

ANALYSE LEXICOMETRIQUE D'UN CORPUS LEXICOGRAPHIQUE EN LANGUE FANG : COMMENT CONTOURNER LES DIFFICULTES LIEES AU PROBLEME D'ECRITURE ?

Thierry Afane-Otsaga

Université Omar Bongo

afanotsaga@gmail.com

Résumé

La lexicométrie est l'étude quantitative du lexique, à l'aide de méthodes statistiques. En lexicographie, cette technique joue un rôle important dans la sélection de la macrostructure et la construction de la microstructure des dictionnaires. Le but de cet article est de montrer l'importance des techniques lexicométriques dans le processus de compilation d'un dictionnaire. Cette démonstration s'appuie sur l'usage d'un programme de lexicométrie dénommé TenkaText. L'article aborde aussi la difficulté des programmes lexicométriques à analyser des corpus en langues gabonaises et formule quelques propositions pour résoudre ce problème.

Mots clés : *Lexicométrie, TenkaText, dictionnaire, corpus.*

Abstract:

Lexicometry is the quantitative study of lexicon with the help of statistical methods. In lexicography, that technics play an important role in the selection of the macrostructure and the construction of the microstructure of dictionaries. The purpose of this article is to show the importance of lexicometry technical in the dictionary compilation process. This demonstration will use of a lexicometric program named TenkaText. The article also talk about the difficulty of lexicometric programs to analyse corpora in Gabonese languages and formulate some proposals in order to solve that matter.

Keywords: *Lexicometry, TenkaText, dictionary, corpora.*

Introduction

La lexicométrie (aussi appelée la logométrie) est l'étude quantitative du lexique, à l'aide de méthodes statistiques. En d'autres termes, la lexicométrie est une branche de la linguistique qui étudie statistiquement l'emploi des mots, c'est-à-dire combien de fois un mot particulier apparait dans un corpus et dans quel contexte. En lexicographie, cette technique permet de mesurer la quantité d'usage

des mots d'une langue par un locuteur ou un groupe de locuteurs tout en identifiant le contexte d'utilisation des termes dont on veut comprendre la signification. La lexicométrie est donc une démarche qui fait appel aussi bien à l'informatique, à la statistique ainsi qu'aux sciences humaines et sociales. En effet, les différentes analyses proposées par cette méthode se basent sur le calcul des mots du corpus. En lexicographie, la lexicométrie joue un rôle primordial dans la sélection des éléments de la macrostructure et la construction de la microstructure qui sont les deux principales composantes du dictionnaire. Le but de cet article est de montrer non seulement l'importance des techniques lexicométriques dans le processus de compilation d'un dictionnaire en langues gabonaises, mais aussi de proposer une piste de solution pour contourner les difficultés rencontrées lors de cette analyse en rapport avec l'écriture. Cette démarche va s'appuyer sur des exemples concrets à travers l'usage d'un programme de lexicométrie appelé TenkaText en comparant les résultats d'analyse de deux langues : le français et le fang. En effet, nous prendrons le soin de mettre en lumière la difficulté des programmes lexicométriques à analyser des corpus en langues africaines en général, plus précisément en langues gabonaises, du fait des systèmes d'écriture utilisés pour ces dernières. Ce constat nous amènera à formuler quelques propositions pour résorber cette difficulté.

1. Les objectifs de la lexicométrie

La lexicométrie (aussi appelée textométrie), « est l'étude quantitative du vocabulaire et permet de caractériser certains usages dans le discours » (Catherine Fuchs, 2022). « C'est une technique automatique de statistique linguistique, de statistique lexicale ou de linguistique quantitative, de statistique textuelle, voire d'analyse des données en linguistique. Elle peut même être une méthodologie d'étude du discours, qui se veut exhaustive, systématique est automatisée » (Muller Charles, 1992). Cette pratique est un inventaire de mots ou un répertoire linguistique, une alliance de l'informatique, de la statistique et des sciences humaines et sociales, vu que les différentes analyses proposées par cette étude se basent sur le calcul entier de tous les mots du corpus. En d'autres termes, la lexicométrie est la science de mesure

