



Centre universitaire Ahmed Zabana de Relizane

L'équipe de recherche CNEPRU du projet « Le manuel scolaire algérien: Analyse lexicométrique des manuels de la langue française (de la 3ème A.P à la 3ème A.S) »

En partenariat avec

Le laboratoire « Langue et communication » du CU de Relizane

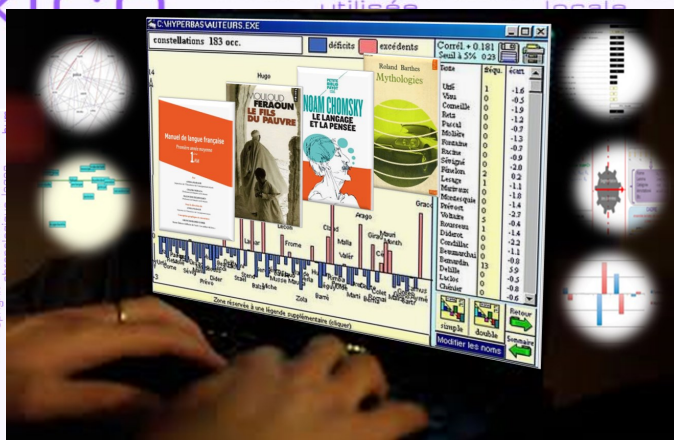
Organise

## 4<sup>èmes</sup> JOURNÉES REGIONALES DE FORMATION EN LOGOMETRIE

### Formation destinée aux enseignants

**Analyse informatisée et interprétation des discours**  
en sciences du langage, didactique des langues et sciences des textes littéraires  
(Initiation au logiciel Hyperbase)

Le mercredi 14 et le jeudi 15 mars 2017



vocabulaire richesse  
recherche  
Logométrie  
sciences lexique du par disc  
globale Weblex les de assistée Corp  
dans Spad-t grande de fréquences Corp  
Alceste DTM sociales Lexic  
calcul Spynx tout texte métrie le Outils

Logométrie  
recherche  
et discours  
sciences lexique du par disc  
globale Weblex les de fréquences Corp  
calcul Spad-t grande de fréquences Corp  
Alceste DTM sociales Lexic  
calcul Spynx tout texte métrie le Outils

## Introduction

## Objectifs des deux journées

Livres, journaux, magazines, revues scientifiques et autres corpus textuels, sont de plus en plus disponibles sous forme numérisée. Il suffit d'un clic pour accéder à une extraordinaire masse de textes couvrant tous les domaines. Ainsi, ce qui n'était hier encore qu'un rêve est aujourd'hui une réalité à portée de la main.

Aujourd'hui donc qu'une grande partie des productions langagières est numérique et est accessible souvent gratuitement par Internet, il serait dommage que nous ne profitions pas de l'apport considérable que nous offre un traitement assisté par ordinateur de ces corpus.

En effet, il serait pour le moins incompréhensible de continuer d'appréhender les textes volumineux, c'est-à-dire composé de centaines de milliers d'occurrences, de la manière traditionnelle, qui exige énormément de temps et qui souvent ne nous permet pas une analyse complète et rigoureuse, alors que nous avons aujourd'hui la possibilité de le faire avec des logiciels de logométrie très performants et riches en informations, qui sont mis à notre disposition le plus souvent gratuitement.

Oui, à présent, il ne nous est plus permis d'ignorer ces puissants outils informatiques. Pour s'en convaincre, songeant par exemple au temps et au travail qu'exigerait une étude des manuels scolaires, du primaire au secondaire <sup>(1)</sup>; l'œuvre complète de l'écrivain Mouloud Mammeri ou encore un corpus constitué des discours de notre président Abdelaziz Bouteflika.

(1) Voir à ce titre, notre projet CNEPRU « Le manuel scolaire algérien: « Analyse lexicométrique des manuels de la langue française (de la 3ème A.P à la 3ème A.S) » sur notre blog: <https://logometrie-algerie.jimdo.com/>

A l'issue de ces deux journées de formation, l'enseignant-chercheur sera capable de :

- Comprendre les fondements théoriques et statistiques des analyses logométriques
- Utiliser concrètement le logiciel Hyperbase

## La logométrie et les sciences humaines et sociales

Considérée comme un prolongement de la lexicométrie (mesure du lexique) et de la textométrie (mesure du texte), la logométrie (logos = discours ; métrie = mesure) est une méthode d'analyse et d'interprétation des discours utilisée dans les sciences humaines et sociales

Cette méthode, assistée par ordinateur, permet de décrire qualitativement et quantitativement un corpus numérique même de grande taille. Elle combine aussi lecture globale (tout le discours) et lecture locale (les unités du discours) pour construire l'interprétation. Elle est basée essentiellement sur des outils de recherche (concordanciers, recherche de contextes, navigation hypertextuelle) et sur des outils statistiques et mathématiques susceptibles de caractériser un texte (dictionnaire de fréquences, calcul du vocabulaire caractéristique, distances intertextuelles, richesse du vocabulaire, accroissement lexical chronologique, etc.).

