

La résurrection du dictionnaire ancien par la déconstruction positive de l'informatique

Rey Christophe, Zaoui Corinne
Université de Provence, Equipe DELIC
Christophe.Rey@up.univ-aix.fr
zaoui@up.univ-aix.fr

Mots-clés: Informatisation, dictionnaires, balisage souple, XML.

Les travaux d'étude et d'informatisation des dictionnaires anciens que poursuit l'un d'entre nous (C. REY) mettent clairement en évidence les problèmes d'informatisation liés à la nature même de ces données, à savoir un certain manque de rigidité structurelle. Ce manque de rigidité occasionne indubitablement une moins bonne lisibilité des données, que les dictionnaires modernes – par le biais de leur informatisation systématique – semblent avoir réglée. Il est ainsi parfois difficile d'identifier les différents champs d'information constitutifs des tous premiers dictionnaires, et de repérer précisément, par exemple, ce qui relève du marquage grammatical, de la définition, etc., afin de pouvoir reconstruire, récupérer, retrouver ou mettre en évidence les informations les plus saillantes qui s'y trouvent.

Les soucis de structuration des données pour une manipulation et une interrogation grâce aux normes de codage informatiques nouvelles prônées par C. ZAOUÏ, permettent d'apporter des solutions nouvelles à notre projet de rétroconversion d'un dictionnaire encyclopédique, le dictionnaire *Grammaire & Littérature* (1782-1786) de l'*Encyclopédie Méthodique* (1782-1832).

L'un de nos objectifs communs étant de proposer une solution qui, à travers ses choix méthodologiques et informatiques, soit généralisable, ou tout au moins facilement réexploitable sur d'autres corpus, nous avons donc fait émerger le concept d'un balisage XML "souple" ou "flottant".

La solution des balisages précédents, soit *minimal* (WOOLDRIDGE 1994 et 1996, LEROY-TURCAN, 1996) soit *analytique* (WIONET et TUTIN) n'étant pas pleinement satisfaisante, cette notion de balisage "souple" nous permettait d'allier la souplesse du premier et la richesse du second, tout en respectant la nature non rigide du document et en autorisant ainsi un balisage identifiant les grands champs informationnels et en mettant en évidence un certain nombre d'informations flottant à l'intérieur de ces grands constituants. L'efficacité de notre type de balisage a pu être mise à l'épreuve à travers la réalisation d'un outil d'interrogation de fichiers XML: CorpXML (<http://www.up.univ-mrs.fr/delic/perso/rey/methodique/index.htm>).

Références bibliographiques:

- REY, Christophe. (2004). Analyse et informatisation des articles traitant de l'étude des sons dans le dictionnaire Grammaire & Littérature de Nicolas Beauzée et Jean-François Marmontel, issu de l'*Encyclopédie Méthodique*. Thèse de doctorat. Aix-en-Provence.
- VÉRONIS, J., & IDE, N. (1996). Encodage des dictionnaires électroniques: problèmes et propositions de la TEI. In D. Piotrowsky (Ed.), *Lexicographie et informatique - Autour de l'informatisation du Trésor de la Langue Française*. Actes du Colloque International de Nancy (29, 30, 31 mai 1995), pp. 239-261, Paris: Didier Erudition.
- WIONET Chantal, TUTIN Agnès. (2001). Pour informatiser le Dictionnaire universel de Basnage (1702) et de Trévoux (1704) Approche théorique et pratique. Honoré Champion.
- WOOLDRIDGE, T.R., LEROY-TURCAN Isabelle. (1996). "Les mots-clefs métalinguistiques comme outil d'interrogation structurante des dictionnaires anciens", *Lexicomatique et dictionnaires* (éd. A. Clas, P. Thoiron & H. Béjoint), Beyrouth: FMA & Montréal: AUPELF-UREF, pp. 307-16.
- WOOLDRIDGE, T.R. (1994). "Projet d'informatisation du Dictionnaire de l'Académie (1694-1935)", Actes du Colloque international Le Dictionnaire de l'Académie française et la lexicographie institutionnelle européenne, Institut de France, novembre 1994; (ed. B. Quemada & J. Pruvost), Paris, Champion: 309-20.