du lexique. Cette technique se réalise au moyen de programmes informatiques appropriés choisis par le chercheur selon ses objectifs. La lexicométrie est donc un outil d'analyse statistique et graphique à partir duquel il est possible de définir la quantité des mots, de calculer la fréquence des unités dans un corpus, de calculer l'écart lexical entre les textes du corpus, de mettre en lumière la richesse du vocabulaire d'une langue et l'état de son évolution. Elle permet même d'identifier la typologie d'un discours, les thèmes transversaux abordés dans un corpus, la cohésion et la cohérence d'un discours, etc.

2. Quelques programmes de lexicométrie en lexicographie

En lexicographie, les techniques ou méthodes lexicométriques sont surtout utilisées pour faciliter le travail du lexicographe dans le processus de sélection de la macrostructure et la rédaction de la microstructure lors de la compilation d'un dictionnaire. Lorsqu'on parle de macrostructure en lexicographie, on fait référence à l'ensemble de mots de la langue retenus comme entrées principales du dictionnaire. En d'autres termes, la macrostructure est constituée par tous les mots qui font faire l'objet d'une définition ou d'une traduction dans un dictionnaire. Quant à la microstructure, elle fait référence aux définitions ou aux traductions en elles-mêmes. Il existe plusieurs programmes ou logiciels de lexicométrie disponibles et gratuits sur la toile. Nous citerons particulièrement Hyperbase, Lexico et TenkaText.

Hyperbase est un logiciel universitaire d'exploration documentaire et statistique des textes. Il est conçu et diffusé par le CNRS et l'Université Nice Sophia Antipolis. Hyperbase est un programme de statistique textuelle et d'exploration documentaire des grands corpus en France. Le logiciel est principalement utilisé pour les thèses et les ouvrages de linguistique, de littérature, d'histoire, de sociologie ou de science politique. Hyperbase associe des fonctions documentaires et statistiques qui permettent à l'analyste de décrire, de caractériser, de classer et d'interpréter les textes (Bonin Emmanuel et Dallo Alain, 2003 : 389-402).

Lexico est un outil de statistiques textuelles ayant plusieurs fonctionnalités. C'est un logiciel ayant plusieurs versions dont l'originalité principale est qu'elle permet à l'utilisateur de garder la

maîtrise sur l'ensemble des processus lexicométriques depuis la segmentation initiale jusqu'à l'édition des résultats finaux. Les unités qui seront ensuite automatiquement décomptées sont exclusivement constituées à partir de la liste des délimiteurs fournie par l'utilisateur, sans recours à des ressources dictionnaires extérieures. Ce programme permet : (1) d'obtenir la liste des formes présentes dans un document (nombre de formes, hapax, fréquence des mots, etc.) ; (2) d'établir le dictionnaire de ces formes (classement lexicographique, ordre alphabétique, fréquence, etc.) ; (3) d'afficher la concordance d'un mot précis (possibilité de trier le contexte à gauche ou à droite par ordre alphabétique, ce qui permet de trouver des expressions ou autres collocations) ; (4) de rechercher des segments répétés dans le texte, d'établir des partitions du texte (en fonction des auteurs, de la chronologie, de la source, du genre, de la pagination originale, etc., ce qui permet d'effectuer par la suite des comparaisons entre parties) ; (5) de déterminer des sections à travers lesquelles il est possible de naviguer ; (6) d'effectuer différents calculs statistiques (Bonin Emmanuel et Dallo Alain, 2003 : 389-402). Mais dans cet article, nous avons choisi d'illustrer notre propos avec *TenkaText*, un programme simple, facile à manipuler et gratuit que nous avons souvent utilisés pour nos statistiques lexicométriques lors de l'analyse de corpus lexicographiques. Pour qu'un logiciel de lexicométrie fournisse des résultats satisfaisants, il est indispensable que les corpus analysés soient constitués de textes d'origine et stabilisés. En d'autres termes, le corpus ne doit subir aucune modification ou transformation par rapport aux données collectées à la source : discours, entretiens, conversations, interviews, etc.

