

Entre interrogation plein texte et balisage fin, façon TLF: la voie moyenne du Féraud informatisé sous STELLA.

La version Philologique du Dictionnaire critique de la langue française informatisé, mise en ligne en 2003 sur les deux sites miroirs de l'ARTFL de Chicago et de l'ATILF est une base de données textuelles interrogeable plein- texte, les seuls balisages étant référentiels (page, colonne, article) ou typographiques. La version STELLA du Féraud informatisé, qui devrait être en ligne à partir de 2008, offre deux nouveautés qui font son originalité :

- elle offrira un lien constant vers les pages scannées et restaurées du Supplément manuscrit compilé par Féraud dans les années révolutionnaires et jusqu'à sa mort en 1807 et édité en 1987 pour le deux centième anniversaire de la parution du Féraud chez Mossy à Marseille. Ainsi toute l'oeuvre sera disponible ;
- une DTD et des automates de transcodage et de balisage, construits en collaboration entre les laboratoires sous la direction scientifique de Philippe Caron ,offrent un texte XML en cours de relecture par Pierrette Marchaudon de METADIF.

Sans jamais forcer le texte en l'insérant dans un maillage trop fin de balises contraignantes, le projet Féraud 2 analyse les grands formants des articles et isole des objets lexicographiques discrets et largement reconnaissables automatiquement. Le résultat sera une interface utilisateur plus déliée qui autorisera des recherches sectorisées et/ou croisées pour ne retenir que les résultats pertinents.

Néanmoins, cette deuxième version reste marquée par la philosophie constamment présente au coeur de l'entreprise :

- fidélité maximale à l'original ;
- mais recherche de récupération maximale de l'information scientifique dans ce cadre ;
- optimisation de l'effort à l'intérieur d'une entreprise qui reste prudente et modérée dans ses moyens humains. Ainsi nous n'avons pas balisé ni lemmatisé les sources, ne pouvant pas du tout envisager humainement de le faire. Pas de lemmatisation non plus de la mouvance graphique de l'oeuvre, pourtant très variable à cause du parti-pris affiché chez Féraud de moderniser l'orthographe de son temps.

L'utilisateur restera donc acteur de sa recherche et il sera dûment averti des moyens à mettre en oeuvre pour récupérer par lui-même des informations présentes sous des formes graphiques ou textuelles variées.

Equipes et personnes physiques concernées :

Direction scientifique du projet dans son entier : Philippe Caron et Louise Dagenais (Université de Montréal)

Direction scientifique de la version 2 : Philippe Caron

Equipe Féraud à l'ATILF : Jean-Yves Kerveillant (DTD et automates de transcription et balisage), Christiane Jadelot (relecture des balises conjecturales de fin d'exemples et implantation des liens vers le Supplément en mode-image), Jacques Dendien puis Etienne Petitjean (chargement de la base et mise en place de l'interface-utilisateur).

METADIF Cergy: Pierrette Marchaudon (expertise des résultats, relecture des liens et de toute la version transcodée et balisée)

M.S.H.S . de Poitiers: bureau et énergies pour 22 mètres carré, maintenance du matériel informatique; mise à disposition du scanner DIGIBOOK de i2S (Pessac, France) ; formation de Pierrette Marchaudon sur le DIGIBOOK et matériel informatique de bureau.

Laboratoire FORELL : consommables papier et téléphone