

“Computational Mapping of Post-Conflict Dynamics in Kasai: An Automated Textual Analysis of Sociopolitical Factors Following the Kamuina Nsapu Insurgency”

LUISHIYE KAPINGA Lucien¹

Faculté des Sciences, Université Pédagogique de Kananga, B.P. 282- Kananga R. D. Congo.

Abstract. The Kamuina Nsapu insurgency (2016–2017) caused lasting disruption to the stability of Central Kasai in particular and the wider Grand Kasai region in general, in the Democratic Republic of the Congo. More than eight years after the end of the violence, humanitarian organizations continue to report deep and persistent socio-economic and institutional consequences. This study offers a computational statistical analysis of institutional and media discourses related to post-conflict dynamics. Drawing on a multi-source corpus composed of reports from non-governmental organizations, United Nations agencies, and international media outlets, a complete text-mining pipeline is applied. The findings reveal the predominance of specific post-conflict factors as well as strong structural relationships among them. Finally, the study proposes a typology of post-conflict discourses and discusses its implications for the design and orientation of public policies.

Keywords: Computational mapping; Post-conflict dynamics; Textual analysis; Sociopolitical factors; Automated analysis; Insurgency; Kamuina Nsapu; Grand Kasai.

Cartographie computationnelle des dynamiques post-conflit au Kasai : une analyse textuelle automatisée des facteurs sociopolitiques à l'issue de l'insurrection Kamuina Nsapu.

Résumé. L'insurrection Kamuina Nsapu (2016-2017) a durablement affecté la stabilité du Kasai Central en particulier et de l'espace du Grand Kasai en général, en République démocratique du Congo. Plus de huit ans après la cessation des violences, les organisations humanitaires continuent de documenter la persistance de profondes séquelles socio-économiques et institutionnelles. Cette étude propose une analyse statistique computationnelle des discours institutionnels et médiatiques relatifs aux dynamiques du post-conflit. A partir d'un corpus multi-sources composé de rapports d'organisations non gouvernementales, d'agences des Nations Unies et d'articles de la presse internationale, un pipeline intégral de fouille de textes est mis en œuvre. Les résultats mettent en évidence la prépondérance de certains facteurs post-conflit ainsi que l'existence de relations structurelles significatives entre ceux-ci. L'étude aboutit enfin à l'élaboration d'une typologie des discours post-conflit et en discute les implications pour l'orientation et la formulation des politiques publiques.

Mots clés : Cartographie computationnelle, Dynamique post-conflit, Analyse textuelle, Facteurs sociopolitiques, Automatisée, Insurrection, Kamuina Nsapu, Grand-Kasai.

Introduction

Les violences liées à l'insurrection *Kamuina Nsapu*, survenues principalement dans la province du Kasai Central entre 2016 et 2017, ont fortement bouleversé l'équilibre social, économique et institutionnel de cette région, ainsi que de l'ensemble de l'espace du Grand Kasai, en République démocratique du Congo. Cette crise armée, d'une intensité exceptionnelle, s'est traduite par des conséquences humaines et matérielles considérables, faisant plus de quatre mille morts, laissant des milliers de personnes portées disparues et provoquant le déplacement interne d'environ 1,4 million de civils. À ces pertes s'ajoutent une destruction massive des infrastructures publiques et communautaires, ainsi qu'un effondrement durable de l'accès aux services sociaux de base, notamment l'éducation, la santé et l'alimentation, aggravant ainsi des vulnérabilités structurelles déjà préexistantes dans la région (HRC, 2018). Bien que les affrontements armés aient officiellement cessé, les séquelles du conflit continuent de peser lourdement sur les dynamiques locales, révélant la complexité et la profondeur des défis post-conflit auxquels fait face le Kasai Central.

Dans la période qui a suivi la fin des violences, de nombreuses organisations de défense des droits humains et institutions internationales telles que les agences des Nations Unies, l'UNICEF, Amnesty International, la Fédération internationale

¹ Chef de Travaux à l'Université Pédagogique de Kananga, Département de Mathématique, Statistique Informatique.

pour les droits humains (FIDH), *Human Rights Watch* ou encore *ReliefWeb*, ont produit une abondante littérature visant à documenter l'ampleur de la crise, les conditions de vie des populations affectées et les besoins humanitaires persistants. À ces rapports institutionnels s'ajoutent les productions médiatiques de la presse internationale, qui ont largement relayé les dimensions humanitaires, sécuritaires et économiques de la crise du *Kasaï*. Toutefois, malgré cette richesse documentaire, force est de constater que ces corpus textuels demeurent rarement exploités de manière systématique et intégrée à l'aide de méthodes statistiques et computationnelles avancées. L'absence d'analyses quantitatives et automatisées appliquées à ces discours limite la capacité à dégager une compréhension globale et structurée des facteurs qui organisent réellement la période post-conflit, ainsi que des relations complexes qu'ils entretiennent entre eux.

C'est précisément à ce niveau que s'inscrit la présente étude, qui se propose de combler cette lacune méthodologique par une analyse statistique et textuelle approfondie des discours institutionnels et médiatiques relatifs à la phase post-insurrectionnelle du *Kasaï* Central. En s'appuyant sur des techniques de collecte automatisée de données documentaires et de *scrapping*, un corpus multi-sources a été constitué à partir de rapports et d'articles issus d'organisations et de médias reconnus, notamment l'*UNICEF*, *OCHA-ReliefWeb*, la *MONUSCO*, *Human Rights Watch*, *Amnesty International*, la *FIDH*, *UN-News*, ainsi que des médias internationaux tels que *RFI*, *BBC*, *Deutsche Welle Afrique*. Ces sources offrent une couverture riche et diversifiée des dimensions humanitaires, sécuritaires, sociales et économiques du post-conflit. L'étude mobilise ensuite des méthodes statistiques, telles que le test du Khi-deux, afin d'examiner les relations de dépendance entre les facteurs identifiés et le lexique mobilisé, ainsi que des techniques d'analyse textuelle automatisée destinées à extraire les structures lexicales dominantes, les thématiques récurrentes et les dynamiques discursives sous-jacentes.