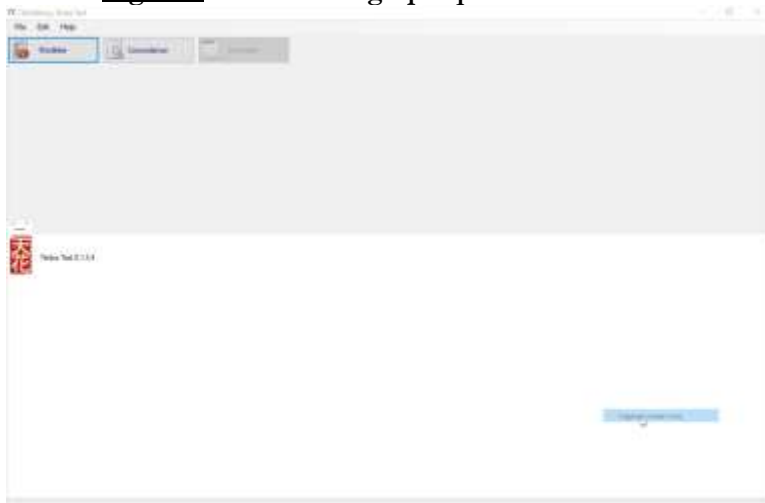
3. TenkaText

3.1. Présentation

Aussi connu sous la dénomination de Corsis, TenkaText est un logiciel libre développé par la compagnie corsis.sf.net (https://fr.osdn.net/projects/sfnet_corsis). Ce programme, disponible en anglais, est compatible avec les systèmes d'exploitation Windows et Linux. TenkaText est un logiciel d'analyse de corpus qui permet au chercheur d'obtenir des informations telles que :

- *La liste des mots*. Elle présente l'ensemble des mots du corpus par ordre de fréquence tout en fournissant des informations sur la représentativité de chaque item dans l'ensemble du corpus. En d'autres termes, elle montre le nombre de fois qu'un item lexical apparait dans le corpus, du plus récurrent au moins fréquent.
- *La liste des concordances*. Elle met en relief les différents contextes d'apparition de chaque item dans le corpus à la demande de l'utilisateur. En d'autres termes, elle met en relief l'environnement contextuel des items lexicaux dans un corpus. Pour cela, il suffit au chercheur de choisir un mot et de demander au programme de lui fournir ses contextes d'utilisation dans le corpus.
- *La liste des mots clés*. Elle présente les mots qui sont le plus en rapport avec un thème particulier.

Figure 1 : L'interface graphique de TenkaText



Source : Afane-Otsaga (2021), Cours d'analyse de données Master 2 Sciences du Langage, Université Omar Bongo.

3.2. TenkaText et macrostructure

« La macrostructure d'un dictionnaire est constituée par l'ensemble des unités lexicales retenues comme lemmes du dictionnaire. Si un dictionnaire est composé d'une seule liste de mots, cela veut dire que c'est cette liste de mots qui constitue la macrostructure du dictionnaire. Par contre, si un dictionnaire est composé d'une liste centrale et d'une liste de mots additionnelle (en annexe par exemple), ces deux listes constituent la macrostructure du dictionnaire » (Hausmann Franz Josef et Wiegand Herbert Ernst, 1989:328). En d'autres termes, la macrostructure du dictionnaire est constituée par l'ensemble des mots décrits dans un dictionnaire. Ces mots peuvent être décrits dans la partie centrale du dictionnaire (nomenclature) ou dans les parties annexes du dictionnaire (pré-textes et post-textes). Rappelons que les lemmes sont des items lexicaux qui font l'objet d'une description ou d'une traduction dans un dictionnaire. En lexicographie, la sélection des signes-lemmes obéit à un certain nombre de critères dont le plus important est la fréquence d'utilisation des items lexicaux dans la langue. En effet, plus une unité lexicale est utilisée dans une langue et plus il est nécessaire que les locuteurs en comprennent le sens et sachent l'utiliser dans la pratique quotidienne de la langue. De même, la fréquence d'utilisation d'un item lexical dans une langue peut susciter la création de contenus sémantiques divers et même sa transformation morphologique. Le lexicographe doit donc tenir compte de tous ces éléments lors de la sélection des items lexicaux retenus comme lemmes du dictionnaire. Ainsi, avec l'aide de TenkaText, le chercheur peut facilement obtenir une liste des items lexicaux les plus utilisés dans le corpus. Il peut aussi analyser des textes d'origines divers et les associer à un même corpus tout en gardant les informations sur la source de chaque texte.

