



ÉCOLE D'ÉTÉ INTERNATIONALE : 2 AU 6 MAI 2016
FORMATION DE MASTER ET DE DOCTORAT : MÉTHODES ET OUTILS NUMÉRIQUES DE
L'ANALYSE DE DISCOURS EN HISTOIRE
Université du Québec À Montréal, Université de Lille (IRHiS umr-CNRS 8529), Université de Paris 1
(LAMOP umr-CNRS 8589)

L'école d'été se déroule sur 5 jours. L'objectif en est de permettre aux étudiants de master et de doctorat, ainsi qu'aux collègues des trois universités de se former aux enjeux et aux notions de base de la lexicométrie (lexicométrie, textométrie, logométrie, corpus, lemmatisation...), de l'exploration des textes et de la statistique textuelle (descriptive, multivariée, exploratoire). Elle s'organise sur un schéma simple : des conférences méthodologiques le matin, des prises en main logiciels et des expérimentations l'après-midi sur des corpus proposés par les formateurs ou par les participants.

Comité scientifique : Pascal Bastien, Benjamin Deruelle, Stéphane Lamassé
Comité d'organisation : Pascal Bastien, Benjamin Deruelle, Stéphane Lamassé

Recherche et formation conjointes :

Il s'agit de former les participants venus des trois universités, non pas à l'informatique, mais à l'utilisation scientifique des méthodes de recherche que l'informatique met à la disposition des chercheurs. Ces méthodes intéressent non seulement les historiens, mais toutes les disciplines des SHS (philosophie, sociologie, psychologie, littérature, civilisation...). Pourtant, elles restent le plus souvent ignorées, voire parfois évitées au prétexte des savoir-faire « techniques » qu'elles imposent de maîtriser. Cet état de fait est d'autant plus paradoxal que les logiciels sont de plus en plus nombreux et de plus en plus simples d'utilisation. Le risque devient alors celui d'un usage en dehors de toute formation méthodologique et épistémologique.

L'objectif de cette formation est donc de dispenser un enseignement à la fois, épistémologique, méthodologique et pratique sur l'usage d'une des méthodes les plus prometteuses dans le domaine de l'analyse du discours et des représentations : la lexicométrie. Elle propose ainsi de s'attacher à des principes et notions fondamentales tels que les enjeux de l'utilisation de l'informatique pour le traitement des sources ou des données d'enquêtes, des règles de la constitution de corpus (métasource), des échelles d'analyse (de la forme au thème), de l'usage de la lemmatisation, ou encore de la production du chiffre et de l'interprétation des résultats statistiques (univers statistiques, critiques des données).

Cette formation se fait ainsi non seulement au travers de la présentation de l'état de l'art et d'études désormais considérées comme exemplaires, mais d'expérimentations menées à partir d'exemples introduits par les formateurs et de corpus proposés par les participants. La semaine se déroulera ainsi autour de conférences méthodologiques, assurées par des spécialistes d'horizons divers (historiens, linguistes, informaticiens), suivies de discussion le matin, et d'ateliers de prise en main des logiciels et d'expérimentations l'après midi. Aucun « pré-requis » n'est indispensable pour suivre cette formation, sauf, peut-être, de la curiosité, un esprit ouvert à l'usage de l'informatique pour la recherche, et une connaissance élémentaire de l'informatique et de la bureautique. Il est souhaitable que les participants aient un corpus déjà constitué, mais un corpus expérimentable sera proposé le cas échéant.

Retombées attendues : rayonnement, partenariat création et renforcement des réseaux

Les retombées attendues sont ainsi multiples. Il s'agit d'abord bien sûr de former des étudiants de master, de doctorat et des collègues. Toutefois, cette école sera aussi l'occasion d'exporter les savoir-faire de nos universités et donc de participer à leur rayonnement et au renforcement de ses liens avec le réseau des universités publiques du Québec, qui sont assez peu développées en histoire.

Le projet propose, en outre, d'intégrer des doctorants des différentes universités partenaires. Pour eux et au-delà de l'aspect formation, le projet est une belle occasion de se créer des réseaux internationaux, de faire connaître leur recherche en Amérique du nord, ainsi que de découvrir une culture et une façon de faire de l'histoire, d'enseigner et d'être formés.



La rencontre des étudiants, enseignants et formateurs doit aussi être l'occasion de créer des synergies, des liens et des échanges scientifiques et pédagogiques entre les universités partenaires. Et ainsi, d'organiser une seconde édition de cette école sur des thèmes définis à partir de cette première expérience, en France dans les deux années qui suivent cette première édition.