Dans ce contexte, la question centrale qui guide cette recherche est la suivante : « quels sont les facteurs post-conflit les plus récurrents, dominants et structurants dans les rapports institutionnels et les discours médiatiques consacrés au *Kasaï* après l'insurrection *Kamuina Nsapu* ? » Cette interrogation générale se décline en plusieurs sous-questions analytiques, visant notamment à identifier les facteurs les plus fréquemment évoqués, à analyser leur organisation statistique en termes d'association ou d'opposition, à dégager des profils distincts de discours post-conflit et à mettre en évidence les clusters thématiques qui structurent les dynamiques observées. À travers cette démarche, l'étude ambitionne de proposer une lecture intégrée et objectivée des discours post-conflit, susceptible d'éclairer tant la recherche académique que l'élaboration et l'orientation des politiques publiques.

Sur le plan de l'organisation, cette recherche est structurée autour de quatre sections principales, en dehors de la présente introduction et de la conclusion. La première section est consacrée au cadre théorique et à la formulation des hypothèses de recherche. La deuxième section expose la méthodologie adoptée, incluant la constitution du corpus, les outils statistiques et les techniques d'analyse textuelle mobilisées. La troisième section présente les résultats empiriques et leur interprétation, tandis que la quatrième section ouvre la discussion sur les apports, les limites et les implications de la recherche.

1. Cadre théorique et hypothèses

1.1. Cadre théorique

1.1.1. Analyse statistique des données textuelles (ASDT)

La littérature consacrée à l'analyse des contextes post-conflit repose traditionnellement sur des approches essentiellement qualitatives, telles que les rapports narratifs, les enquêtes de terrain ou les études de cas approfondies. Si ces méthodes offrent une compréhension fine des expériences locales et des dynamiques sociales, elles présentent néanmoins des limites lorsqu'il s'agit d'appréhender de manière systématique de vastes ensembles de discours produits sur une longue période. Or, la multiplication croissante des rapports émanant des organisations non gouvernementales, des agences internationales, ainsi que des articles de presse et documents institutionnels, constitue aujourd'hui une masse textuelle considérable. Cette abondance de données ouvre de nouvelles perspectives méthodologiques, en permettant une exploitation statistique et computationnelle des discours relatifs aux dynamiques post-conflit.

Dans ce contexte, l'analyse statistique des textes, communément désignée sous le terme de « *textométrie* », apparaît comme un outil méthodologique particulièrement pertinent pour objectiver les discours. Elle permet d'identifier les régularités lexicales, de mettre en évidence des structures thématiques récurrentes et de tester l'existence de relations statistiquement significatives entre différents facteurs discursifs (Lebart, Salem & Berry, 1998). En dépassant la simple lecture interprétative des textes, la *textométrie* offre la possibilité de fonder l'analyse sur des indicateurs mesurables et reproductibles, tout en conservant une articulation étroite avec l'interprétation qualitative des résultats.

La *textométrie* peut être définie comme une branche spécifique de l'analyse statistique des corpus textuels, visant à produire des cartographies thématiques et relationnelles à partir de textes. Elle mobilise pour ce faire des méthodes telles que les analyses factorielles, les classifications automatiques, les contrastes statistiques ou encore des indicateurs d'évolution lexicale. Cette approche constitue le socle méthodologique des outils qui génèrent des représentations visuelles des discours, en combinant rigueur statistique et capacité exploratoire. En ce sens, la *textométrie* articule des démarches quantitatives et qualitatives, permettant d'explorer les corpus de manière systématique tout en rendant compte de la dynamique et de la complexité des phénomènes discursifs.

Plus largement, l'analyse de données textuelles, qu'elle relève de la *textométrie* ou de la *lexicométrie*, s'appuie sur un ensemble d'outils complémentaires. Ceux-ci incluent des instruments *lexicométriques* tels que les analyses de fréquences et de cooccurrences, des tests statistiques comme le khi-deux, ainsi que des méthodes factorielles notamment l'analyse factorielle des correspondances (AFC), l'analyse en composantes principales (ACP) ou l'analyse des correspondances multiples (ACM). L'intégration de ces techniques permet de représenter les relations structurelles et thématiques au sein des corpus sous forme de cartes sémantiques et visuelles. Comme le soulignent Lebart et al. (1998), ces méthodes rendent possible le passage d'un texte brut à une représentation mathématique interprétable des discours, facilitant ainsi l'analyse comparative et longitudinale des dynamiques discursives.

1.1.2. Cartographie computationnelle

L'analyse computationnelle des textes s'inscrit dans cette continuité méthodologique en visant la production de représentations cartographiques des motifs, des thèmes et des relations présentes au sein de grands corpus textuels. Elle repose sur la combinaison de traitements statistiques et de flux de travail computationnels reproductibles, permettant d'assurer la robustesse et la transparence des analyses (Stoltz & Taylor, 2024). Cette démarche ne se limite pas à la description des contenus textuels, mais cherche à modéliser les structures profondes des discours et leurs évolutions dans le temps. La cartographie computationnelle peut ainsi être définie comme un ensemble de méthodes textométriques et de visualisations factorielles visant à représenter, sous forme de cartes, les transformations discursives observées dans des corpus associés à un contexte donné. Elle mobilise des outils statistiques et informatiques afin de rendre lisibles les relations entre thèmes, acteurs et facteurs, tout en mettant en évidence les dynamiques et les ruptures discursives propres à une période ou à un événement spécifique.

Dans le cadre de la présente étude, la cartographie computationnelle des dynamiques post-conflit s'articule autour de plusieurs composantes essentielles : la constitution d'un corpus textuel pertinent et multi-sources, le choix de méthodes et de techniques d'analyse adaptées, ainsi que la production de résultats sous forme de cartes interprétables. Cette approche vise à offrir une lecture structurée et objectivée des discours post-conflit, tout en fournissant des éléments analytiques utiles à la compréhension des enjeux sociopolitiques et à l'élaboration de politiques publiques fondées sur l'analyse des données.