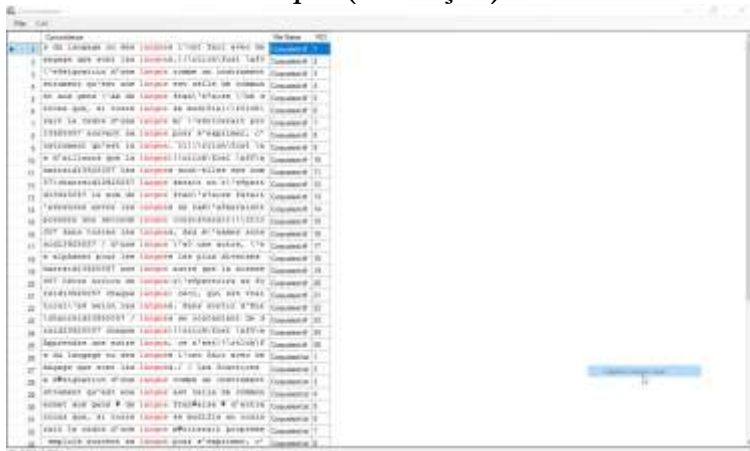
dans une langue, la représentativité (pourcentage) et l'échantillonnage (quantité des textes que contient le corpus) sont des critères importants pour la sélection des items lexicaux qui vont constituer la macrostructure du dictionnaire.

3.2. TenkaText et microstructure

En lexicographie, lorsqu'on parle de microstructure, on fait référence à l'ensemble des commentaires servant à définir ou à traduire un signe-lemme dans un dictionnaire. Pour les dictionnaires monolingues, la microstructure concerne essentiellement l'orthographe, les définitions et les exemples. Pour les dictionnaires bilingues ou plurilingues, la microstructure renvoie généralement aux traductions équivalentes, à la transcription phonétique de l'équivalent et aux exemples.

Pour réaliser de bonnes définitions ou proposer des traductions équivalentes conformes, il est nécessaire que le lexicographe ait à sa disposition des données capables de lui fournir les différents contextes d'utilisation des lemmes à définir ou à traduire. C'est dans ce contexte que l'utilisation d'un programme de lexicométrie est intéressante à plus d'un titre. En effet, à l'instar de TenkaText, un logiciel de lexicométrie fournit au chercheur les différentes collocations des items lexicaux du corpus.

Figure 3 : Aperçu des concordances du mot « langue » dans un corpus (en français).



Source : Afane-Otsaga (2021), Cours d'analyse de données Master 2 Sciences du Langage, Université Omar Bongo.

Dans la figure 3, on peut voir le résultat de l'analyse de TenkaText pour le mot « langue » en rouge. Le programme présente le mot en question avec tous ses environnements contextuels dans le corpus. On peut ainsi voir les phrases dans lesquelles l'item lexical est utilisé, ainsi que les mots qui l'entourent généralement. Grâce à ces informations, le lexicographe dispose de données nécessaires pour réaliser la construction de sa microstructure.

4. Lexicométrie en langues gabonaises

4.1. La difficulté liée à l'écriture

La quasi-totalité des programmes lexicométriques réalisent une analyse des corpus en se basant sur les unités lexicales de la langue analysée. Par conséquent, pour qu'une analyse lexicométrique soit cohérente, il est indispensable que les unités analysées soient clairement identifiées. Si cette clarté lexicale est bien présente dans des langues comme le français et l'anglais, elle n'est pas toujours évidente en ce qui concerne

les langues gabonaises. En effet, les langues gabonaises présentent deux difficultés majeures concernant leur analyse lexicométrique :

1. Elles n'ont pas encore de système orthographique reconnu et utilisé par tous ;
2. Ces langues fonctionnent globalement par système de classes préfixales qui introduisent le genre, le nombre et le temps. Du fait de la prononciation, ces préfixes sont souvent accolés aux radicaux des mots lors de l'écriture de ces langues.

