

## LEXICOMETRIE<sup>(\*)</sup>

Philippe GALIANA

### I) QUE FAIT-ON AVEC LEXICOMETRIE ?

La lexicologie est l'étude scientifique du vocabulaire d'un texte. Lorsque cette étude scientifique d'un texte est faite avec l'outil informatique, on parle alors de lexicométrie.

Un logiciel de lexicométrie, comme celui du même nom de chez NATHAN, est un programme informatique plurilinguistique et pluridisciplinaire qui permet bien sûr de saisir un texte (le module de traitement de texte de LEXICOMETRIE étant assez peu puissant, nous lui préférons l'intégré WORKS dont nous utilisons également la base de données.), mais surtout de faire le décompte des formes (ou mots) employés dans ce texte. Pour chaque forme, le logiciel ne fait que compter le nombre de fois qu'elle apparaît. Le même travail pourrait se faire manuellement, mais il serait fastidieux voire impossible pour de longs textes ou de gros livres. Une fois les comptes terminés, l'utilisateur aura une somme d'informations avec laquelle il pourra au mieux étayer des hypothèses et diriger ses recherches.

(Lorsqu'il s'agit de traiter de gros ouvrages, la saisie au clavier est hors de question et, dans ce cas, nous faisons appel au scanner.)

Une fois que le logiciel LEXICOMETRIE a terminé son travail de comptage de chacune des formes du texte nous obtenons des index et des concordanciers. (voir annexes)

Un index n'est ni plus ni moins que la liste des formes du texte accompagnées d'un nombre représentant la fréquence d'apparition dans le texte.

Alors que l'index est une liste hors contexte, le concordancier, donne lui, les différents contextes dans lesquels on a utilisé une forme.

---

(\*) LEXICOMETRIE distribué par Cedic/Nathan et le C.A.R.F.I de Versailles. EGA/CGA Hercule

En ce qui concerne par exemple une forme comme PIGEON, il est primordial de pouvoir la situer dans son contexte afin de la désambiguïser.

Les index sont généralement de deux types, les index alphabétiques et les index hiérarchiques.(voir annexes)

## **ID VERS LES PREMIERES INTERPRETATIONS LINGUISTIQUES**

Les index et les concordanciers peuvent à leur tour donner lieu à des calculs statistiques et à des représentations graphiques. En effet, toujours en prenant comme exemple la forme PIGEON, le concordancier de cette forme, donné par le logiciel LEXICOMETRIE nous permet de déterminer dans quel cas du texte il s'agit du substantif et dans quel autre cas du texte il s'agit de l'adjectif. C'est la désambiguïstation. Tous les contextes donnant lieu à une utilisation en tant que substantif sont alors répertoriés dans une base de données (hors de LEXICOMETRIE) et tous les contextes donnant lieu à une utilisation en tant qu'adjectif sont répertoriés dans un autre fichier de la base de données. De là, peuvent être tirées des statistiques pour chaque emploi de la forme étudiée.

Les index et les concordanciers peuvent être globaux (portant sur tout le texte) ou sélectifs (portant sur un seul personnage ou une seule forme). Ils peuvent nous faire apprécier l'importance d'un personnage par rapport aux autres personnages ou par rapport à la totalité de l'oeuvre. Un personnage qui a lui seul représenterait 25% des formes de l'oeuvre est, sans nul doute un personnage qu'il vaudrait mieux étudier de près. Les index permettent de confronter la richesse du vocabulaire du personnage A avec la richesse du vocabulaire du personnage B. De même, en examinant de près le vocabulaire d'un personnage, on peut tirer des renseignements intéressants sur son niveau socio-culturel...

Avec les index, nous pouvons aussi nous intéresser à l'étude des temps et des personnes verbales.

Bref. Il serait trop long ici de détailler tous les axes de recherche vers lesquels peuvent nous mener le "saucissonnage" d'un texte par un logiciel de mesures lexicales comme LEXICOMETRIE.

A Paris VIII, dans le laboratoire d'études linguistiques hispaniques, les étudiants de 2ème et 3ème cycles utilisent donc les résultats du logiciel LEXICOMETRIE comme un point de départ dans

Philippe GALIANA LE BULLETIN DE L'EPI

l'étude littéraire d'un ouvrage; les index et les concordanciers fournissent les axes de recherche nécessaires à une lecture linguistique et scientifique.