1.2. Hypothèses de recherche

Afin d'examiner de manière systématique l'organisation et la structuration des discours relatifs aux dynamiques post-conflit, cette étude s'appuie sur un ensemble d'hypothèses visant à mettre en évidence les régularités statistiques, les profils discursifs et les configurations thématiques sous-jacentes au corpus analysé. Il s'agit notamment :

- La première hypothèse postule que les facteurs caractérisant le post-conflit ne sont pas distribués de manière aléatoire au sein du corpus. Elle suggère au contraire que certains facteurs présentent une dominance statistique, reflétant leur centralité dans les discours institutionnels et médiatiques.
- La deuxième hypothèse avance que les unités lexicales associées à certains facteurs post-conflit affichent des valeurs du test du khi-deux statistiquement significatives. Cette hypothèse repose sur l'idée que des liens forts existent entre des groupes lexicaux spécifiques et des thématiques particulières du post-conflit.
- La troisième hypothèse soutient que les discours post-conflit s'organisent autour de profils discursifs distincts et relativement homogènes. Ces profils seraient étroitement associés à des facteurs spécifiques, traduisant des logiques discursives différencierées selon les contextes et les acteurs de production des textes.
- Enfin, la quatrième hypothèse émet l'idée que les dynamiques post-conflit se structurent en clusters thématiques clairement différencierés et statistiquement significatifs. Ces ensembles thématiques refléteraient l'existence de grandes configurations discursives organisant la représentation du post-conflit au sein du corpus étudié.

Dans leur ensemble, ces hypothèses visent à guider l'analyse computationnelle des discours post-conflit en articulant rigueur statistique et interprétation substantielle des résultats. Leur validation ou leur infirmation permettra non seulement d'objectiver les logiques discursives dominantes, mais également de mieux comprendre la manière dont les facteurs sociopolitiques du post-conflit sont construits, hiérarchisés et mis en relation dans les productions institutionnelles et médiatiques. Ce cadre hypothétique constitue ainsi un fondement analytique essentiel pour éclairer les enjeux de gouvernance et d'élaboration des politiques publiques dans les contextes post-conflit.

2. Méthodologies

2.1. Corpus et préparation des données

Le corpus est composé d'extraits textuels provenant de rapports d'organisations non gouvernementales, d'articles de presse et de documents institutionnels relatifs à la période post-conflit. Chaque unité d'analyse est caractérisée par les variables suivantes : la source du document, la date de publication, l'extrait textuel, les facteurs post-conflit identifiés, ainsi que la période considérée comme la période post-conflit.

2.2. Prétraitement

Les étapes ci-après ont été mises en œuvre : nettoyage linguistique des textes (suppression de la ponctuation et des chiffres), élimination des mots vides (*stopwords*), lemmatisation, puis construction d'une matrice terme-document binaire. Les corpus ont ainsi été nettoyés, normalisés et transformés en matrices lexicales exploitables à l'aide du logiciel Python 3.8.10, en mobilisant plusieurs bibliothèques spécialisées, telles que Pandas, NLTK, et autres outils dédiés à l'analyse textuelle.

2.3. Méthodes et techniques d'analyse mobilisées

2.3.1. Nuage de mots (Word Cloud)

Le nuage de mots, ou *Word Cloud*, est une méthode de visualisation exploratoire des données textuelles qui permet de représenter graphiquement les unités lexicales d'un corpus, pondérées en fonction de leur fréquence ou de leur importance statistique. Les mots apparaissant le plus fréquemment sont généralement affichés avec une taille proportionnelle à leur occurrence, offrant ainsi une lecture immédiate des thématiques dominantes au sein du corpus.

Selon Heimerl et ses collaborateurs (2014), le nuage de mots constitue une forme d'analyse visuelle destinée à « condenser une grande masse textuelle en une représentation synthétique et intelligible » (Heimerl, F., Lohmann, S., Lange & Ertl, T., 2014). Dans le champ de l'analyse statistique des données textuelles (ASDT), le nuage de mots est généralement considéré comme un outil descriptif préliminaire, permettant d'appréhender les éléments lexicaux saillants avant de recourir à des méthodes multivariées plus formelles, telles que l'analyse factorielle des correspondances (AFC), l'analyse des correspondances multiples (ACM) ou l'analyse en composantes principales (ACP).

D'après Lebart, Salem et Berry (1998), l'analyse textuelle s'effectue habituellement en trois grandes étapes : Prétraitement et description lexicale ; Analyse statistique exploratoire ; et Analyse factorielle et modélisation. Le nuage de mots s'inscrit pleinement dans la première étape, aux côtés des analyses de fréquences lexicales, des listes de mots caractéristiques et des indices de spécificité. Il permet d'identifier les thèmes dominants, de repérer les lexiques structurants et de détecter d'éventuelles redondances discursives. Par ailleurs, la construction d'un Word Cloud repose essentiellement sur la fréquence d'occurrence des formes lexicales, notée :

$$f(w_i) = \sum_{i=1}^N x_{ij}$$

Où : w_i est le mot i

x_{ij} est le nombre d'occurrences du mot i dans le document j

N est le nombre total de documents

2.3.2. Tests statistiques du *chi-deux* (χ^2)

Le test du *khi-deux* (χ^2) est un test statistique non paramétrique introduit par Karl Pearson en 1900, destiné à évaluer l'existence d'une association significative entre deux variables qualitatives (Saporta, 2011). Ce test repose sur la comparaison entre les fréquences effectivement observées dans un tableau de contingence et les fréquences théoriques attendues sous l'hypothèse d'indépendance des variables. Lorsque les écarts entre les fréquences observées et attendues sont faibles, cela suggère que les variables considérées sont indépendantes. En revanche, si ces écarts sont importants, ils indiquent l'existence d'une association statistiquement significative entre les variables. Considérons un tableau de contingence comportant r lignes et c colonnes. La statistique du *khi-deux* est alors définie par la formule suivante :

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Où : O_{ij} = effectif observé
 E_{ij} = effectif théorique attendu
 $E_{ij} = \frac{(n_i \times n_j)}{n}$

Le test suit asymptotiquement une loi du Khi-deux à :
 $(r - 1)(c - 1)$ Degré de liberté

2.3.3. Le test du *khi-deux* en analyse de données textuelles

En analyse textuelle, le test du *khi-deux* repose sur la construction d'un tableau de contingence lexical, qui permet de croiser les unités lexicales avec les catégories d'analyse. Dans ce tableau, les lignes représentent généralement les documents, les périodes, les catégories ou les facteurs étudiés, tandis que les colonnes correspondent aux mots ou lemmes. Les cellules indiquent la fréquence d'apparition de chaque mot dans une catégorie donnée, constituant ainsi la base de calcul des statistiques. Cette structuration permet d'analyser de manière systématique la répartition lexicale au sein des différents ensembles textuels.

