

PRES Université Paris Est  
Créteil Marne la Vallée  
Ecole Doctorale LSHSS (E.D.403)

**Séminaire d'informatique textuelle 2009-2010.**  
Samedi 9h 30 - 16h30 ; salles informatiques CMC 227 et 342

### Traitement automatisé des données textuelles Problèmes de sémantique et méthodes informatiques pour les sciences humaines

De nombreux outils informatiques multimédias sont désormais proposés aux chercheurs en sciences humaines et sociales pour le recueil, le traitement des données, et la présentation des résultats). En liaison avec ces applications (présentées dans le cours de JM Leblanc), le séminaire doctoral d'informatique textuelle propose des outils et des problématiques qui permettent d'approfondir les champs de recherche en sémantique, syntaxe, morphologie, énonciation, argumentation, ou de reformuler des réflexions méthodologiques classiques comme les diverses typologies textuelles, l'articulation, la conservation et la transmission des textes, des sons et des images.

Les six séances du séminaire présentent une série de problématiques théoriques en les associant à la présentation d'outils pratiques, à partir de l'analyse de corpus textuels sociaux et politiques. Chaque séance comprend la présentation et la discussion d'une problématique et la mise en oeuvre pratique des technologies qui l'illustrent. Des exercices d'analyse de corpus permettent de tester des fonctionnalités informatiques, et de répondre aux questions pratiques. Déroulement des séances mensuelles: 3 heures de cours; 3 heures de TD.

Le séminaire s'adresse aux doctorants de l'Ecole doctorale « Cultures et sociétés », aux étudiants de Master de Paris 12, et des autres universités dans la limite des places disponibles, titulaires du C2i ou qui suivent le cours des outils multimédias. Il nécessite une inscription préalable.

#### Calendrier des séances.

1. Samedi 24 octobre 2009: **Lexicologie historique, étymologie sociale, lexicométrie.** Bases textuelles littéraires, médiatiques, historiques ; logiciels et filtres simples d'interrogations.
2. Samedi 21 novembre 2009: **Outils grammaticaux, sémantiques, stylistiques.** Pourquoi et comment étiqueter, quantifier, décrire les textes (Logiciels Cordial, Tree Tagger, Hyperbase). Rechercher des formes, des contextes, des schémas syntaxiques, des propriétés sémantiques dans un corpus textuel.
3. Samedi 19 décembre 2009 : **Analyse de contenu et analyse de discours : deux approches de la discoursivité.** Construire et analyser des corpus lexicométriques. Fonctions documentaires et fonctions statistiques (lexico3); typologies génériques Hypothèses, corrélations et causalités interprétatives. Analyses lexicométriques et analyses de contenu (Tropes).
4. Samedi 23 janvier 2010 : **Questions de philologie numérique.** Recueil et établissement des textes ; identification, standardisation et format ; conservation, préservation et mise à disposition ; cumul et comparaison des données et de résultats
5. Samedi 20 février 2010 : **Expérimentations textométriques,** distance intertextuelle. topographies textuelles, Cooccurrences dans les énoncés et mondes sémantiques (Alceste). ). Voisinages lexicaux ; cooccurrences dans les textes ; tableaux de propriétés textuelles (Weblex et les grammaires MPC).
6. Samedi 20 mars 2010: **Cartographies de textes. Visualisation des résultats** (Hyperbase, Weblex, Astartex, 3D, et autres outils). Séance organisée en relation avec une demi-journée d'étude du CEDITEC (vendredi 19 mars 2010).

Site de recherche, de formation, et de ressources : <http://textopol.org>

[fiala@univ-paris12.fr](mailto:fiala@univ-paris12.fr) Linguistique et informatique, Université Paris 12 Val de Marne 94010 Créteil