

LE TERME CORPUS EN SCIENCES HUMAINES ET SOCIALES, Entretien avec Marie-Luce DEMONET, professeur (Université de Tours)

CORPUS, BREVE HISTOIRE D'UN MOT

Le terme a d'abord désigné l'ensemble des textes juridiques de Justinien, avant d'être utilisé, à la fin du Moyen Âge, pour les œuvres complètes d'un auteur ou pour un ensemble thématique, organisés "comme un corps". L'arrangement est effectué par un disciple ou un éditeur. (60 mots)

A QUOI CELA SERT-IL ?

Les grands corpus épigraphiques, archéologiques et linguistiques ont été les premiers à offrir la possibilité d'exploiter systématiquement des ensembles d'objets ou de textes, ce que les traitements informatiques ont naturellement facilité avec les outils appropriés (balisage, automatisation, requêtes). (60 mots)

QUELS EN SONT LES CRITERES D'ELABORATION ?

Pour qu'il y ait "corpus", il faut que l'ensemble présente une cohérence dans son objet, mais pas nécessairement dans sa forme. Un "corpus d'auteur" peut être constitué d'œuvres littéraires, d'images, de documents divers.

PRESENTATION D'UN CORPUS

Le consortium « CAHIER » (Corpus d'Auteurs pour les Humanités) : Informatisation, Édition, Recherche) est une composante de l'infrastructure CORPUS créée au cours de l'année 2011. (25 mots)
Il a pour objectif de constituer un réseau des projets destinés à fournir des éditions numériques ou des portails consacrés à un auteur, une œuvre, une école, un ensemble textuel, une thématique, souvent à partir de documents originaux mis en ligne. (25 mots)
On y trouve par exemple les projets Flaubert, Stendhal, Montesquieu, Desanti, la *Quête du Graal*, *l'Astrée*, Machiavel, Rabelais, les traités de rhétorique, Horace etc (25 mots)

Chaque consortium établit son propre mode de fonctionnement, qui peut varier selon les objets d'étude, plus ou moins homogènes. Les projets des équipes se trouvant à différents stades d'élaboration, l'infrastructure permet de rendre les données numériques partageables et réutilisables, de tester des outils, de confronter des méthodologies à l'intérieur d'un même consortium ou dans la transversalité des disciplines. (100 mots)

QUELS ASPECTS JURIDIQUES FAUT-IL PRENDRE EN CONSIDERATION ?

La question de la propriété intellectuelle préoccupe tous les consortiums, soit pour la protection des données produites par les équipes, soit pour les documents sous droits que les consortiums sont conduits à utiliser et qui ne sont pas tous dans le domaine public (images, enregistrements, œuvres ou éditions récentes). Des groupes de travail ont commencé à se réunir pour s'informer sur la jurisprudence et rédiger des recommandations. (70 mots)

1 GRANDE IMAGE

Ou

2 PETITES IMAGES

A PROPOS DE L'INFRASTRUCTURE DE RECHERCHE CORPUS

L'IR CORPUS réunit, sous la responsabilité de Laurent Douset (Université d'Aix-Marseille/EHESS), des consortiums disciplinaires qui se sont organisés eux-mêmes. Actuellement, quatre sont labellisés: anthropologie, corpus écrits, corpus oraux, corpus d'auteurs, et plusieurs autres sont en cours de constitution: musicologie, archéologie, histoire ancienne, histoire moderne, géographie... (40

Pour en savoir plus...

<http://www.corpus-ir.fr/>

<http://www.ilf.cnrs.fr/>

www.cahier.paris-sorbonne.fr