Par ailleurs, le χ^2 global en analyse textuelle mesure l'écart général entre la distribution lexicale observée et une distribution théorique uniforme ou indépendante, permettant d'évaluer si le vocabulaire est structuré différemment selon les catégories analysées (Lebart & Salem, 1994). Concrètement, il répond à la question suivante : les discours sont-ils lexicalement distincts d'une catégorie à l'autre ? L'interprétation du χ^2 global se fait selon le seuil de signification statistique. Un χ^2 global significatif ($p < 0,05$) indique que les catégories présentent des discours lexicalement distincts, ce qui suggère une structuration thématique ou factorielle différenciée des textes. En revanche, un χ^2 global non significatif ($p \geq 0,05$) traduit une homogénéité générale des discours, indiquant que les variations lexicales entre catégories ne sont pas statistiquement marquées.

Par contre, le test du *khi-deux* par mot (χ^2 lexical), ou test du *khi-deux* appliqué à chaque mot, permet d'identifier les termes surreprésentés ou sous-représentés dans une catégorie particulière. Il répond à la question suivante : quels mots caractérisent de manière significative un facteur ou un type de discours ? ... Chaque mot est ainsi testé individuellement par rapport à sa distribution attendue au sein des différentes catégories. Pour un mot donné m et une catégorie c , le χ^2 lexical permet de quantifier l'écart entre la fréquence observée et la fréquence attendue, révélant les termes qui structurent et différencient les discours. Cette approche offre une vision fine de la manière dont le vocabulaire se concentre autour de certains thèmes ou facteurs, contribuant à la cartographie des dynamiques discursives au sein du corpus étudié.

$$\chi^2(m) = \sum \frac{(O_m - E_m)^2}{E_m}$$

où : O_m = fréquence observée du mot
 E_m = fréquence attendue

Tableau 1: Interprétation de χ^2 par mot

Valeur χ^2	Interprétation
Élevée + $p < 0,05$	Mot fortement caractéristique
Faible + $p > 0,05$	Mot faiblement discriminant
Très faible	Mot banal, transversal

Interprétation : Le tableau ci-dessus illustre l'interprétation des valeurs du χ^2 lexical. Une valeur élevée avec $p < 0,05$ désigne un mot fortement caractéristique d'une catégorie, tandis qu'une valeur faible avec $p > 0,05$ indique un mot peu discriminant. Les valeurs très faibles correspondent à des mots banals ou transversaux, présents dans toutes les catégories. Ce classement permet de hiérarchiser les mots selon leur pertinence pour analyser les dynamiques discursives.

2.4. Analyse lexicale (réseau lexical)

Le réseau lexical constitue un outil d'analyse permettant de représenter les cooccurrences de mots au sein des mêmes unités textuelles, qu'il s'agisse de phrases, de documents ou de segments spécifiques. Dans ce type de réseau, chaque mot est représenté par un nœud, tandis que chaque lien, ou arête, traduit le fait que deux mots apparaissent ensemble dans un même contexte textuel, que ce soit une phrase, un document ou un tableau binaire (Polanco, 1995). Les cooccurrences ainsi identifiées permettent de mettre en évidence des champs sémantiques, des thématiques dominantes ou des logiques discursives sous-jacentes. Il convient de préciser qu'il ne s'agit pas d'un réseau sémantique au sens strictement linguistique, mais d'un réseau statistico-discursif. Dans cette représentation, les nœuds correspondent aux mots ou formes lexicales, les liens indiquent une cooccurrence significative ou fréquente entre deux mots, la densité du réseau reflète le degré d'interconnexion du discours, et la centralité mesure l'importance discursive d'un terme, jouant ainsi un rôle de pont sémantique. Ce type de réseau n'a pas pour objectif de tester une hypothèse statistique ; il vise avant tout à cartographier l'organisation sémantique du discours.

L'interprétation d'un réseau lexical repose sur plusieurs notions clés. La centralité lexicale permet d'identifier les mots fortement connectés, souvent transversaux à plusieurs thèmes ou structurants du discours. Les clusters lexicaux, ou communautés lexicales, désignent des sous-ensembles de mots fortement interconnectés. Chaque cluster peut être interprété comme un champ thématique latent, émergent sans hypothèse préalable, et révélant ainsi la structuration interne du discours. Par ailleurs, certains mots jouent le rôle de passerelles, ou brokers lexicaux, en reliant plusieurs clusters et en assurant une médiation discursive entre différentes thématiques. Enfin, la densité du réseau constitue un indicateur de redondance thématique : un réseau dense reflète un discours homogène, avec des thématiques fortement interconnectées et une spécification lexicale relativement faible.

3. Résultats et Interprétations

3.1. Statistiques descriptives principales

Le corpus étudié est constitué de dix-neuf documents, incluant des rapports d'organisations non gouvernementales, des bulletins humanitaires ainsi que des articles de presse, couvrant la période post-insurrection. Ce corpus permet d'appréhender de manière représentative les discours institutionnels et médiatiques relatifs aux dynamiques post-conflit. Le nombre total de documents analysés est donc de $N = 19$. Les sources principales comprennent des institutions internationales et des ONG reconnues, telles que l'UNICEF, *Human Rights Watch* (HRW), le Bureau de la coordination des affaires humanitaires de l'ONU (OCHA), le Haut-Commissariat des Nations Unies pour les réfugiés (UNHCR), *Amnesty International* et *ReliefWeb*, ainsi que des médias internationaux tels que *RFI*, *BBC*, *UN-News* et *DW Afrique*.