Par exemple, dans une langue comme le fang, la phrase « je mange du manioc » s'écrirait de la manière suivante : « madzi mbong ». Cette transcription ne permet pas de mettre en lumière le fait que « madzi » est en réalité composé des morphèmes **mə+adzi**. Ici le préfixe **mə-** fait office de pronom personnel (je, me ou moi) et **-adzi** constitue le radical du verbe manger.

Cette tendance, qui impose automatiquement l'usage d'une écriture conjonctive pour ces langues, pose le problème de la distinction entre les différentes unités de la langue, faussant ainsi les analyses lexicométriques. En effet, rien dans un corpus écrit de manière conjonctive ne permet de distinguer de manière claire toutes les unités qui participent au fonctionnement de la langue. Car une analyse lexicométrique doit pouvoir distinguer chaque unité linguistique afin de pouvoir les classer dans les différents paradigmes des parties du discours (noms, verbes, adjectifs, adverbes, prépositions, etc.).

Pour surmonter cette difficulté, et en attendant l'adoption de leurs systèmes et modes d'écriture, nous proposons que les corpus en langues gabonaises soient transcrits avec un système hybride qui associerait à la fois une démarche conjonctive et disjonctive. En effet, nous proposons que l'usage de l'apostrophe soit systématique dans les situations où les unités linguistiques analysées seraient composées de deux (2) ou plusieurs morphèmes et que ces morphèmes renverraient à des unités distinctives significatives dans la langue. L'apostrophe va non seulement permettre de remplacer le son assimilé (celui qui ne se prononce pas), mais aussi maintenir la fluidité de la prononciation. Par conséquent, sa présence (l'apostrophe) indiquerait au programme lexicométrique utilisé le nombre de morphèmes présents dans l'unité lexicale en question, à condition que les morphèmes correspondent à

des éléments des parties du discours dans le fonctionnement de la langue.

Par exemple la phrase : « je vais à l'école tous les jours » peut s'écrire en fang des deux (2) manières suivantes :

1. En écriture conjonctive : « **makə asikolo məlu məsə.**
2. En écriture disjonctive : « **ma kə a sikolo mə lu mə sə.**

Comme nous pouvons le voir, l'écriture conjonctive met l'accent sur la prononciation sans tenir compte des unités qui composent la phrase. Dans celle-ci il est quasi impossible de distinguer le sujet du verbe, l'article du nom ou le singulier du pluriel. Avec cette écriture le programme lexicométrique TenkaText n'identifiera que quatre unités linguistiques dans cette phrase, alors qu'en réalité la phrase contient plus d'unité que cela.

L'écriture disjonctive elle fait bien la séparation des unités significatives distinctives (les morphèmes), mais ne respecte pas la langue telle qu'elle est prononcée dans son usage par les locuteurs.

À notre avis, non seulement il est important que toutes les unités significatives distinctives de la langue soient mises en évidence pour permettre au logiciel lexicométrique de les répertorier, mais aussi de respecter la manière dont la langue est prononcée. Pour concilier donc ces deux aspects, nous proposons l'usage d'une écriture hybride, qui, non seulement va maintenir ensemble les éléments qui doivent le rester, et va séparer ceux dont la distinction est nécessaire.

À cet effet, nous proposons d'utiliser l'apostrophe comme élément séparateur et distinctif. En d'autres termes, l'usage de l'apostrophe permettrait non seulement de mettre en lumière les différentes unités lexicales présentes dans la phrase, mais donnerait aussi au programme lexicométrique utilisé la possibilité d'analyser le comportement de chacune d'elle. Il est important de préciser que nous préconisons l'usage de l'apostrophe uniquement si le morphème qu'il permet de mettre en évidence existe réellement dans la langue analysée. Avec une écriture hybride ayant recours à l'apostrophe, on peut faire la distinction des unités de la phrase tout en respectant la prononciation.

