubrique obligatoire en l'absence d'un contexte et vice-versa);
- Contexte (rubrique obligatoire en l'absence d'une définition et vice-versa);
- Sources (référence unique pour l'ensemble des données de la fiche, pour la définition et le contexte ou référence ou pour une de ces deux rubriques);
- Compilateur ou terminographe (rubrique obligatoire);

Communauté française de

Belgique

- Réviseur ou auteur des modifications éventuelles (rubrique obligatoire, dans ce cas);
- Fiabilité, sous forme de code unique pour l'ensemble de la fiche, de la définition et du contexte, ou fiabilité propre à chacune de ces rubriques (rubrique obligatoire);
- Remarques ou observations encyclopédiques, linguistiques diverses (rubrique facultative);
- Synonymes (rubrique facultative);
- Antonymes (rubrique facultative);
- Génériques (rubrique facultative);
- Spécifiques (rubrique facultative);
- Termes apparentés ou co-hyponymes (rubrique facultative);
- Restrictions, ou marques d'usage géographiques, temporelles (rubrique facultative);
- Abréviation ou acronyme (rubrique facultative);
- Formules mathématiques, chimiques, graphes, dessins (rubrique facultative);
- Date de saisie (rubrique obligatoire);
- Date de la dernière mise à jour (rubrique obligatoire si la rubrique «réviseur» existe).

2.1 Fonctions de consultation

La consultation peut être ponctuelle ou interactive c'est-à-dire se faire par lot et en différé. Dans le second cas, la réponse à l'écran et/ou par listage est obtenue le lendemain de la frappe de la liste au terminal. La réponse à une interrogation groupée reste stockée dans la mémoire de l'ordinateur du CTI pendant une semaine; passé ce délai, elle est effacée, corrigeant ainsi les contre-temps qui peuvent empêcher l'utilisateur de prendre connaissance des informations demandées.

La consultation se fait dans la langue de l'utilisateur (français, néerlandais, anglais ou allemand). Des cheminements sont prévus dans la banque (saut à des fiches logiquement

reliées, à des fiches d'autres langues, à d'autres parties d'une fiche).

Tant pour les interrogations interactives que par lots, deux profils d'utilisateurs ont été prévus par défaut, afin d'économiser le volume des données transmis et partant, les coûts d'interrogation:

- Le traducteur, le rédacteur... bref, tous ceux qui cherchent des renseignements linguistiques sur un terme. Ces utilisateurs introduisent le terme inconnu, sa langue-source et la ou les langues d'arrivée requises. *Belgoterm* fournit alors, dans la ou les langues précisées, les équivalents, le libellé du domaine (affiché *in extenso*), la définition et/ou le contexte.

- Le chercheur, le documentaliste, le journaliste, l'enseignant, l'étudiant... c'est-à-dire ceux qui sont en quête d'informations de nature encyclopédique sur le terme. Ils dactylographient le terme et la langue-source. *Belgoterm* affiche en réponse la totalité de la fiche consacrée à la notion correspondante pour cette langue. Le saut à la fiche d'un hétéronyme, hyponyme, co-hyponyme, antonyme, synonyme affiché peut être demandé par la pression d'une touche.

Si le terme demandé est polysémique, le système affiche, en mode d'interrogation «encyclopédique», les domaines et, pour chacun d'eux, la définition et/ou le contexte correspondant. L'utilisateur peut demander l'affichage de toutes les informations sur un terme, pour chacun de ces domaines. En mode d'interrogation de type «traduction», le terme polysémique et le domaine sont seuls affichés dans chaque langue demandée. La frappe d'une touche permet l'affichage de toutes les données de la fiche pour un des équivalents. Le saut à la fiche d'un hétéronyme, hyponyme, co-hyponyme, antonyme, synonyme affiché peut être demandé par la pression d'une touche.

Les touches de fonction du clavier ont été reprogrammées pour les fonctions fréquemment utilisées: appel de l'écran d'aide, demande de découpe en syllabes et de reconstitution (*cf. infra*), saut à l'écran précédent, à l'écran de sélection ou au menu principal, listage des composés et des syntagmes comportant le terme demandé, demande d'affichage de la fiche complète, défilement arrière ou avant, demande d'un nouveau terme pour les mêmes langues, affichage du ou d'un des graphiques illustrant le terme demandé.

2.2 Optimisation de la recherche

Si le terme pluri-lexématique recherché ne se trouve pas tel quel dans la banque, le système procède, dans l'ordre:

1. À une recherche par masque (élimination des mots vides);
2. À une recherche au départ de tous les mots constituant le terme;
3. À une recherche au départ de la forme phonétique de tous les lexèmes constituant le terme;
4. À la découpe syllabique des lexèmes non trouvés aux étapes précédentes, si l'utilisateur le demande (*cf. infra*) et à la recherche sur les mots obtenus par concaténation des syllabes ainsi obtenues.

