

École doctorale Lille – Nord de France, spécialité « Études européennes »

Séminaire doctoral transfrontalier « Vie des mots, dynamique des langues : l'Europe, du Moyen Âge à l'époque moderne »

adossé au projet ANR-DFG Eurolab (<http://www.eurolab.meshs.fr>)

Contact : matthieu.marchal@meshs.fr

Langues, glossaires et bases de données informatiques

21 mars 2011

Université Lille 3, Maison de la Recherche, salle 008

10h15 : Accueil des participants par Marie-Madeleine CASTELLANI, directrice d'ALITHILA.

Bases de données lexicographiques informatisées

Séance présidée par Nathalie GASIGLIA (Lille 3, SAVOIRS TEXTES LANGAGE)

10h30 : Pascale RENDERS (Université de Liège), « L'informatisation par rétroconversion du FEW (*Französisches Etymologisches Wörterbuch*) ».

11h00 : Claire DUCARME (Université de Liège), « Les *Curiositez françoises* d'Antoine Oudin : un trésor lexicographique. Création d'une base de données au service de l'analyse linguistique ».

11h30 : Esther BAIWIR (Université de Liège), « On veut des chiffres ! Quelques dangers d'une exploitation non raisonnée des ressources lexicographiques informatisées ».

Discussion.

12h30-13h30 : déjeuner.

Bases de données linguistiques et rédaction de glossaires

Séance présidée par Marie-Madeleine CASTELLANI (Lille 3, ALITHILA)

13h30 : Marie STEFFENS (Université de Liège), « Dictionnaires et structuration sémantico-lexicale du français : l'antonymie ».

14h00 : Jennifer GABEL (Université de Aberystwyth), « La structuration sémantique du glossaire de texte et du dictionnaire : étude contrastive ».

14h30 : Marc KIWITT (Université de Heidelberg), « Le traitement des termes scientifiques dans le DEAF (*Dictionnaire Étymologique de l'Ancien Français*) et le DEAF électronique : quelles leçons en tirer pour la rédaction de glossaires ? »

Discussion suivie d'une pause café.

salle A5 206

Atelier pratique pour les doctorants

Séance menée par Nathalie GASIGLIA

16h00-18h00 : Nathalie GASIGLIA (Université Lille 3) propose aux doctorants et étudiants intéressés une initiation à la structuration de documents en XML pour l'extraction d'informations à partir d'articles du *Dictionarium latinogallicum* de Robert Estienne (1531).

Résumés des communications

Pascale RENDERS (Université de Liège), « L'informatisation par rétroconversion du FEW (*Französisches Etymologisches Wörterbuch*) »

L'informatisation du FEW est en cours de réalisation. Nous résumerons dans cette communication les étapes du projet et le résultat attendu, en abordant plus spécifiquement deux questions, qui concernent toute base informatisée : quelle utilisation est voulue du dictionnaire informatisé, et en quoi cet objectif influence-t-il la modélisation de l'objet lexicographique ?

Claire DUCARME (Université de Liège), « Les *Curiositez françoises* d'Antoine Oudin : un trésor lexicographique. Création d'une base de données au service de l'analyse linguistique »

Rédigées en 1640, les *Curiositez françoises* d'Antoine Oudin constituent un recueil précieux du vocabulaire de la langue non normée de la première moitié du 17^e siècle. L'auteur nous livre un dictionnaire comportant quelque 8000 entrées, qui n'ont, jusqu'à ce jour, pas été traitées systématiquement. Afin de pallier ce manque et de pouvoir réaliser des analyses linguistiques pertinentes sur l'ensemble du corpus, le balisage informatique de chaque article du dictionnaire semble constituer une étape nécessaire. Celui-ci ne se suffit pas à lui-même, mais doit être appréhendé comme un moyen au service de l'analyse linguistique. C'est pourquoi, lors de ce séminaire, après avoir présenté brièvement l'auteur et son œuvre, nous exposerons les réflexions linguistiques qui sous-tendent notre balisage informatique (projet en cours). Les apports de l'informatisation seront mis au jour et illustrés de manière théorique au moyen de plusieurs curiosités.

