

En partenariat avec

L'équipe de recherche du projet PRFU intitulé "L'intégration des TICE dans l'enseignement / apprentissage du FLE en Algérie"

Et l'équipe de recherche du projet PRFU intitulé : « Vers une approche enactive dans la conception des dispositifs de formation professionnelle au niveau de l'enseignement supérieur par l'analyse de l'activité »

La maison du doctorant, Université Oran 2 Mohamed Ben Ahmed

Le mercredi 15 mars 2023

ATELIER DOCTORAL DE FORMATION EN LEXICOMETRIE

Analyse informatisée et interprétation des discours
en sciences humaines et sociales
(Initiation au logiciel Lexico 03)

Formateurs

Dr. DOULATE SERDURI Hamida, Dr. TIOUIDIQUINE Abdelouahid,
Dr. YOUSFI Boulanouar



Aujourd'hui qu'une grande partie des textes (écrits ou oraux) qui circulent dans le monde, existe sous forme numérisée et est accessible par Internet, il ne nous est plus permis d'ignorer l'apport considérable que nous offre un traitement assisté par ordinateur. Livres, journaux, magazines et autres productions langagières sont de plus en plus disponibles sous forme électronique. Pourquoi alors continuer de nous comporter à leur égard à la manière traditionnelle, épuisante, lente et incomplète, alors que des logiciels de lexicométrie très performants sont mis à notre disposition et le plus souvent gratuitement?

En effet, ces outils informatiques permettent une exploration rigoureuse de corpus textuels énormes, chose impossible à l'homme. Pour s'en convaincre, songeant par exemple au temps et au travail que nécessiterait une recherche sur les discours de Boumediène, sur l'œuvre complète de Yasmina Khadra, sur un corpus annuel d'un journal comme « El Watan », ou sur un corpus de manuels scolaires, sans l'aide d'un ordinateur.

A l'issue de cette journée de formation, le doctorant sera capable de :

- Comprendre les fondements théoriques et statistiques des analyses lexicométriques
- Utiliser concrètement le logiciel Lexico3

La lexicométrie 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. »⁽¹⁾

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

⁽¹⁾ *L'analyse du discours assistée par ordinateur Formation – Damon Mayaffre (CNRS – UMR 6039 « Bases, corpus et Langage »*

http://eprints.aidenligne-francais-universite.auf.org/19/1/pdf_Formation_Mayaffre_Alexandrie_dec09_.pdf

Exemples d'application de Lexico3

Le logiciel Lexico3 peut être employé, par exemple, pour calculer la taille (ou longueur ou étendue) d'un texte, c'est-à-dire le nombre de mots qui le composent, dresser la liste de ces mots et correspondre à chacun d'eux sa fréquence absolue dans ce texte, c'est-à-dire le nombre de ses apparitions dans celui-ci.

De cette liste pouvant être obtenue, soit par ordre alphabétique, soit par ordre de fréquence décroissante, nous pouvons dégager les vocables fréquemment utilisés, ceux qui le sont rarement et ceux de fréquence moyenne. Ce qui nous donnerait une première idée sur la teneur du vocabulaire employé par le scripteur. Les fréquences peuvent également diriger notre attention sur des phénomènes qui sont invisibles à l'œil nu. Comment pourrait-on, par exemple, deviner dans un corpus de grande taille comme l'œuvre de Boudjedra, que tel vocable a été excessivement utilisé sans l'aide d'un ordinateur ?

Lexico3 permet aussi de quantifier la ponctuation utilisée dans notre texte ou notre corpus. Ce sont des informations importantes pour une analyse textuelle. En effet, de particulier, il nous serait possible de déterminer et d'accéder rapidement à toutes les phrases composant le texte étudié. D'attribuer à ces dernières une longueur moyenne, ce qui est important dans une étude stylistique.

On pourrait également extraire tous les contextes d'un mot qui nous intéresse, en particulier, afin de déterminer les termes qui sont « trop » employés dans son voisinage, et ceux qui le sont « insuffisamment ». Ce qui nous révélerait entre autres le sens que donne le scripteur à ce mot.

Ce logiciel sert, par ailleurs, à comparer deux ou plusieurs textes entre eux ou avec un corpus numérisé comme celui de la base *Frantext*.^[2]

Il nous informe, par exemple, si un texte T1 est plus riche en vocabulaire qu'un texte T2. C'est-à-dire quel est celui qui contient plus de termes différents que l'autre. En d'autres termes, s'il y a répétition ou renouvellement des vocables utilisés.

Il nous dit comment évolue l'accroissement du vocabulaire en chacun d'eux et détermine les moments de variation significatifs dans l'apport de ce vocabulaire. En particulier, la statistique lexicale, par un simple découpage du texte en tranches, nous permet de déceler les parties qui ont connu un afflux de vocables nouveaux et nous indique la position où cette augmentation a eu lieu. Ce qui permet de savoir à quels moments de son discours le scripteur renouvelle son vocabulaire ou se répète. Le retour au texte nous révélerait alors les thèmes et les raisons de cette attitude.

