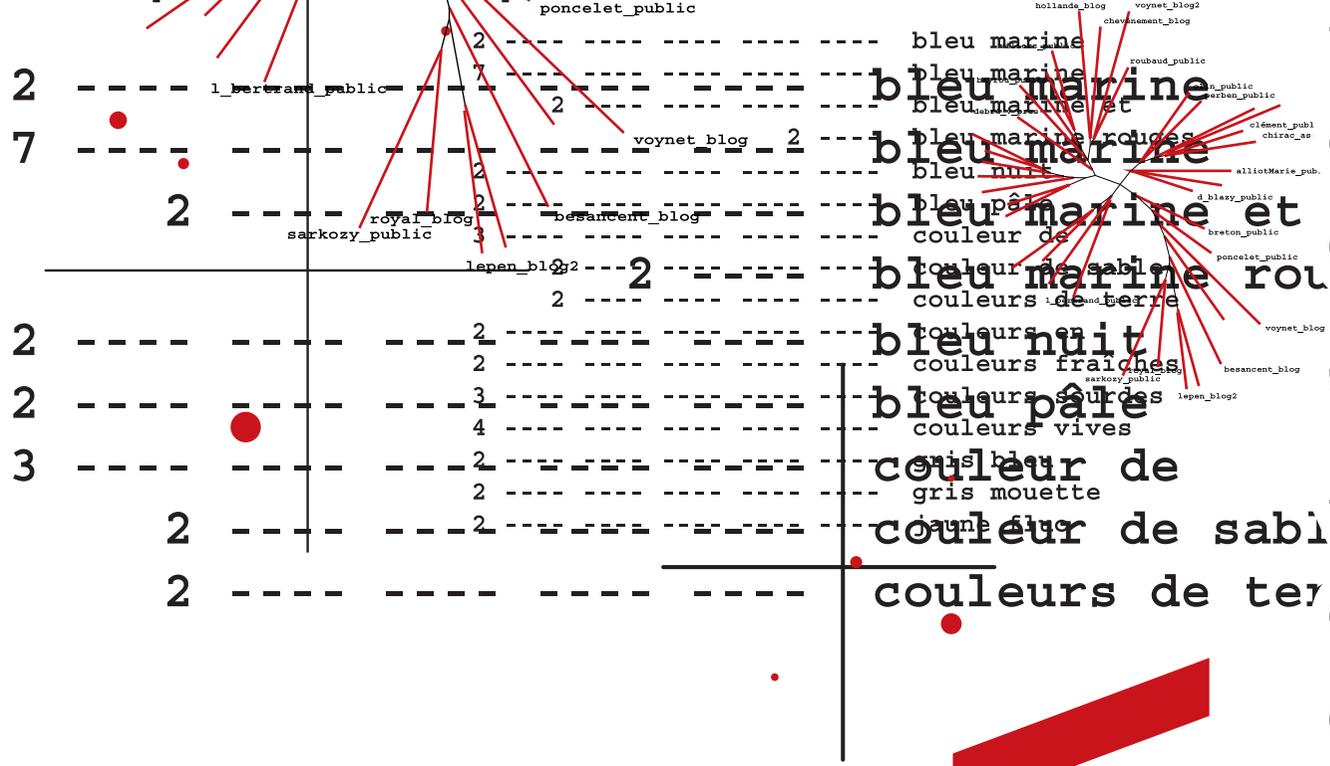


Cette formation doctorale constitue une introduction à l'analyse quantitative et qualitative des corpus textuels et à la sémantique discursive, qui s'adresse à tous les doctorants de SHS dont la recherche porte sur l'analyse de textes littéraires, le dépouillement d'enquêtes et de formulaires, l'analyse des discours sociaux et politiques. Chaque séance, outre la prise en main d'un logiciel et des problématiques théoriques qui s'y rattachent comprend des exercices pratiques en linguistique de corpus et en statistique textuelle qui permettent de tester les fonctionnalités, de répondre aux questions pratiques. Les doctorants et les étudiants en Master 2 de l'UPEC, ou d'autres universités dans la limite des places disponibles, peuvent y trouver des réponses concrètes dans leurs recherches individuelles ou collectives.