Esther BAIWIR (Université de Liège), « On veut des chiffres ! Quelques dangers d'une exploitation non raisonnée des ressources lexicographiques informatisées »

Depuis quelques décennies, divers projets d'informatisation d'ouvrages lexicographiques fleurissent ; la plupart permettront de nouveaux types de questionnement, impossibles ou fastidieux à effectuer dans les ouvrages imprimés. Néanmoins, cette vague de nouveaux outils s'accompagne parfois de pratiques dangereuses. En effet, la facilité d'accès des matériaux peut laisser croire que le travail d'analyse a été effectué par la machine et que les résultats si peu chèrement obtenus sont définitifs. Au travers d'un exemple - le cas des arabismes dans le *Trésor de la Langue française informatisé* (TLFi), nous proposerons la prudence et la réflexion comme meilleures armes contre ces abus.

Marie STEFFENS (Université de Liège), « Dictionnaires et structuration sémantico-lexicale du français : l'antonymie »

Après avoir rappelé brièvement l'importance de l'antonymie dans la structuration du lexique français, nous exposerons les résultats de notre analyse des renvois antonymiques dans deux dictionnaires généraux, le *Grand Robert de la langue française* et le *Trésor de la langue française*. Nous envisagerons ensuite des pistes pour améliorer la représentation structurée du lexique dans les dictionnaires, notamment grâce aux outils informatiques.

Jennifer GABEL (Université de Aberystwyth), « La structuration sémantique du glossaire de texte et du dictionnaire : étude contrastive »

La communication s'attachera à comparer la structuration sémantique générale d'un article de dictionnaire et celle exigée pour une entrée d'un glossaire d'édition de texte.

Marc KIWITT (Université de Heidelberg), « Le traitement des termes scientifiques dans le DEAF (*Dictionnaire Étymologique de l'Ancien Français*) et le DEAF électronique : quelles leçons en tirer pour la rédaction de glossaires ? »

Après une brève présentation du *Dictionnaire Étymologique de l'Ancien Français* (DEAF), et de sa version électronique qui est actuellement mise en ligne de manière progressive, la communication abordera plusieurs aspects du traitement des termes scientifiques dans le DEAF et dans le DEAFél, et notamment des questions relevant de la définition et de la datation. La méthode lexicographique adoptée par le DEAF sera confrontée au traitement de termes scientifiques dans d'autres dictionnaires historiques et dans des glossaires d'éditions de textes, pour examiner ainsi les outils qu'elle peut fournir aux éditeurs de textes scientifiques médiévaux.

Atelier pratique pour les doctorants

Nathalie GASIGLIA (Université Lille 3), « Initiation à la structuration de documents en XML pour l'extraction d'informations à partir d'articles du *Dictionarium latinogallicum* de Robert Estienne (1531) »

L'observation d'une sélection d'articles du *Dictionarium latinogallicum* de Robert Estienne (1531) nous permettra d'identifier la nature de différents types d'éléments informationnels qu'ils comportent et des indices typographiques ou positionnels qui permettent de délimiter les segments de texte relevant de chacun de ces types (et par la suite de les baliser en XML - *eXtensible Markup Language*).

Partant de ces observations, nous enrichirons le texte de ces articles de délimitateurs (des balises XML) pour chacun de ces segments, en procédant par étapes successives (de l'application de styles dans un traitement de texte à l'affinement, dans un éditeur XML, du balisage issu des styles précédemment appliqués).

Le document structuré en XML obtenu sera enfin exploité pour extraire des sélections d'informations susceptibles d'être prises en compte dans le cadre d'analyses du contenu textuel des articles. Sans entrer dans le détail de ces analyses, nous verrons comment manipuler le document structuré avec des petits programmes conçus spécifiquement pour lui (des *transformations* écrites en XSL - *eXtensible Stylesheet Language*).

Durant l'atelier, la sélection d'articles sera traitée du texte nu au document structuré dont seront extraites des informations ciblées, de manière telle que les participants les plus autonomes informatiquement pourront reproduire le mode de traitement proposé sur d'autres documents et que les autres auront au moins réalisé par eux-mêmes une partie des manipulations présentées et en auront perçu le principe et les potentialités.