Sans aller aussi loin dans la recherche linguistique, LEXICOMETRIE, peut également convenir à l'instituteur de cours préparatoire qui, soucieux d'une bonne progression dans les textes qu'il donne à ses élèves, examine la difficulté du vocabulaire, la richesse du vocabulaire dans les index et les concordances que lui fournira LEXICOMETRIE.

D'ailleurs, la fameuse A.F.L (Association Française pour la Lecture), a compris qu'il fallait absolument élargir le cercle des initiés à ce genre de lecture scientifique, et commercialise dès le mois de Février 91 le logiciel ELMO INTERNATIONAL (français, allemand, anglais, italien, espagnol) qui devrait connaître un bon succès dans les collèges et les écoles primaires.

Un peu plus connu que LEXICOMETRIE et ELMO INTERNATIONAL, le logiciel PISTES diffusé par le CNDP, donne également des index et des concordanciers.

Aux pages suivantes, vous trouverez des exemples de concordanciers et d'index de cet article par LEXICOMETRIE ainsi que le schéma de nos propositions méthodologiques actuelles à PARIS VIII.

GALIANA Philippe  
 Instituteur chargé de cours  
 à PARIS VIII.

## INDEX 1 : 848 MOTS / 357 FORMES

a	9	CGA	1	donnant	2
à	12	chacune	1	donné	1
absolument	1	chaque	2	donne	3
accompagnées	1	CHARGE	1	données	3
actuelles	1	chez	1	donner	1
adjectif	2	clavier	1	dont	1
afin	1	CNDP	1	doute	1
agit	3	collèges	1	du	14
ailleurs	1	comme	5	écoles	1
ainsi	1	commercialise	1	effet	1
allemand	1	compris	1	EGA	1
aller	1	comptage	1	également	3
alors	3	compter	1	élargir	1
alphabétiques	1	comptes	1	élèves	1
anglais	1	concerne	1	elle	1
apparaît	1	concordances	1	ELMO	2
apparition	1	concordancier	2	ème	2
appel	1	concordanciers	6	emploi	1
apprécier	1	confronter	1	employés	1
article	1	connaître	1	en	6
assez	1	connu	1	espagnol	1
association	1	contexte	2	est	9
au	3	contextes	3	et	21
aura	1	convenir	1	étant	1
aussi	2	cours	1	étayer	1
autre	2	COURS	1	être	2
autres	1	cycles	1	étude	4
aux	2	d'	13	études	1
avec	4	dans	15	étudiants	1
AVEC	1	de	48	étudiée	1
axes	2	DE	1	étudier	1
b	1	décompte	1	examinant	1
base	3	départ	1	examine	1
bien	1	des	17	exemple	2
bon	1	dès	1	exemples	1
bonne	1	désambiguïsation	1	f	2
bref	1	désambiguïser	1	faire	3
c	1	détailler	1	faisons	1
c'	1	déterminer	1	fait	1
calculs	1	deux	1	FAIT-ON	1
cas	3	devrait	1	faite	1
ce	4	différents	1	fallait	1
cedic	1	difficulté	1	fameuse	1
celui	1	diffusé	1	.....	
cercle	1	diriger	1		
cet	1	distribué	1		
cette	2	donc	1		

## INDEX 2 : 848 Mots / 357 Formes

de	48	contextes	3	nombre	2
les	24	donne	3	oeuvre	2
et	21	données	3	page	2
le	21	également	3	PARIS	2
la	19	faire	3	permet	2
un	19	fois	3	peu	2
des	17	hors	3	peut	2
dans	15	lecture	3	PIGEON	2
du	14	lieu	3	plus	2
index	14	même	3	portant	2
d'	13	on	3	près	2
LEXICOMETRIE	13	recherche	3	quel	2
à	12	richesse	3	rapport	2
l'	12	s'	3	répertoriés	2
texte	12	son	3	sans	2
une	12	sont	3	serait	2
il	10	sur	3	ses	2
a	9	tous	3	seul	2
est	9	VIII	3	statistiques	2
nous	9	adjectif	2	substantif	2
logiciel	8	aussi	2	tant	2
par	8	autre	2	textes	2
que	8	aux	2	travail	2
forme	7	axes	2	utilisation	2
personnage	7	cette	2	voir	2
concordanciers	6	chaque	2	absolument	1
en	6	concordancier	2	accompagnées	1
vocabulaire	6	contexte	2	actuelles	1
comme	5	donnant	2	afin	1
ou	5	ELMO	2	ailleurs	1
peuvent	5	ème	2	ainsi	1
qu'	5	être	2	allemand	1
qui	5	exemple	2	aller	1
avec	4	f	2	alphabétiques	1
ce	4	gros	2	anglais	1
étude	4	i	2	apparaît	1
formes	4	informatique	2	apparition	1
lui	4	INTERNATIONAL	2	appel	1
pour	4	lesquels	2	apprécier	1
scientifique	4	lexicométrie	2	article	1
agit	3	linguistique	2	assez	1
alors	3	liste	2	association	1
au	3	mais	2	aura	1
base	3	mieux	2	autres	1
cas	3	ni	2	.....	

