

Introduction

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 logométrie 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 Bouteflika, sur l'œuvre complète de Boudjedra, sur un corpus annuel d'un journal comme « El Watan », ou sur un corpus de manuels scolaires, sans l'aide d'un ordinateur.

Objectifs des deux journées

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 Lexico3

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. » ⁽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://lprints.aixmarseille.fr/francis-universite.auf.org/19/11/pdf_Formation_Mayaffre_Alexandrie_de09_.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 celles-ci, 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 abordons 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.

Programme prévisionnel

09h00 - 09h30	Séance d'ouverture	Allocutions de M.M. Les Directeurs du CUR (Dr. RAHIEL), de l'I.L.L. (Dr. BRAHIMI) et du Laboratoire « Langue et communication » (Dr. MEFLAH)
09h30 - 09h45	Conférence plénière	Mohamed BOUDAUD « Usages de la logométrie en sciences du langage, didactique des langues et sciences des textes littéraires »
09h45-10h00	Conférence plénière	Abdelouahid TIOUIDIOUNE « Synthèse des travaux en logométrie effectués par les enseignants et les étudiants du département de français du C.U. Ahmed Zabana de Relizane »
10h00 - 10h15	Conférence plénière	Madjid MASMUDI « Initiation à l'analyse textométrique »
10h15 - 10h30	Pause-café	
10h30 - 12h30	Séances en atelier Cours de démonstration ⁽³⁾ Formateur : Mohamed BOUDAUD Corpus : « Topographie idéale pour une agression caractérisée », roman de Rachid Boudjedra	Séance n°01 1. Intégrer un texte dans la base Réaliser une concordance 2. Créer une partition du corpus 3. Constituer un groupe de formes 4. Créer une carte des sections
12h30 - 14h00	Déjeuner	
14h00 - 16h00	Reprise du cours de démonstration	Séance n°02 6. Créer un inventaire des segments répétés 7. Calculer les spécificités d'une partie 8. Réaliser une Analyse Factorielle des Correspondances 9. Naviguer parmi les résultats 10. Enregistrer des résultats dans le rapport
08h00 - 10h00	Séances en atelier Cours d'application Formateur : Mohamed BOUDAUD Corpus : « Le manuel scolaire de français de la 3 ^{ème} AS »	Séance n°03 1. Intégrer un texte dans la base 2. Réaliser une concordance 3. Créer une partition du corpus 4. Constituer un groupe de formes
10h00 - 10h15	Pause-café	
10h15-12h15	Reprise du cours d'application	Séance n°04 5. Créer une carte des sections 6. Créer un inventaire des segments répétés 7. Calculer les spécificités d'une partie
12h15-14h00	Déjeuner	
14h00-16h00	Reprise du cours d'application	Séance n°05 8. Réaliser une Analyse Factorielle des Correspondances 9. Naviguer parmi les résultats 10. Enregistrer des résultats dans le rapport
16h00 - 1630	Séance plénière	- Synthèses des travaux et recommandations - Remise des attestations - Cérémonie de clôture

⁽³⁾ Lamalle, C., Martinez, W., Flury, S., Sakon, A., Kancova, A., & Maisondieu, A. (2001). *Dis premiers pas avec Lexico3. Manuel d'utilisation abrégé*, à télécharger à partir du site : <http://www.tal.univ-paris3.fr/lexico/lex3-10p.as/index.htm>

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

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 pour les intervenants et les participants sont assurés par le Laboratoire « Langue et communication » et l'Institut des Lettres et des langues durant les deux journées.
4. Les frais de déplacement et d'hébergement sont à la charge des participants.

Calendrier

• Dates d'envoi des dossiers de candidature à algerie.logometrie@yahoo.fr	Du 14 avril au 14 mai 2015
• Etablissement de la liste des candidats retenus et de la liste d'attente	Le 15 mai 2015
• Confirmation de participation	Le 16 mai 2015
• Dates des journées d'étude	Le mercredi 20 et le jeudi 21 mai 2015

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.

Responsables scientifiques

- Dr. Benabdellah MEFLAH, Directeur du Laboratoire « Langue et communication »
- Dr Boudaoud BRAHIM, Directeur de P.L.L.

Coordinateurs des journées

- Mohamed Boudaoud
[\[medboudaoud@yahoo.fr\]](mailto:medboudaoud@yahoo.fr)
- Abdelouahid Tiouidiouine
[\[wahid.tiou_univ@yahoo.fr\]](mailto:wahid.tiou_univ@yahoo.fr)
- Madjid MASMOUDI
[\[elmadjido@yahoo.fr\]](mailto:elmadjido@yahoo.fr)

Comité d'organisation

- **Membres du laboratoire LANGUE & COMMUNICATION :** Mohamed Saoudi SOUIG, Nabil Zoheir BERREHAL, Kahina AIT YAHIA, Khansa SMAIL-MESSOUS
- **Membres du projet CNEPRU (code projet: U00206220140011) :** Dr Malika BENSEKAT, Nafissa ZEROUALI, Mounia OSMANI
- **Etudiants (mastérants):** Aicha ALIANE, Aoued CHEMLAL, Sofiane BENGUERFI, Boulenouar YOUSFI, Nour el Houda BELKHODJA, Sabrina BENHENNI, khadidja EDDRIEF

Lieux

Salle de conférences du centre universitaire et salles de TD du département de français