Avec une écriture hybride la phrase s'écrirait de la manière suivante : « **m'akə a sikolo mə'lu mə'sə.**

Cette distinction est d'autant plus importante que lors de la sélection des signes lemmes, le lexicographe est renseigné sur le fait que les unités lexicales telles que **m'akə**, **mə'lu** et **mə'sə** sont composés de plusieurs morphèmes et que certains de ces morphèmes peuvent jouer le rôle de pronom personnel ou être la marque du genre ou du nombre. Par conséquent, ces unités doivent faire l'objet d'un traitement particulier lors de la sélection de la macrostructure et la rédaction de la microstructure du dictionnaire. Dans la présentation macrostructurale, le lexicographe se doit de fournir les données morphologiques de ces unités linguistiques pour une meilleure compréhension des usagers.

4.2. Exemple d'analyse avec une écriture conjonctive

Figure 4 : Liste de fréquence des mots à partir d'un corpus fang en écriture conjonctive avec TenkaText.

Rank	Word	Freq	%	Tests	%	Lemmas	Set
1	MERA	13	6.22 %	1	100.00 %		
2	ADA	11	5.28 %	1	100.00 %		
3	ABAN	8	4.31 %	1	100.00 %		
4	ABAN	3	2.03 %	1	100.00 %		
5	AGANA	4	1.91 %	1	100.00 %		
6	BAN	4	1.91 %	1	100.00 %		
7	A	3	1.44 %	1	100.00 %		
8	ZAGAH	3	1.44 %	1	100.00 %		
9	ADAL	3	1.44 %	1	100.00 %		
10	ADAL	3	1.44 %	1	100.00 %		
11	B	3	1.44 %	1	100.00 %		
12	BON	3	1.44 %	1	100.00 %		
13	M	3	1.44 %	1	100.00 %		
14	MEGAH	3	1.44 %	1	100.00 %		
15	MEGAH	3	1.44 %	1	100.00 %		
16	MEGAH	3	1.44 %	1	100.00 %		
17	MEGAH	3	1.44 %	1	100.00 %		
18	R	3	1.44 %	1	100.00 %		
19	AD	2	0.96 %	1	100.00 %		
20	ABAGHA	2	0.96 %	1	100.00 %		
21	ABACHLE	2	0.96 %	1	100.00 %		
22	ABALA	2	0.96 %	1	100.00 %		
23	AGANE	2	0.96 %	1	100.00 %		
24	ATARE	2	0.96 %	1	100.00 %		
25	AV	2	0.96 %	1	100.00 %		
26	BA	2	0.96 %	1	100.00 %		
27	GAM	2	0.96 %	1	100.00 %		
28	KAM	2	0.96 %	1	100.00 %		
29	KE	2	0.96 %	1	100.00 %		
30	L	2	0.96 %	1	100.00 %		
31	ME	2	0.96 %	1	100.00 %		
32	ADARE	1	0.48 %	1	100.00 %		
33	ADAGA	1	0.48 %	1	100.00 %		
34	ADAGLI	1	0.48 %	1	100.00 %		
35	ADAGAN	1	0.48 %	1	100.00 %		

129 time needed by core 5.017, time needed by gui 5.012

Source : Afane-Otsaga (2021), Cours d'analyse de données Master 2 Sciences du Langage, Université Omar Bongo.

Comme nous le voyons dans la figure 4, le résultat de l'analyse du corpus fang en écriture conjonctive ne permet pas de mettre en évidence le fait que les unités linguistiques dans cette langue sont généralement composées de préfixes introduisant le nombre, le genre ou le temps.

4.3. Exemple d'analyse avec une écriture disjonctive

Figure 5 : Liste de fréquence des mots à partir d'un corpus fang en écriture disjonctive avec TenkaText.