## Recherche de : la , dans : ARTICLE

8[ 1] LA lexicologie est l'étude scientifique du voc  
 9[ 49]é WORKS dont nous utilisons également LA base de données.), mais surtout de faire le  
 10[ 8]l s'agit de traiter de gros ouvrages, LA saisie au clavier est hors de question et,  
 12[ 9] Un index n'est ni plus ni moins que LA liste des formes du texte accompagnées d'un  
 12[ 19]accompagnées d'un nombre représentant LA fréquence d'apparition dans le texte.  
 13[ 38]PIGEON, il est primordial de pourvoir LA situer dans son contexte afin de la désambi  
 13[ 45]r la situer dans son contexte afin de LA désambiguïser.  
 21[ 28]et, toujours en prenant comme exemple LA forme PIGEON, le concordancier de cette for  
 21[ 67] texte il s'agit de l'adjectif. C'est LA désambiguïstation. Tous les contextes donn  
 21[ 111] répertoriés dans un autre fichier de LA base de données. De là, peuvent être tirées  
 21[ 116]tre fichier de la base de données. De LA, peuvent être tirées des statistiques pour  
 21[ 126]es statistiques pour chaque emploi de LA forme étudiée.  
 23[ 42]x autres personnages ou par rapport à LA totalité de l'oeuvre. Un personnage qui a 1  
 23[ 76]s. Les index permettent de confronter LA richesse du vocabulaire du personnage A avec  
 23[ 84]e du vocabulaire du personnage A avec LA richesse du vocabulaire du personnage B. De  
 30[ 6] Sans aller aussi loin dans LA recherche linguistique, LEXICOMETRIE, peut é  
 30[ 32]tes qu'il donne à ses élèves, examine LA difficulté du vocabulaire, la richesse du vo  
 30[ 36]examine la difficulté du vocabulaire, LA richesse du vocabulaire dans les index et le  
 31[ 2] D'ailleurs, LA fameuse A.F.L (Association Française pour la  
 31[ 8]use A.F.L (Association Française pour LA Lecture), a compris qu'il fallait absolument

20 occurrence(s) trouvée(s)

## Recherche de : dans , dans : ARTICLE

9[ 64]écompte des formes (ou mots) employés DANS ce texte. Pour chaque forme, le logiciel r  
 10[ 17]e au clavier est hors de question et, DANS ce cas, nous faisons appel au scanner.)  
 12[ 23] présentant la fréquence d'apparition DANS le texte.  
 13[ 16], donne lui, les différents contextes DANS lesquels on a utilisé une forme. En ce qui  
 13[ 40] est primordial de pourvoir la situer DANS son contexte afin de la désambiguïser.  
 21[ 45]LEXICOMETRIE nous permet de déterminer DANS quel cas du texte il s'agit du substantif  
 21[ 56] du texte il s'agit du substantif et DANS quel autre cas du texte il s'agit de l'adj  
 21[ 84]que substantif sont alors répertoriés DANS une base de données (hors de LEXICOMETRIE)  
 21[ 106]en tant qu'adjectif sont répertoriés DANS un autre fichier de la base de données. De  
 28[ 4] A Paris VIII, DANS le laboratoire d'études linguistiques hisp  
 28[ 29]LEXICOMETRIE comme un point de départ DANS l'étude littéraire d'un ouvrage; les index  
 30[ 5] Sans aller aussi loin DANS la recherche linguistique, LEXICOMETRIE, L  
 30[ 23]qui, soucieux d'une bonne progression DANS les textes qu'il donne à ses élèves, exami  
 30[ 40]cabulaire, la richesse du vocabulaire DANS les index et les concordances que lui four  
 31[ 49] qui devrait connaître un bon succès DANS les collègues et les écoles primaires.

15 occurrence(s) trouvée(s)

