

Atelier « Édition analytique »

CAHIER / Équipe ICAR - LinCoBaTo / Labex ASLAN

Dates : 30 juin – 4 juillet 2014

Lieu : ENS de Lyon, site Buisson (IFÉ)

Salle de réunion n° 2

19 allée de Fontenay

Lyon 7e

Plan d'accès : <http://ife.ens-lyon.fr/ife/institut/informations-pratiques>

Objectif : Initier les participants à une chaîne éditoriale basée sur une démarche analytique

Résumé :

A partir d'exemples d'usages analytiques de corpus littéraires (caractériser le style d'un auteur, un genre littéraire, retrouver des motifs), nous présenterons les principes méthodologiques et les outils de travail (préparation, analyse et diffusion de corpus) qui permettent de produire et d'utiliser des éditions :

- de bonne qualité philologique et esthétique ;
- adaptées à des usages analytiques ;
- portant des annotations linguistiques et philologiques ;
- pouvant être intégrées dans de grands corpus ou des bibliothèques numériques.

Programme

Lundi 30 juin

10h - 12h

- Réunion du comité de pilotage du consortium CAHIER

12h - 13h30

- Déjeuner

13h30 - 13h45

- Accueil des participants

13h45 - 14h30

- Ouverture de l'atelier (A. Lavrentiev, S. Heiden)
- Tour de table des participants

14h30 - 15h30

- Introduction à la textométrie, présentation de TXM (B. Pincemin)

15h30 - 16h

- Pause café

16h - 16h45

- Introduction à la philologie numérique progressive dans l'environnement TXM : "établissement du texte/corpus pour le logiciel" (S. Heiden)
 - pour l'analyse : lecture, navigation, recherche, accès en ligne / téléchargement
 - progressif : presse-papier / texte brut / XML / TEI ↔ modèle de données
 - multi-modal : texte / image / vidéo / audio
 - (multilingue)

16h45 - 17h30

- Présentation de recherches linguistiques en diachronie du français et démonstration du portail de la Base de français médiéval et de l'édition de la *Queste del saint Graal* (C. Guillot & A. Lavrentiev)

Mardi 1 juillet

9h - 9h45

- Présentation de recherches stemmatologiques outillées (J.-B. Camps)

9h45 - 10h45

- Présentation de recherches sur un corpus de romans (M. Bernard)

10h45-11h15

- Pause café

11h15 - 12h30

- Modèle de données pour l'analyse avec TXM créé par l'import des sources du corpus (S. Heiden & A. Lavrentiev)
 - unités textuelles et leurs métadonnées
 - structures internes aux textes et leurs propriétés
 - unités lexicales et leurs propriétés (tokenisation, annotation par des outils de Traitement Automatique de la Langue)
 - plans textuels et éléments « hors texte »
 - textes multilingues et corpus parallèles
 - éditions : facsimilé || transcription [page]

12h30 - 14h

- Déjeuner

14h - 15h30

- Numérisation et édition avec un logiciel traitement de texte, conversion en XML (F. Glorieux & V. Jolivet)
 - chaîne Algone
 - outils TEI (Oxgarage)

15h30 - 16h

- Pause café

16h - 17h30

- Numérisation et édition avec un logiciel traitement de texte, conversion en XML (suite) (A. Lavrentiev, S. Heiden, M. Decorde)
 - pratiques de la BFM (éditions standard / synoptiques / multi-facettes)
 - outil TXM (Import ODT et lemmatization automatique avec TreeTagger)

Mercredi 2 juillet

9h - 9h45

- Présentation du “Corpus 14” (G. Luxardo)

9h45 - 10h30

- Présentation du corpus-exemple EA2014 et des corpus des participants
- Édition XML (nettoyage, enrichissement) (A. Lavrentiev, S. Heiden)
 - logiciel Oxygen
 - préparation à la segmentation lexicale
 - enrichissement des métadonnées

10h30 - 11h

- Pause café

11h - 12h30

- Édition XML (suite)

12h30 - 14h

- Déjeuner

14h - 17h30

- Atelier d’import progressif du corpus exemple EA2014 et de corpus de participants dans TXM (txt, xml/pb, xml/tei-p, xml/w) (S. Heiden)
- Stylage des éditions TXM (A. Lavrentiev)

Jeudi 3 juillet

9h - 12h30

- Publication du corpus exemple EA2014 dans un portail TXM (A. Lavrentiev)
- Analyse en ligne et en local du corpus avec TXM (B. Pincemin)

12h30 - 14h

- Déjeuner

14h - 17h30

- Publication et analyse avec les outils de Grenoble (T. Lebarbé, projet Les Manuscrits de Stendhal)
- Aspects juridiques et éditoriaux (D. Pierrot)
 - Propriété intellectuelle (texte, lexique, balisage)
 - Licences, conditions de diffusion, etc...

Vendredi 4 juillet

9h - 10h30

- Publication avec le kit SynopsX (J.-P. Magué et M. Ingarao)

10h30 - 11h

- Pause café

11h - 12h

- Bilan de l'atelier

Contact : Alexei Lavrentiev

alexei.lavrentev@ens-lyon.fr

tél. 04 37 37 63 10 / 06 67 34 33 95