Lexico3 nous donne, par ailleurs, des outils mathématiques pour mesurer la distance lexicale entre deux ou plusieurs textes, c'est-à-dire en gros le nombre de vocables qu'ils n'ont pas employé en commun.

Ce sont là quelques applications du Lexico3 que nous aborderons lors des ateliers.

Nous ajouterons que les chiffres et les graphes obtenus par celui-ci, s'ils sont riches en informations, ne sont pas suffisants. En effet, l'étape capitale en statistique est l'interprétation des données chiffrées, obtenues lors de l'expérience réalisée sur un échantillon ou la population entière des unités concernées par l'analyse.

^[2] *Frantext est le nom d'une base de données de textes français : textes littéraires et philosophiques, mais aussi scientifiques et techniques (environ 10%), développée et maintenue au sein de l'ATILF-CNRS. Elle a été créée autour d'un noyau de mille textes, dans les années 70, afin de fournir des exemples pour le Trésor de la Langue Française. Une fois le dictionnaire terminé, elle a continué à évoluer. Elle est régulièrement enrichie et comporte aujourd'hui près de 4000 références. Elle est la seule à proposer des recherches sur des textes qui vont de 1180 à 2009*
^[3] Lornalle, C., Martinez, W., Fleury, S., Salem, A., Kuncova, A., & Maisondieu, A. (2001). *Dix premiers pas avec Lexico3. Manuel d'utilisation abrégé*, à télécharger à partir du site : <http://www.tal.univ-paris3.fr/lexico/lex3-10pas/index.htm>

Programme prévisionnel		
Le mercredi 15 mars 2023		
08h30-09h00	Accueil des participants, distribution des badges, affichage de la liste officielle des participants	
09h00 - 09h15	Cérémonie d'ouverture par Pr. BABA HAMED Nassim, Vice-Recteur de la formation supérieure de troisième cycle, l'habilitation universitaire, la recherche scientifique et la formation supérieure de post-graduation, de l'Université d'Oran 2	
09h15 - 09h30	Dr. DOULATE SEROURI Hamida Communication : « <i>Le traitement lexicométrique des corpus oraux: outils et étapes</i> »	
09h30-09h45	Dr. TIOUIDIOUINE Abdelouahid et Pr AISSA Khaldia Communication : « <i>Travaux en lexicométrie effectués par les étudiants et les enseignants du département de français de l'Université de Relizane</i> »	
09h45-10h00	Pause-café	
10h00 - 12h00	Cours de démonstration et d'application - Corpus intitulé : « Le débat en classe de FLE » et constitué par Dr. DOULATE SEROURI Hamida - Logiciel Lexico 3.6	Séance n°01 1.La préparation 2.L'exploration 3.L'exploitation documentaire 4.L'exploitation statistique. Les calculs 5.Le menu DISTRIBUTION
12h00 - 12h15	Pause-déjeuner	
12h15- 14h30	Suite du cours de démonstration et d'application - Corpus intitulé : « Le débat en classe de FLE » et constitué par Dr. DOULATE SEROURI Hamida -Logiciel Lexico 3.6	Séance n°02 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
14h30-15h00	Cérémonie de clôture	- Synthèses des travaux et recommandations -Photo collective

INFORMATIONS PRATIQUES

Public ciblé

- 25 doctorants
- De préférence un tiers des participants provient des établissements universitaires limitrophes de la wilaya d'Oran
- Tous les doctorants inscrits pour une thèse de doctorat (classique/LMD) de langue française, options : sciences du langage, didactique des langues et sciences des textes littéraires en particulier et en sciences humaines et sociales en général.

Modalités de candidature

1. Inscription via le lien :
<https://forms.gle/Fqt6d3NNgQzXfb9s6>
2. Les candidats retenus doivent se munir d'un PC portable et d'un flash disque.
3. La formation est gratuite.

Calendrier

Date limite d'envoi des dossiers de candidature via le lien :

<https://forms.gle/Fqt6d3NNgQzXfb9s6>

09 mars 2023

Établissement de la liste des candidats retenus

Le 11 mars 2023

Confirmation de participation

Le 12 mars 2023

Date de la journée d'étude

Le mercredi 15 mars 2023

Téléchargement et configuration conseillée

- Télécharger la version 3.6 de lexico à partir du lien :
<http://lexi-co.com/Produits.html>
- 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.

Lieu

BUC (Bibliothèque universitaire centrale, Maison du doctorant, Université d'Oran 2 Mohamed Ben Ahmed



Fiche élaborée par : Dr. Abdelouahid TIOUJOUINE
Université de Relizane