3.2. Visualisation descriptives : Word Cloud (Nuage de mots)



Figure 1 : Word cloud global du corpus

Interprétation : Le nuage de mots (*WordCloud*) illustre les termes les plus souvent employés dans les extraits de rapports et d'articles traitant de la situation post-conflit liée à l'insurrection Kamuina Nsapu. La taille de chaque mot est proportionnelle à sa fréquence d'apparition : plus un mot est affiché en grand, plus il occupe une place centrale dans le discours des documents analysés. Les mots les plus dominants mettent ainsi en évidence les priorités du post-conflit, telles que la santé, la réintégration, la réparation et la justice.

3.3. Thèmes dominants mis en évidence

1. Protection et besoins essentiels

1.1. Protection et besoins essentiels

Les mots les plus saillants du corpus, tels que «enfants», «services», «besoin», «accès», «protection», «humanitaire», «santé», «déplacés» ou encore «retour», mettent en évidence une préoccupation centrale pour la survie et la protection des populations vulnérables. Cette fréquence lexicale traduit le poids humanitaire des contextes post-conflit, dans lesquels la priorité est accordée à la fourniture des services essentiels, incluant les soins de santé, l'éducation, l'alimentation, le logement ainsi que la sécurité physique. Parmi ces populations, les enfants apparaissent comme les principaux bénéficiaires des interventions, mais également comme les victimes les plus exposées, soulignant l'importance de stratégies ciblées pour répondre à leurs besoins spécifiques dans un environnement fragile et instable.

2. Réparation et justice

Les termes tels que « justice », « réparation », « enquêtes », « atrocités », « exécutions », « violences », « demandes » ou « droits » reflètent une forte aspiration à la justice transitionnelle. Ils traduisent une mémoire collective profondément marquée par le conflit, ainsi qu'un besoin persistant de reconnaissance des victimes, de réparation des préjudices subis et de lutte contre l'impunité. Par ailleurs, la présence d'expressions telles que « enquêtes indépendantes » ou « mécanismes psychologiques » souligne l'importance accordée à la documentation rigoureuse des violations commises, ainsi qu'à la prise en charge psychologique et au soutien des survivants, illustrant la dimension humaine et réparatrice de ces discours.

3. Traumatismes et santé psychosociale

Les termes tels que « traumatisme », « stress », « soutien », « précaire », « survivants », « vulnérables » et « psychosociaux » témoignent de l'impact durable du conflit sur la santé mentale et émotionnelle des populations affectées. La prévalence marquée de ce vocabulaire souligne l'importance cruciale de mettre en place des programmes de réhabilitation et d'accompagnement psychologique adaptés, notamment pour les enfants soldats, les femmes victimes de violences sexuelles, les déplacés internes et les orphelins, qui constituent les groupes les plus exposés et vulnérables aux séquelles psychosociales du conflit.

4. Réintégration et reconstruction communautaire

4. *Réintégration et reconstruction communautaire*

Des termes tels que « réintégration », « cohésion », « initiatives », « recrutés », « communautés », « retour » ou « villages » reflètent la reconstruction progressive du tissu social. Ils traduisent les efforts déployés en matière de réconciliation, le retour volontaire des populations déplacées et la réhabilitation des infrastructures locales. Le terme « cohésion », en particulier, met en évidence la fragilité persistante de la confiance entre communautés, laquelle reste largement tributaire du soutien humanitaire et de l'application de mesures de justice.

5. Facteurs structurels et socio-économiques

Les termes tels que « pauvreté », « subsistance », « agriculture », « foncières », « emploi », « ressources » et « éducation » mettent en évidence que le relèvement économique et social constitue un enjeu central dans le contexte post-conflit. Ils reflètent notamment la dépendance persistante à l'aide extérieure, la reprise lente de la production agricole ainsi que la fragilité du marché local. Par ailleurs, les tensions foncières se révèlent être une source récurrente de conflits résiduels, contribuant à complexifier le processus de reconstruction et de stabilisation socio-économique.

3.4. Structure du discours post-conflit

L'analyse du *Word Cloud* (nuage de mots) révèle que les unités lexicales s'organisent de manière implicite autour de trois grands pôles thématiques :

Tableau 2: Structure du discours post-conflits

Pôle	Principaux termes associés	Interprétation
Humanitaire & Protection	enfants, services, besoin, santé, déplacés, appel, humanitaire	Urgence de survie et assistance immédiate
Justice & Réparation	justice, réparation, enquêtes, victimes, atrocités	Besoin de reconnaissance, d'enquêtes et de justice
Reconstruction & Cohésion	réintégration, cohésion, retour, initiatives, villages	Relèvement, réconciliation et développement durable

Interprétation : Le tableau met en évidence trois pôles thématiques majeurs qui structurent le discours post-conflit. Le premier pôle, *Humanitaire & Protection*, regroupe des termes tels que « enfants », « services » ou « déplacés », soulignant l'urgence de répondre aux besoins immédiats de survie et d'assistance. Le second pôle, *Justice & Réparation*, rassemble des notions comme « justice », « enquêtes » et « victimes », mettant en avant la nécessité de reconnaissance, de réparation et d'investigation sur les atrocités subies. Enfin, le pôle *Reconstruction & Cohésion* inclut des termes tels que « réintégration », « retour » et « villages », reflétant les enjeux de relèvement, de réconciliation et de développement durable au sein des communautés affectées. Dans l'ensemble, ces trois pôles traduisent une articulation cohérente entre urgence humanitaire, justice réparatrice et perspectives de reconstruction socio-politique.