Grands axes du programme :

I. De la source à la metasource : stocker et manipuler l'information

- réflexions épistémologiques sur les enjeux et les pratiques de la lexicométrie
- notion de corpus
- Les enjeux de la lemmatisation (textes anciens et contemporains)

II. L'exploitation des textes

- Initiation à la constitution et à l'analyse de corpus (principes, segmentation, partition)
- Initiation à l'exploitation de corpus
 - o Première approche descriptive : repérage des formes et des spécificités
 - o La diachronie : l'évolution sémantique
 - o Analyse multivariée (analyse factorielle, méthodes post-factorielles)
 - o Repérage des thématiques
 - o Des co-occurrences aux poly co-occurrences
- codages et recodages de l'information textuelle

PROGRAMME

Premier jour — La linguistique de corpus

Tour de table (9h)

Conférence de Stéphane Lamassé (Paris 1, LAMOP) et Benjamin Deruelle (Lille SHS, HiRIS), *Principes et expériences de la linguistique de corpus en histoire : historiographie et méthode* (10h-12h)

TP (13h30-15h, salle informatique)

- Présentation des corpus des participants
- Présentation et évaluation des logiciels les plus courants (TXM, Lexico3, Iramuteq, Alceste, R)

TP (15h30-17h30, salle informatique)

- Installation des logiciels utilisés durant la formation (clés USB, Lexico3, Coocs 2, Iramuteq, Alceste, R)
- Gestion des fichiers (partition, balisage, expressions rationnelles)

Deuxième jour — Exploration des textes et mesures du langage

Tour de table (9h) : Débriefing de la veille et tour de table autour des lectures fournies aux participants

Lectures commentées :

- GUILLHAUMOU, Jacques, « Le corpus en analyse de discours : perspective historique », dans *Corpus*, n°1 : Corpus et recherches linguistiques, novembre 2002, disponible sur le [www](http://www.corpus.revue.org), mise en ligne le 15 décembre 2003, [<http://corpus.revue.org>], 28 p.
- MAYAFFRE, Damon, « Les corpus politiques : objet, méthode et contenu. Introduction », dans *Corpus*, n°4 : Les corpus politiques : objet, méthode et contenu, décembre 2005, disponible sur le [www](http://www.corpus.revue.org), mis en ligne le 5 septembre 2006, [<http://corpus.revue.org>].

Conférence de Élias Rizkallah (UQAM), *Du corpus à l'exploration textuelle* (10h-12h)

TP (13h30-16h30, salle informatique) : Études de cas

TP optionnel (16h30-17h30, salle informatique)

- Mise en œuvre de Treetagger pour les langues anciennes et contemporaines

Troisième jour — Analyse multivariée et méthodes de classification

9h : Débriefing de la veille et tour de table autour des lectures fournies aux participants

Lectures commentées :

- COURDESSE, Lucile, « Blum et Thorez en mai 1936 : analyses d'énoncés », *Langue française*, 9 (1971), pp. 22-33. Disponible en ligne [http://www.persee.fr/doc/lfr_0023-8368_1971_num_9_1_5569]
- SALEM, André, « Segments répétés et analyse statistique des données textuelles », dans *Histoire et Mesure*, I (1-2), 1986, pp. 5-28.

Conférence de Marie Cottrell (PR Université Paris 1), *Les méthodes multidimensionnelles appliquées à l'analyse textuelle (analyse factorielle, classification hiérarchique, cartes de Kohonen)* (10h-12h)

TP (13h30-17h, salle informatique) : Étude de cas (chansons de geste, arithmétiques médiévales)

- La norme et les écarts
- Des formes aux étiquettes
- Identifier des thématiques

Quatrième jour — Le système des mots

9h : Débriefing de la veille et tour de table autour des lectures fournies aux participants (9h)

Lectures commentées :

- PROST, Antoine, « Les mots », dans *Pour une histoire politique*, René Rémond (dir.), Paris, Seuil, 1988, pp. 255-286.

Conférence Stéphane Lamassé (Paris 1, LAMOP) et Benjamin Deruelle (Lille SHS, HiRIS), *Les cooccurrences, les poly-occurrences et les trames en analyse du discours* (10h-12h)

TP (13h30-16h30, salle informatique) : Étude de cas (corpus contemporain)

- Analyse des cooccurrences (cooccurrences, polycooccurrences, trames)

Cinquième jour — La carte et les mots

9h : Débriefing de la veille et tour de table autour des lectures fournies aux participants (9h)

Lectures commentées :

- MAYAFFRE, Damon, « L'Herméneutique numérique », HAL, en ligne [<https://hal.archives-ouvertes.fr/hal-00586512>]
- Magali GUARESI, « La forme député. L'approche co-occurrence, un bond qualitatif ? L'environnement lexical du lemme "député" dans les Professions de foi des candidates à la députation (1958 – 2002) » *Corela* [En ligne], HS-15 | 2014, mis en ligne le 15 octobre 2014, consulté le 11 mars 2016. URL : <http://corela.revues.org/3586>

Conférence de Gaëtan Bonnot (Paris 1, PIREH) et Léo Dumont (Paris 1, PIREH, Centre d'histoire du XIXe siècle), *Des cartes et des mots : études de cas*

TP (13h30-15h30, salle informatique)

- Projeter des mots sur une carte.

Feedback (15h30-16h30, salle de conférence)