

Informatique & Bible, asbl - Belgique
Rue de Maredsous, 11 B5537 Denée - Belgique
Tél:+32(0)82.69.96.47 Fax:+32(0)82.22.32.69
cib@cibmaredsous.be



Interface n° e-95 Juin 2004

Les Dictionnaires électroniques

La revue "Traitement automatique des langues" consacre son 44e volume (n° 2/2003) aux dictionnaires électroniques. Ceux-ci apportent une valeur surajoutée à une version papier, car ils permettent un accès rapide à l'information recherchée et sous des formes diverses (ordre alphabétique, thématique, niveau de langue, fréquence, etc.). Grâce à une requête multicritère, l'utilisateur applique un filtre aux données, pour repérer plus efficacement ce qu'il cherche. L'informatique libère le dictionnaire de sa camisole papier. En effet, pour des raisons économiques ou pratiques, le dictionnaire-livre est limité en taille, souvent "monotâche" (spécificité) et "monocritère" (classement alphabétique). Les possibilités qu'offre le traitement automatique de la langue sont énormes. Mais pour les exploiter, il faut d'abord effectuer une bonne analyse des besoins – de l'homme et de la machine, et ensuite se débarrasser des réflexes liés aux contraintes des anciens supports!

Une première série d'articles concerne les problèmes rencontrés lors de la construction d'un dictionnaire électronique. Les auteurs y présentent leur méthodologie, leur manière de structurer les données lexicales pour construire une base de données. (Voir ci-dessous l'exemple d'informatisation du TLF). Dans la seconde partie du volume, les auteurs expliquent comment ils sont partis du point de vue de l'utilisateur, possédant des connaissances linguistiques plus ou moins limitées, pour élaborer leur stratégie d'accès à l'information. (Voir infra le cas du DAFLES).

Le TLF

J. Dendien et J.-M. Pierrel (ATILF, Université de Nancy) présentent la version électronique du *Trésor de la Langue Française* (TLF), dictionnaire de référence du français, réalisé par le CNRS, entre le début des années 60 et la fin des années 90. L'article expose les différents aspects de l'informatisation d'un dictionnaire déjà existant. La première opération fut d'encoder sur support électronique la totalité du texte (16 volumes, 35 millions caractères). La deuxième étape du processus a consisté à injecter dans le texte saisi "au kilomètre", des balises textuelles de type SGML, afin d'obtenir une version structurée. Étant donné l'importance du volume à traiter, il était hors de question de prétendre réaliser cette tâche manuellement. Il a donc fallu créer des automates capables de reconnaître les différents objets textuels (définitions, exemples, indications stylistiques, etc.). La reconnaissance automatique a été guidée par les éléments suivants: la typographie, les éléments ouvrants-fermants (comme les guillemets, par exemple), la nomenclature spécifique à quelques rubriques, la succession normée de certaines données...

Cette première version balisée du dictionnaire a demandé un volume certain de travail. Ensuite, la validation de la structure a été réalisée par un compilateur spécifique (écrit en YACC). Le résultat de cette phase de balisage est une version du TLF formalisé en XML.

La version numérisée possède les avantages du livre, renforcés par les possibilités d'une navigation en ligne, grâce au logiciel d'exploitation STELLA. Ceci dit, on a constaté que la majorité des utilisateurs n'exploitent pas les potentialités de cette version électronique, ils ont surtout recours au dictionnaire en ligne pour vérifier le sens ou l'orthographe d'un mot. Prenant en compte ce fait, la page d'accueil du site du TLF propose trois possibilités pour la recherche d'un mot. Premièrement: une case blanche, dans laquelle on tape le mot recherché. Une certaine marge d'erreur est tolérée (omettre un accent, une cédille...; taper une forme fléchie et non le lemme). Deuxième solution: une liste défilante, mais c'est assez fastidieux. Troisièmement, il est possible de faire une saisie phonétique du mot, ce qui constitue un intérêt pour les non-francophones.

Le TLF, ce prestigieux dictionnaire, est [accessible gratuitement sur Internet](#). Ses

concepteurs ont le souhait que cette formidable base de données lexicographiques devienne le fondement de projets de recherche et d'exploitations diverses.

DAFLES

Th. Selva, J. Binon et S. Verlinden (K.U. Leuven) ont travaillé à la réalisation d'un [nouveau dictionnaire accessible en ligne](#), le DAFLES, outil destiné à l'apprentissage du français langue étrangère ou seconde langue . Les dictionnaires électroniques traditionnels permettent de mener des recherches transversales. C'est leur principal atout. Cependant, quand l'utilisateur parvient à l'information recherchée, il peut être confronté aux mêmes problèmes qu'avec une version papier: quantité plus ou moins importante de texte à lire, enchevêtrement des rubriques au sein de l'article... une masse de données qui peut être lourde pour un non-natif.

Le DAFLES présente un certain nombre de fonctionnalités pour aider l'apprenant dans sa consultation. Le dictionnaire multiplie les accès à l'information. En d'autres termes, il met en œuvre différentes stratégies de recherche autour d'une information qui peut être dans le chef de l'utilisateur plus ou moins vague. Car comment trouver ce que l'on ne connaît pas? Le DAFLES permet alors d'accéder aux articles en français à partir d'une requête en cinq langues étrangères. L'utilisateur peut aussi faire sa recherche à partir des "domaines d'appartenance" (champs lexicaux). Le DAFLES a aussi le souci d'adapter le contenu aux besoins de l'utilisateur: deux types de définitions sont proposées (courte ou longue). Par ailleurs, un réseau d'hyperliens a été créé pour accéder à des informations complémentaires. L'exploitation de la synonymie et l'activation du schéma actantiel (qui?, quoi?, comment?, etc.) sont des moyens d'étendre le vocabulaire.

Ce dictionnaire électronique de deuxième génération a tiré profit de l'outil informatique pour aider à la consultation. Cela est rendu possible grâce à une structuration souple et cohérente des données lexicales, dont l'unité de base est la lexie, c'est-à-dire l'acception d'un vocable. Le DAFLES est donc une base de données qui s'articule suivant des liens sémantiques. En outre, cette base de données lexicographiques destinée en priorité à des utilisateurs humains, peut être exploitée par la machine pour générer (semi-)automatiquement des exercices de vocabulaire et leur correction.

Anne-Françoise Denamur