3.5 Score par document et facteur dominant

3.5.1 Pourcentage de Score global de facteur dominant par document

Tableau 3 : Score par document et facteurs dominants

Doc_Index	Score_global_0_100	Dominant_factor	Score_traumatisme	Score_pauvreté	Score_reconstruction	Score_justice	Score_réconciliation	Score_sécurité	Score_déplacement	Score_santé	Score_éducation
0	37,4	sécurité	0	0	8,5	0	5,9	9,8	6,5	6,5	0
1	39,4	justice	0	0	0	31,3	0	0	8,0	0	0
2	45,0	réconciliation	0	0	0	10,4	14,4	12,0	0	8,0	0
3	78,2	déplacement	0	10,7	19,9	9,9	6,9	0	23,0	7,6	0
4	42,8	justice	0	0	0	20,9	0	12,0	0	0	9,8
5	32,1	déplacement	0	0	0	0	0	0	24,1	8,0	0
6	38,6	traumatisme	12,8	0	10,4	0	7,2	0	0	8,0	0
7	39,9	reconstruction	12,2	0	19,9	0	0	0	0	7,6	0
8	49,9	déplacement	12,8	0	10,4	10,4	0	0	16,0	0	0
9	100	traumatisme	25,7	11,2	20,9	0	0	24,1	8,0	0	9,8
10	41,5	justice	0	11,8	0	22,0	7,6	0	0	0	0

11	39,8	reconstruction	0	0	9,5	0	6,6	0	7,3	7,3	8,9
12	55,1	pauvreté	0	35,6	11,0	0	0	0	8,4	0	0
13	78,2	traumatisme	14,3	12,5	0	11,6	8,0	13,4	8,9	8,9	0
14	43,3	reconstruction	0	0	19,1	9,5	0	0	7,3	7,3	0
15	26,8	justice	0	0	0	19,9	6,9	0	0	0	0
16	47,0	traumatisme	23,5	0	9,5	0	6,6	0	0	7,3	0
17	32,9	déplacement	0	10,7	0	0	6,9	0	15,3	0	0
18	66,9	santé	0	0	0	19,9	6,9	0	7,6	23,0	9,3

Interprétation : Le tableau révèle une forte hétérogénéité des situations, avec des scores globaux variant de 26,8 à 100, indiquant des niveaux très contrastés de vulnérabilité. Les facteurs dominants les plus fréquents (traumatisme, déplacement, justice et reconstruction) traduisent un contexte marqué par des crises complexes et durables. Les valeurs élevées associées au traumatisme, au déplacement et à la santé montrent que les impacts psychosociaux, les mobilités forcées et les fragilités sanitaires constituent des enjeux majeurs, souvent combinés à des déficits en sécurité, en justice et en réconciliation. L'éducation apparaît faiblement représentée, tandis que la pauvreté, bien que moins dominante, atteint des niveaux critiques, soulignant son rôle structurel. L'accumulation de plusieurs facteurs chez les scores les plus élevés suggère enfin des vulnérabilités imbriquées, appelant des réponses intégrées plutôt que sectorielles.

3.5.2. Scores moyens par facteur (rapportés)

Top termes (global): services, enfants, rapport, besoins, moyens, réparations, subsistance, stigmatisation, humanitaire, santé, réintégration, tensions, protection, foncières, retours, enquêtes, décrivant, survivants, soutien, victimes, besoin, locales, limité, appel, cohésion, accès, traumatismes, déplacés, l'accès, justice.

Tableau 4: Scores moyens par facteur

Facteurs	Score moyen
déplacement, reconstruction, gouvernance:	78,27
gouvernance, reconstruction, sécurité:	49,90
justice, gouvernance, traumatisme:	41,57
justice, santé, traumatisme:	78,29
justice, sécurité, gouvernance:	26,86
justice, sécurité, traumatisme:	42,86
justice, traumatisme, sécurité:	39,44
pauvreté, santé, sécurité:	38,63
reconstruction, déplacement, santé:	32,19
reconstruction, gouvernance, déplacement:	43,35
reconstruction, pauvreté, emploi:	39,86
reconstruction, pauvreté, sécurité:	33,00
santé, pauvreté, déplacement:	37,46
santé, reconstruction, gouvernance:	66,95
sécurité, gouvernance, déplacement:	45,07
traumatisme, santé, éducation:	39,90
éducation, réinsertion, traumatisme:	55,14
éducation, traumatisme, réinsertion:	73,51

Interprétation : Le tableau met en évidence de fortes disparités entre les différentes combinaisons de facteurs, révélant que certaines articulations produisent des effets nettement plus significatifs que d'autres. Les scores les plus élevés concernent notamment les associations « justice, santé, traumatisme » (78,29) et « déplacement, reconstruction, gouvernance » (78,27), ce qui suggère que l'intégration simultanée des dimensions psychosociales, institutionnelles et matérielles constitue un levier majeur dans les contextes post-crise ou post-conflit. À l'inverse, des combinaisons pourtant centrales comme « justice, sécurité, gouvernance » (26,86) ou « reconstruction, déplacement, santé » (32,19) affichent des scores relativement faibles, traduisant soit une faible synergie entre ces facteurs, soit des politiques fragmentées qui peinent à produire des résultats cohérents. Les regroupements autour de la pauvreté et de la sécurité présentent globalement des scores moyens à faibles, indiquant que l'absence d'une dimension de gouvernance, de

justice ou de santé limite leur impact global. Par ailleurs, les résultats liés à l'éducation et à la réinsertion montrent une dynamique positive, en particulier pour « éducation, traumatisme, réinsertion » (73,51), soulignant le rôle structurant de l'éducation dans les processus de résilience et de reconstruction sociale. Dans l'ensemble, ce tableau suggère que les approches intégrées et multisectorielles, articulant gouvernance, justice, santé et prise en charge des traumatismes, sont celles qui génèrent les effets les plus significatifs, tandis que les interventions isolées ou faiblement coordonnées tendent à produire des résultats plus limités.