Cette méthodologie est de plus en plus employée comme outil d'investigation. En effet :

« L'ensemble des Sciences humaines et sociales a centralement affaire aux mots, aux textes, aux discours. Que le texte soit considéré comme archive par l'historien ou objet pour le littéraire, que le discours soit perçu comme témoin ou comme acteur pour le politologue, que le langage soit posé comme commencement pour la psychanalyse ou comme fin pour les Sciences du langage... : nous travaillons tous avec ou sur du matériau linguistique. [...] Partant, la réflexion sur les textes ou les discours –leur composition, les méthodes pour les traiter, leur compréhension ou leur interprétation– n'est pas un luxe mais une nécessité, sauf à s'abandonner à la lecture instinctive, à l'interprétation libre et à une démarche intuitive infra scientifique. Précisément, il s'agira de présenter une méthode d'approche globale des textes susceptible de décrire le lexique et la grammaire, les réseaux thématiques et les structures rhétoriques ; susceptible, surtout, de baliser les parcours de lecture pour objectiver, autant que faire se peut, l'interprétation. » <sup>(2)</sup>

Plusieurs logiciels sont utilisés en logométrie tels Hyperbase, Lexico3, Astartex-diatag, Weblex, Alceste, Cordial, Tropes, Spad-t, DTM, Sphinx Lexica...

<sup>(2)</sup> L'analyse du discours assistée par ordinateur Formation – Damon Mayaffre (CNRS – UMR 6039 « Bases, corpus et Langage » ([http://eprints.aideenligne-francais-universite.auf.org/19/1/pdf\\_Formation\\_Mayaffre\\_Alexandrie\\_dec09\\_.pdf](http://eprints.aideenligne-francais-universite.auf.org/19/1/pdf_Formation_Mayaffre_Alexandrie_dec09_.pdf))

## Le logiciel Hyperbase

## Programme prévisionnel

### Création

*Hyperbase* est un logiciel d'analyse de donnée textuelle (ADT) développé au laboratoire « UMR 7320 : Bases Corpus Langage » à Nice, depuis sa création, par Etienne BRUNET.

### Fonctions

Hyperbase allie deux types de fonctionnalités, documentaires et statistiques, et permet au chercheur la description, la caractérisation, le classement ainsi que l'interprétation des textes.

Les **fonctions documentaires** permettent, par exemple :

- ◆ un retour au texte plein ou lemmatisé pour une lecture naturelle du corpus;
- ◆ une navigation hypertextuelle dans le corpus par mots-clés;
- ◆ une recherche et tri des contextes et des concordances d'une unité;
- ◆ un Index et dictionnaires des formes, des lemmes, des codes et des fréquences.

Quant aux **fonctions statistiques**, elles permettent, entre autres :

- ◆ un calcul des spécificités et graphes de distribution des unités linguistiques du corpus
- ◆ des indices de richesse lexicale et d'accroissement du vocabulaire
- ◆ un traitement et représentation factoriels de matrices lexicales ou grammaticales complexes dans la lignée des travaux de Jean-Paul Benzécri
- ◆ un calcul de distances entre textes, classification et représentation arborées
- ◆ une extraction des phrases typiques et des segments répétés
- ◆ un calcul et une représentation de cooccurrences et de réseaux thématiques
- ◆ une comparaison statistique avec le Trésor de la langue française, GoogleBooks, le British National Corpus.

### Utilisation

Hyperbase est utilisé dans plusieurs disciplines des sciences humaines et sociales afin de décrire et d'interpréter des corpus de textes numériques (œuvres littéraires, archives historiques, discours politiques, questionnaires et entretiens sociologiques). Il permet également d'établir des chronologies endogènes dans des corpus socio-historiques, de faire des typologies d'auteurs dans des corpus littéraires ou de caractériser les genres dans des corpus linguistiques. En outre, il est employé pour déceler les thématiques ou isotopies des discours, notamment grâce au calcul des univers lexicaux d'un mot-pôle ou le traitement généralisé des cooccurrences. Enfin, le logiciel offre au chercheur l'opportunité d'enrichir et de contrôler l'interprétation, en associant le quantitatif (statistique textuelle) au qualitatif (retour assisté et systématique au texte.)

### Sources :

1. « Méthodes qualitatives informatisées ».Page professionnelle de Christophe Lejeune. 06 mar 2017, 18:20 <<http://www.squash.ulg.ac.be/logiciels/>> .
2. « Hyperbase »Wikipédia, l'encyclopédie libre. 2 mar 2017, 12:40 UTC. 05 mar 2017, 18:06 <<http://fr.wikipedia.org/w/index.php?title=Hyperbase&oldid=135025095>>.
3. « HYPERBASE (Traitement documentaire et statistique des corpus textuels) ».Réalisation: Etienne Brunet (Institut National de la langue française).06 mar 2017, 78:20 <<http://ancilla.unice.fr/>>.