Des informations partielles sont donc données en l'absence de réponse complète:

- L'élimination des mots-outils est utile pour les syntagmes, dans lesquels l'usage des prépositions varie fréquemment.
- La recherche phonétique permet de faire aboutir une recherche de terme mal orthographié ou dont l'orthographe varie d'un pays ou d'une région à l'autre.
- Lorsque *Belgoterm* ne comporte pas de traduction pour la ou les

langues demandées mais bien pour d'autres, il le signale et le système fournit les traductions dans les langues non demandées, à titre d'information. L'intérêt de cette démarche pour des langues très proches comme l'allemand et le néerlandais est manifeste.

– Lorsque le fichier contient des parties du terme demandé, il présente les termes pluri-lexématiques comportant la vedette uni-lexématique demandée. La vedette, le domaine, et, pour l'interrogation en mode «encyclopédique», la définition et/ou le contexte, sont alors affichés pour chaque composé ou syntagme ainsi trouvé; l'action d'une touche permet l'affichage de la totalité de la fiche correspondante.

2.3 Découpe syllabique

Lorsque *Belgoterm* ne comporte pas le terme uni-lexématique dont on demande la traduction ou des renseignements d'ordre encyclopédique et que la recherche de termes partiels n'a pas donné suffisamment d'informations, l'utilisateur peut demander une découpe en syllabes, suivie de la reconstitution des lexèmes.

Chaque terme reconstitué trouvé ainsi est affiché dans la langue-source, suivi par les rubriques *Domaine* et *Vedette* (pour l'interrogation de type «traducteur») et par les rubriques *Définition/contexte* (pour l'interrogation de type «encyclopédique») de chaque équivalent, pour chaque langue demandée. L'absence de traduction pour un terme reconstitué, dans une langue demandée, est signalée. L'action d'une touche permet l'affichage de toutes les informations sur l'équivalent ou le terme reconstitué correspondant.

Cette routine est précieuse pour les langues agglutinantes comme le néerlandais et l'allemand: elle permet

par exemple au système de chercher les constituants de *boldrieboeksmeeetkunde*, à savoir *drieboek, meekunde, boldrieboek, boeksmeeetkunde*; en effet, le programme est capable d'identifier les voyelles et consonnes de liaison (comme *s, n*, dans les langues germaniques) et de ne pas en tenir compte dans sa découpe.

2.4 Gestion de la banque

Les mises à jour de *Belgoterm* se font en différé, par le truchement d'un fichier intermédiaire. Certains contrôles sont exécutés dès la saisie des éléments de la fiche, d'autres s'opèrent en différé. Des listages sont imprimés automatiquement pour permettre aux responsables de l'alimentation de la banque de corriger dans le fichier intermédiaire les erreurs signalées par le système, et de contrôler l'exactitude des homonymes. La saisie définitive des fiches n'est opérée par le système que lorsque toutes les anomalies et erreurs ont été rectifiées dans le fichier intermédiaire.

Le système remplit lui-même, pour toutes les fiches d'une collection donnée, certaines rubriques dont le contenu est répétitif: domaine, date de saisie ou de mise à jour...

Des routines, en cours de programmation, ajouteront en outre les termes génériques et les co-hyponymes lorsque les termes qui sont les termes spécifiques d'une vedette lui seront communiqués. D'autres routines de conversion sont actuellement programmées pour les formats standards des fichiers confectionnés sur micro-ordinateur (*dBase, Termex, Mater*)⁽¹⁾.

(1) Ces programmes sont rédigés par le Centre de terminologie de Bruxelles.

3. *Belgoterm* et le Rint

Le public-cible de *Belgoterm* est, dans un premier temps, constitué tant par les diffuseurs (rédacteurs de manuels, traducteurs, documentalistes, journalistes), que les étudiants et les spécialistes en informatique.

Comme *Cézeau*, à l'Université de Clermont-Ferrand, *Belgoterm* est une banque spécialisée. Les responsables mettent l'accent sur l'aspect encyclopédique de la banque et sur son exhaustivité dans certains domaines; le but est d'assurer un taux de réponse maximal dès le début de la diffusion de la banque. La représentation visuelle des liens notionnels serait très utile, notamment aux étudiants: la didactique des sciences a montré que l'organisation de la mémoire est associative et que, pour la saisie des concepts scientifiques, la compréhension de la structure d'une science est aussi importante qu'une définition précise des termes. Aussi la représentation en trois dimensions des liens logiques entre les termes d'un sous-domaine est-elle à l'étude.

Dans le cadre du Rint, on pourrait imaginer que *Belgoterm* soit un des canaux de collecte et des interfaces d'échange de terminologie avec d'autres «partenaires informatiques» du réseau, notamment les banques de terminologie spécialisées en informatique.

Pour ce même domaine, la banque pourrait peut-être également servir de trait d'union entre le Rint et le Sant (*Samenwerkingsverband Nederlandstalige Terminologie*), le réseau international de terminologie néerlandophone, dont les travaux terminographiques comporteront très souvent des vedettes en langue française.

Caroline de Schaetzen,
Centre de terminologie de Bruxelles,
Institut libre Marie Haps.