3.5.3. Graphique de Scores moyens par facteur (rapportés)

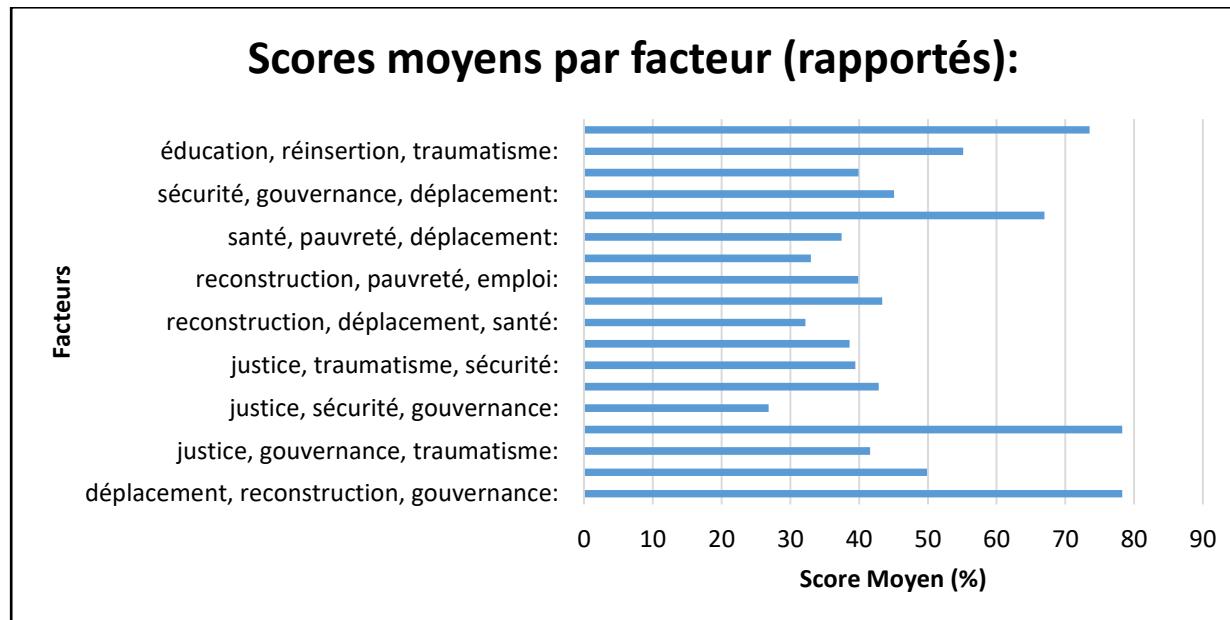


Figure 2 : Scores moyens par facteur

Interprétation : Ce graphique présente un score synthétique (%) indiquant la prépondérance relative des facteurs dans les textes, sans mesurer leur significativité statistique, mais en reflétant leur hiérarchisation discursive. Les combinaisons les plus dominantes ($\approx 70\text{--}85\%$), « justice-gouvernance-traumatisme », « déplacement-reconstruction-gouvernance » et « santé-pauvreté-déplacement », structurent fortement le discours post-conflit. Les facteurs intermédiaires ($\approx 40\text{--}60\%$), tels que « justice-sécurité-gouvernance », « éducation-réinsertion-traumatisme » et « reconstruction-déplacement-santé », occupent une place notable mais secondaire. Enfin, les combinaisons les moins dominantes ($\approx 30\text{--}40\%$), notamment « reconstruction-pauvreté-emploi » et « justice-traumatisme-sécurité », apparaissent plus marginales dans les textes analysés.

3.6. Tests d'association Chi-deux global et par mot

3.6.1. χ^2 Global

Le test du χ^2 de Pearson a été appliqué à la table de contingence, donnant les résultats suivants :

Indicateur statistique	Valeur
Chi-deux de Pearson (χ^2)	2769,196
p-value	0,596
Degrés de liberté (ddl)	2788

Interprétation : Les résultats du test du χ^2 de Pearson indiquent une absence de relation statistiquement significative entre les variables de la table de contingence, comme le montre la p-value élevée ($p = 0,596$), largement supérieure au seuil conventionnel de 0,05. On ne rejette donc pas l'hypothèse nulle d'indépendance globale, ce qui signifie que la distribution des mots n'est pas significativement différente selon les facteurs post-conflit considérés dans leur ensemble. Bien que la valeur du chi-deux soit élevée ($\chi^2 = 2769,196$), le nombre important de degrés de liberté (ddl = 2788) atténue son interprétation et suggère que les écarts observés restent compatibles avec une indépendance statistique. Globalement, ce résultat traduit une certaine homogénéité structurelle des discours post-conflit au Kasaï Central : les rapports d'ONG et humanitaires utilisent un lexique transversal commun à plusieurs facteurs tels que la sécurité, la justice, la santé, le déplacement ou la reconstruction. Cela reflète la nature multidimensionnelle des

situations post-conflit, où les problèmes sont imbriqués et rarement isolés, soulignant la complexité des enjeux à la fois sociaux, institutionnels et humanitaires.

3.6.2. χ^2 par mot

Tableau 5: χ^2 par mot et p-value respective

mot	khi2	p_value
services	5,652173913	0,99523443
enfants	5,090909091	0,99750236
besoins	4,285714286	0,99917954
rapport	4,285714286	0,99917954
traumatismes	3,578947368	0,99975846
victimes	3,2	0,99988983
témoignages	1,789473684	0,99999853
destructions	1,789473684	0,99999853
communautés	1,789473684	0,99999853
crise	1,789473684	0,99999853
conflits	1,789473684	0,99999853
peinent	1,789473684	0,99999853
perturbée	1,789473684	0,99999853
réparation	1,789473684	0,99999853

Interprétation : L'analyse de la table de contingence à l'aide du test du χ^2 de Pearson montre que tous les mots présentent des valeurs de chi² faibles, accompagnées de p-values extrêmement élevées ($\approx 0,995$ à $0,999$). Cela indique qu'aucun mot n'est statistiquement discriminant pour un facteur spécifique pris isolément, et qu'aucun terme ne permet à lui seul d'expliquer une différence significative entre les facteurs. Des mots tels que services, enfants, victimes, traumatismes, destructions ou communautés apparaissent dans presque tous les facteurs, constituant un noyau lexical commun aux discours post-conflit de Kamuina Nsapu. Ces résultats suggèrent que le discours post-conflit n'est pas fragmenté en registres spécialisés, mais qu'il est structuré autour d'un lexique humanitaire unifié, reflétant la multidimensionnalité et l'interdépendance des enjeux sociaux, humains et institutionnels dans la région.