### Le mercredi 26 avril 2017

09h00 - 09h30	Séance d'ouverture	Allocutions de Pr. Benabdellah MEFLAH , Directeur du laboratoire « Langue et communication » et Dr. Malika BENSEKAT, chef de projet CNEPRU sur l'analyse lexicométrique du manuel scolaire
09h30 - 09h45	Conférence plénière	Mohamed BOUDAUD « <i>La logométrie appliquée en sciences du langage, didactique des langues et sciences des textes littéraires</i> »
09h45- 10h00	Conférence plénière	Abdelouahid TIOUIDIOUNE « <i>Travaux en logométrie effectués par les étudiants du département de français du C.U. Ahmed Zabana de Relizane</i> »
10h00 - 10h15	Conférence plénière	Dr. Malika BENSEKAT « <i>Travaux en logométrie effectués par les enseignants du département de français du C.U. Ahmed Zabana de Relizane</i> »
10h15 - 10h30	Pause-café	
10h30 - 12h30	Séances en atelier <u>Cours de démonstration</u> Formateur : Mohamed BOUDAUD - Corpus : l'œuvre de Guy de Maupassant	<b>Séance n°01</b> 1.La préparation 2.L'exploration 3.L'exploitation documentaire 4.L'exploitation statistique. Les calculs 5.Le menu DISTRIBUTION
12h30 - 14h00	Déjeuner	
14h00 - 16h00	Reprise du cours de démonstration	<b>Séance n°02</b> 6.Les menu GRAPHIQUE et LISTE 7.Le menu SPÉCIFICITÉS 8.Les menus THÈME et ASSOCIATION 9.Le menu TOPOLOGIE 10.Le menu SEGMENTS RÉPÉTÉS

### Le jeudi 27 avril 2017

09h00 - 10h00	Séances en atelier <u>Cours d'application</u> Formateur : Mohamed BOUDAUD Corpus : l'œuvre de <i>Mouloud Mammeri</i>	<b>Séance n°03</b> 1.La préparation 2.L'exploration 3.L'exploitation documentaire
10h00 - 10h15	Pause-café	
10h15- 12h15	Reprise du cours d'application	<b>Séance n°04</b> 4.L'exploitation statistique. Les calculs 5.Le menu DISTRIBUTION 6.Les menu GRAPHIQUE et LISTE
12h15- 14h00	Déjeuner	
14h00- 16h00	Reprise du cours d'application	<b>Séance n°05</b> 7.Le menu SPÉCIFICITÉS 8.Les menus THÈME et ASSOCIATION 9.Le menu TOPOLOGIE 10.Le menu SEGMENTS RÉPÉTÉS
16h00 - 1630	Séance plénière	- Synthèses des travaux et recommandations - Remise des attestations - Cérémonie de clôture

# INFORMATIONS PRATIQUES

## Public ciblé

Ces deux journées de formation s'adressent à tous les enseignants-chercheurs exerçant dans les universités, les centres universitaires et les centres de recherche au niveau régional (ouest de l'Algérie), préparant un mémoire de magister ou une thèse de doctorat (classique/LMD) de langue française, options : sciences du langage, didactique des langues et sciences des textes littéraires

## Chef du projet CNEPRU

- Dr. Malika BENSEKAT, Directrice du laboratoire DFP de l'Université de Mostaganem

## Directeur du Laboratoire « Langue et communication »

- Pr. Benabdellah MEFLAH

## Coordinateurs des journées

- Mohamed BOUDAUD  
[medboudaoud@yahoo.fr]
- Abdelouahid TIOUIDIOUINE  
[wahid.tioui\_univ@yahoo.fr]

## Modalités de candidature

1. Envoyer en français un CV scientifique (1 page), une lettre de motivation (1 page) et la dernière attestation d'inscription à [algerie.logometrie@yahoo.fr]
2. Les candidats retenus doivent se munir d'un PC portable et d'un flash disque.
3. Les frais de restauration, de déplacement et d'hébergement sont à la charge des participants.
4. Le nombre des places disponibles est limité.

## Comité d'organisation

- **Autres membres du projet CNEPRU (code projet: U0206220140011) :** Nafissa ZEROUALI, Mounia OSMANI
- **Etudiants :** Boulenouar YOUSFI, Kada BOUTRICHA, BELHOCINE Fatima, Abdelmoundir LARBI -DAOUADJI, Saïd YOUSFI et Ahmed MOUSSA SASSI

## Calendrier

• Dates d'envoi des dossiers de candidature à [algerie.logometrie@yahoo.fr]	Du 01 février au 01 mars 2018
• Etablissement de la liste des candidats retenus et de la liste d'attente	Le 02 mars 2018
• Confirmation de participation	Les 08 mars 2018
• Dates des journées d'étude	Le mercredi 14 et le jeudi 15 avril 2018

## Lieu

Salle de réunion de la bibliothèque centrale du Centre universitaire de Relizane

## Configuration conseillée

- Windows 98 et postérieurs
- Pentium II, 8Mo de mémoire vive
- 3 Mo libres sur le disque dur
- Windows NT 3.51 et 4.0.
- Nous conseillons vivement de regrouper programme et corpus dans un même dossier sur le disque dur.