Overall	Frequency	Alphabetical	Statistics	Notes
1	A	10	4.38 %	1 100.00 %
2	ABA	10	4.38 %	1 100.00 %
3	ME	10	4.38 %	1 100.00 %
4	MIBA	10	4.38 %	1 100.00 %
5	BAM	8	3.51 %	1 100.00 %
6	B	6	2.63 %	1 100.00 %
7	ABAM	5	2.19 %	1 100.00 %
8	ABAN	5	2.19 %	1 100.00 %
9	BA	5	2.19 %	1 100.00 %
10	ABANA	4	1.75 %	1 100.00 %
11	BAN	4	1.75 %	1 100.00 %
12	M	4	1.75 %	1 100.00 %
13	ABAK	3	1.32 %	1 100.00 %
14	ABAL	3	1.32 %	1 100.00 %
15	ABAN	3	1.32 %	1 100.00 %
16	BAGH	3	1.32 %	1 100.00 %
17	BAGHLE	3	1.32 %	1 100.00 %
18	BON	3	1.32 %	1 100.00 %
19	MIBANA	3	1.32 %	1 100.00 %
20	E	3	1.32 %	1 100.00 %
21	AB	2	0.88 %	1 100.00 %
22	ABAGH	2	0.88 %	1 100.00 %
23	ABAGHA	2	0.88 %	1 100.00 %
24	ABCLA	2	0.88 %	1 100.00 %
25	ABAME	2	0.88 %	1 100.00 %
26	AY	2	0.88 %	1 100.00 %
27	DAM	2	0.88 %	1 100.00 %
28	E	2	0.88 %	1 100.00 %
29	KAN	2	0.88 %	1 100.00 %
30	KE	2	0.88 %	1 100.00 %
31	L	2	0.88 %	1 100.00 %
32	MIBAGH	2	0.88 %	1 100.00 %
33	MIBAN	2	0.88 %	1 100.00 %
34	NE	2	0.88 %	1 100.00 %
35	ARA	1	0.44 %	1 100.00 %

529. Time needed by core: 0.015. Time needed by gui: 0.009

Source : Afane-Otsaga (2021), Cours d'analyse de données Master 2 Sciences du Langage, Université Omar Bongo.

Dans la figure 5, le résultat de l'analyse du corpus fang en écriture disjonctive montre que même les morphèmes grammaticaux, qui ne

sont pas des unités lexicales autonomes dans cette langue, sont répertoriés comme étant des morphèmes libres par TenkaText, alors qu'il s'agit en réalité de morphèmes liés. Par exemple le préfixe #me-# (ligne 3) qui peut faire référence aux pronoms personnels *je, me, moi*, etc. est souvent accolé à un radical verbal conjugué : #me-adzi# « medzi » (Je vais manger). Par conséquent, le fait qu'il soit répertorié dans l'analyse du corpus peut pousser le lexicographe à le retenir comme lemme d'un dictionnaire, faussant ainsi les critères de sélection.

4.4. Exemple d'analyse avec une écriture hybride

Figure 6 : Liste de fréquence des mots à partir d'un corpus fang en écriture disjonctive avec TenkaText.