Les résultats produits par les outils textométriques intègrent désormais divers modules de mises en forme des données et des résultats demandant aux utilisateurs une maîtrise des outils de codage et de visualisation et une réflexion ergonomique. La formation comprend la mise en oeuvre d'outils s'appuyant sur les compétences bureautiques universitaires nécessaires à la constitution et gestion des corpus (recueil et documentation des données, exploitation et construction de bases de données).

n , la liberté , la société , et donné à ce re , elle a miné la société sous tous ses nombreux cas où la société civile ne peut ' état mais par la société civile . cela n ut - on dire que la société civile s' est e , au motif que la société spectaculaire télévisuelle si la société se délite , ell du terrorisme , la société civile italienn , niché dans notre société civile ? peut - ttoisation de notre société . nous ne sauri n radicale de notre société qui , en se tra réveille , et notre société illuminée s' é orial - la première société avancée qui , e



Formation Doctorale

APPROCHES TEXTOMÉTRIQUES, ERGONOMIES NUMÉRIQUES

2016-2017

Web

http://ceditec.u-pec.fr
http://textopol.u-pec.fr

Contact

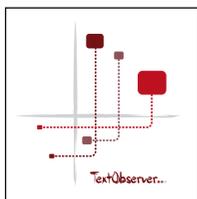
jean-marc.leblanc@u-pec.fr

Localisation

Salle multimédia, I2-317
bâtiment i - Campus centre (CMC)
61 avenue du Général de Gaulle
94010 Créteil Cedex

Horaire

samedi 10h-13h
6 séances - 20 heures



Stages de formation à TextObserver

Préparation de corpus, prise en main sur les corpus des participants.

En semaine, deux séances (dates à préciser).

5 novembre 2016 : Introduction - Approches textométriques des discours.

Présentation et typologie pratique de logiciels standard et d'analyseurs de données textuelles. Options théoriques, principes méthodologiques, limites interprétatives.

Initiation à **Lexico 3 / 5** :

- Fonctions documentaires, décomptes statistiques, modèles probabilistes.
- Distributions statistiques, distributions linguistiques.
- Analyse factorielle des correspondances, spécificités, fréquences...

10 décembre 2016 : Présentation et prise en main de TextObserver [10h-16h]

- Expliciter l'analyse factorielle des correspondances
 - Analyser la variation lexicométrique
 - Introduction aux opérations de catégorisation
- http://textopol.u-pec.fr/textobserver
- Recension de corpus et balisage semi-automatisé : présentation de la base Textopol

21 janvier 2017 : Des corpus textuels aux corpus multimodaux (annoter, catégoriser, étiqueter, visualiser, interpréter).

- Transformer des textes pour les soumettre à des traitements automatisés. Repérer les régularités d'un document, extraire des motifs textuels, concaténer des fichiers.
- Forme graphique, catégories morphosyntaxiques, univers sémantiques.
- Prise en main de quelques catégoriseurs, évaluateurs, étiqueteurs (**cordial, treetager, tropes**)
- Base de données textuelles et outils de constitution et de balisage de corpus.
- Outils de visualisation, de caractérisation de corpus : **Gephi, R, Xlstat, Textstat.**

25 février 2017 : De la lexicométrie au traitement automatique des langues (TAL) [10h-16h]

- P. Gambette (LIGM-MLV) : Les nuages arborés dans **TextObserver** et **Treecloud**.
- C. Martineau (LIGM-MLV) : Présentation et prise en main du logiciel **UNITEX**.

PROGRAMME 2015-2016

PROGRAMME 2015-2016

25 mars 2017 : Corrélations et causalités interprétatives. Expérimentations, distance intertextuelle et voisinages.

Distances, cooccurrences, voisinage
Présentation et prise en main d'**Hyperbase** : de la lexicométrie à la stylométrie.
Présentation d'**Hyperbase** en ligne.

22 avril 2017 : De la textométrie à l'analyse des données, quels outils pour quels usages?

Cooccurrences généralisées et mondes lexicaux : analyses comparées **Alceste** et **Iramuteq**.
Outils de dépouillement d'enquêtes et de formulaires : **Sphinx, Modalisa**.
Ontologies et mondes sémantiques (**Tropes, Alceste, Astartex**)
Quantifier les données en sciences sociales : **Prospero, Nvivo**

Séances spécifiques

Certaines séances spécifiques sont organisées sur la journée entière (10h-16h).
Le samedi après-midi sera consacré aux questions des participants (sur rendez-vous).

Courant 2016 - 2017 :

3e journée de présentation des outils de traitements de corpus textuels

Gramlab/Unitex, Treecloud, TextObserver/WebObserver, Cortext...
Echanges, présentations, ateliers. Outils développés par les équipes de recherche de Paris-Est.
Le programme détaillé sera mis en ligne sur Textopol.

