

PAROLE DE CHERCHEURS

> Denis Vigier

CollEx-Persée et la recherche en linguistique de corpus

Dans mon laboratoire, projet après projet, nous avons construit un réseau de recherche interdisciplinaire national et international autour du discours encyclopédique en France entre 1751 (parution du vol. I de l'Encyclopédie de Diderot et d'Alembert) et aujourd'hui. Notre objectif linguistique consiste à cibler quels changements énonciatifs ont progressivement modifié la manière de présenter aux lecteurs les connaissances au sein de cette synthèse idéalement totalisante du savoir qu'est une encyclopédie.

Depuis les Lumières se sont succédé en France d'impressionnantes entreprises intellectuelles et humaines destinées à rivaliser avec l'œuvre audacieuse et géniale de Diderot, d'Alembert et de Jaucourt – troisième architecte souvent moins reconnu de l'Encyclopédie. Parmi ces entreprises figure une œuvre majeure publiée entre 1885 et 1902 : La Grande Encyclopédie. Inventaire raisonné des sciences, des lettres et des arts par une société de savants et de gens de lettres. Elle a été réalisée sous la direction de savants parmi lesquels le grand chimiste, M. Berthelot. De nombreuses autres personnalités de tous domaines y ont participé – M. Barrès, R. de Gourmont, M. Mauss, F. Brunetière, F. Buisson, entre autres.

Grâce à CollEx-Persée qui a financé Disco-LGE, cette œuvre jusqu'alors partiellement disponible en version ocrisée sur Gallica, le sera intégralement avec le concours de la BnF. Nous travaillons maintenant à l'encodage du texte suivant les standards internationaux de

la TEI ; après quoi nous enrichirons le texte d'annotations linguistiques (mots, lemmes, catégories). Notre objectif est double : i) livrer à la communauté des chercheurs et des internautes une version structurée et annotée de La Grande Encyclopédie, libre de droits et exploitable sur des plateformes automatiques de navigation et de calcul ; ii) intégrer, à Lyon, dans notre corpus des encyclopédies, ce maillon précieux qui s'inscrit dans le mouvement positiviste puis scientifique du XIXe siècle, afin de mieux cerner la singularité énonciative de cette œuvre dans la grande chaîne historique des encyclopédies françaises.



Démarche modélisante et répliquabilité

Pour atteindre cet objectif, nous mettons au point les premières briques d'une chaîne de traitement qui permettra à terme, à partir de versions ocrisées en format pdf d'ouvrages volumineux, de les structurer (TEI) et les annoter linguistiquement de la façon la plus automatisée possible. Pour les étapes manuelles restantes, nous distribuerons les états des fichiers avant et après intervention afin que nos résultats soient répliquables en repartant de n'importe laquelle de ces étapes intermédiaires. Les logiciels et les développements informatiques engagés dans ce projet respectent les préconisations de la « science ouverte » en utilisant des licences libres et pourront être réutilisés et améliorés par la communauté des chercheurs.

> Denis Vigier

Enseignant-chercheur : Maître de conférences habilité à diriger des recherches, UMR 5191, ICAR, Université Lyon 2 – Lumière ENS de Lyon