Rank	Word	Frequency	Percentage	Other Stats
1	me	100	4.38 %	1 1000.00 %
2	adzi	100	4.38 %	1 1000.00 %
3	me adzi	100	4.38 %	1 1000.00 %
4	me adzi	100	4.38 %	1 1000.00 %
5	me adzi	8	2.81 %	1 1000.00 %
6	me adzi	8	2.81 %	1 1000.00 %
7	me adzi	8	2.81 %	1 1000.00 %
8	me adzi	8	2.81 %	1 1000.00 %
9	me adzi	8	2.81 %	1 1000.00 %
10	me adzi	4	1.76 %	1 1000.00 %
11	me adzi	4	1.76 %	1 1000.00 %
12	me adzi	4	1.76 %	1 1000.00 %
13	me adzi	3	1.32 %	1 1000.00 %
14	me adzi	2	1.32 %	1 1000.00 %
15	me adzi	2	1.32 %	1 1000.00 %
16	me adzi	2	1.32 %	1 1000.00 %
17	me adzi	2	1.32 %	1 1000.00 %
18	me adzi	2	1.32 %	1 1000.00 %
19	me adzi	2	1.32 %	1 1000.00 %
20	me adzi	2	1.32 %	1 1000.00 %
21	me adzi	2	1.32 %	1 1000.00 %
22	me adzi	2	1.32 %	1 1000.00 %
23	me adzi	2	1.32 %	1 1000.00 %
24	me adzi	2	1.32 %	1 1000.00 %
25	me adzi	2	1.32 %	1 1000.00 %
26	me adzi	2	1.32 %	1 1000.00 %
27	me adzi	2	1.32 %	1 1000.00 %
28	me adzi	2	1.32 %	1 1000.00 %
29	me adzi	2	1.32 %	1 1000.00 %
30	me adzi	2	1.32 %	1 1000.00 %
31	me adzi	2	1.32 %	1 1000.00 %
32	me adzi	2	1.32 %	1 1000.00 %
33	me adzi	2	1.32 %	1 1000.00 %
34	me adzi	2	1.32 %	1 1000.00 %
35	me adzi	2	1.32 %	1 1000.00 %
36	me adzi	2	1.32 %	1 1000.00 %
37	me adzi	2	1.32 %	1 1000.00 %
38	me adzi	2	1.32 %	1 1000.00 %
39	me adzi	2	1.32 %	1 1000.00 %
40	me adzi	2	1.32 %	1 1000.00 %

Source : Afane-Otsaga (2021), Cours d'analyse de données Master 2 Sciences du Langage, Université Omar Bongo.

Dans la figure 6, l'analyse du corpus transcrit en écriture hybride permet de respecter trois critères lexicographiques importantes : (1) le respect de la prononciation de la langue telle qu'elle est réalisée par les locuteurs ; (2) la mise en évidence des préfixes, malgré le fait qu'ils restent accolés aux radicaux ; et (3) l'identification des unités lexicales telles qu'elles existent dans l'utilisation de la langue. Ce qui permet de

surmonter les difficultés rencontrées lors des analyses avec l'écriture conjonctive ou disjonctive.

Conclusion

La lexicométrie est une importante méthode d'analyse de corpus lexicographiques. Car elle permet au chercheur d'obtenir des informations nécessaires sur la langue, particulièrement pour la sélection des unités lemmes de la macrostructure et pour la construction de la microstructure. Cependant, force est de constater que les applications utilisées jusque-là dans le processus lexicométrique ne donnent pas des résultats recherchés avec les langues gabonaises lorsqu'elles sont utilisées de la même manière qu'avec les langues occidentales à l'instar du français. Pour que ces applications fournissent des résultats fiables ou adaptées, nous proposons que les corpus en langues gabonaises soient transcrits dans une écriture hybride, qui est, en quelque sorte, une association de l'écriture conjonctive et de l'écriture disjonctive. L'analyse des données avec un corpus transcrit en écriture hybride de la langue fang nous a permis d'arriver à des résultats probants.

Références bibliographiques

Afane-Otsaga Thierry (2021), *Cours d'analyse de données*. Libreville, Université Omar Bongo.

Bonin Emmanuel et Dallo Alain (2003), « Hyperbase et Lexico 3, outils lexicométriques pour l'historien », *Histoire & mesure*, XVIII - 3/4 : 389-402.

Fuchs Catherine (2022), « La lexicométrie », <https://www.universalis.fr/dictionnaire/lexicometrie>, 21-01-2022.

Hausmann Franz Josef et Wiegand Ernest Herbert (1989), *Component Parts and Structures of General Monolingual Dictionaries: A Survey*,

Hausmann, F. J. et al. (Eds.). (1989-1991): 328-360, Berlin: Walter de Gruyter.

Corsis (2022), https://fr.osdn.net/projects/sfnet_corsis/23-01-2022.

Muller Charles (1992), *Principe et méthodes de la statistique linguistique*, Paris, Honoré Champion.