3.7. Interprétation du réseau lexical (cooccurrences)



Figure 3 : Réseau lexical

Interprétation : L'analyse du réseau lexical du discours post-conflit dans le Kasaï Central révèle un graphe dense et fortement interconnecté, caractérisé par une abondance de liens, peu de sous-graphes isolés et de nombreux chemins indirects entre les mots, ce qui suggère un discours peu segmenté avec des thématiques enchevêtrées, cohérent avec le χ^2 global non significatif, les χ^2 par mot faibles et des p-values très élevées. Aucun facteur ou catégorie ne structure donc de manière différenciée le lexique, qui reste homogène tout en étant émotionnellement et socialement chargé. Au sein de ce réseau, certains mots-pivots tels que victimes, services, besoins, traumatismes, justice, santé, déplacements, destructions, enfants et femmes occupent une position centrale et agissent comme des ponts sémantiques reliant le registre humanitaire, psychosocial, des droits humains et socio-économique, montrant que le discours s'articule avant tout autour de la figure de la victime plutôt que d'un clivage institutionnel, politique ou communautaire. Même en l'absence de clusters statistiques nets, on distingue des pôles thématiques implicites et récurrents : un pôle humanitaire et de survie, centré sur des termes tels qu'aide, humanitaire, urgence, nutritionnelle, distributions, subsistance et besoins, exprimant une logique de survie immédiate ; un pôle traumatique et violences, regroupant traumatismes, violences, exécutions, atrocités, extrajudiciaires et stigmatisation, montrant que le traumatisme reste lié aux victimes, aux enfants et aux femmes ainsi qu'aux mécanismes de justice ; et un pôle justice-reconstruction-réparation, comprenant justice, réparations, réintégration, mécanismes, retour et réparation, moins dense mais connecté aux pôles traumatiques et humanitaires, indiquant que la justice apparaît comme un prolongement du traumatisme plutôt qu'un champ discursif autonome. Ces pôles, bien qu'interprétatifs, reflètent des zones de cooccurrence lexicale récurrente sans former de véritables clusters statistiques.

4. Discussion des résultats

L'analyse du *WordCloud* met en évidence que les rapports et témoignages issus du contexte post-conflit Kamuina Nsapu insistent particulièrement sur la protection des enfants et des groupes vulnérables, la recherche de justice et de vérité, la prise en charge des traumatismes persistants, ainsi que sur la reconstruction communautaire et économique fragile. À cela s'ajoute le besoin de cohésion et de réconciliation entre les populations affectées. Autrement dit, le vocabulaire post-conflit met en avant un triangle critique composé de l'Humanité, de la Justice et de la Reconstruction, trois dimensions étroitement interdépendantes pour consolider une paix durable et stabiliser le tissu social.

Les résultats statistiques complètent cette lecture lexicale. Le χ^2 global apparaît non significatif et le χ^2 par mot révèle une faible capacité discriminante, ce qui traduit l'absence d'une structuration conflictuelle forte dans le discours. En parallèle, la hiérarchisation des scores moyens indique que la gouvernance constitue un facteur central, tandis que le déplacement des populations structure de manière importante le post-conflit, en lien avec la santé, la pauvreté et la reconstruction. Le traumatisme, quant à lui, traverse plusieurs dimensions telles que la justice, l'éducation et la réinsertion, révélant sa transversalité dans le discours. La combinaison des analyses lexicales et statistiques montre que le *WordCloud* domine par les termes liés aux besoins fondamentaux et à la justice, tandis que le réseau lexical révèle une forte imbrication thématique. Le graphique des scores moyens met en évidence que certaines combinaisons de facteurs, comme déplacement-reconstruction-gouvernance, justice-gouvernance-traumatisme et santé-pauvreté-déplacement, obtiennent des scores élevés. Cette interconnexion est confirmée par le réseau lexical, qui illustre que ces facteurs circulent entre plusieurs dimensions et qu'aucun facteur ne domine isolément. Ainsi, les scores moyens élevés traduisent davantage une intensité discursive qu'une spécialisation thématique.

En conséquence, le réseau lexical permet de révéler une structuration relationnelle du discours que les tests statistiques seuls ne peuvent identifier. Il met en évidence la centralité des victimes et des besoins sociaux, la complémentarité entre analyses statistiques, graphiques et sémantiques, ainsi qu'un discours post-conflit non polarisé, caractérisé par l'absence de segmentation nette. Ces observations indiquent que le discours reflète une crise globale et intégrée, dans laquelle les facteurs post-conflit au Kasaï Central ne s'opposent pas lexicalement, mais se structurent par cooccurrence et intensité discursive. Globalement, ces résultats dessinent un discours post-conflit cohérent et intégré, centré sur la survie des populations, la réparation des traumatismes et la reconstruction sociale. Ils soulignent que la parole des acteurs du post-conflit ne se limite pas à un registre humanitaire ou judiciaire isolé, mais articule de manière continue les dimensions humanitaires, sociales et économiques, offrant ainsi une lecture complète et relationnelle des enjeux post-Kamuina Nsapu.

Conclusion

Cette étude, fondée sur une analyse statistique des données textuelles relatives au post-conflit Kamuina Nsapu, met en évidence une structure discursive cohérente et intégrée, caractérisée par une dominance lexicale des besoins humanitaires et de la justice, une absence de segmentation lexicale forte et une interconnexion étroite des facteurs sociaux, sanitaires et institutionnels. L'utilisation combinée du *WordCloud*, des tests du χ^2 , des scores moyens et du réseau lexical montre que le discours post-conflit au Kasaï central est moins marqué par des oppositions thématiques que par une logique de cumul et de complémentarité des enjeux, où les différents facteurs se renforcent mutuellement plutôt qu'ils ne s'excluent.

Le réseau lexical des cooccurrences révèle une structuration dense et transversale, dominée par des termes liés aux victimes, aux besoins humanitaires et aux traumatismes. Cette forte interconnexion confirme l'absence de différenciation

statistiquement significative observée dans le test du χ^2 global, traduisant un discours homogène mais intensément chargé sur le plan humanitaire et psychosocial. Le nuage de mots global met en évidence une forte récurrence de facteurs tels que « traumatisme », « sécurité », « déplacement », « reconstruction » et « éducation », soulignant la concentration thématique autour des conséquences humanitaires et sociales immédiates du conflit. Cette dominance lexicale n'est pas aléatoire, mais directement liée à la nature des sources utilisées, notamment les rapports d'ONG et les documents humanitaires, ainsi qu'au contexte post-conflit encore instable, ce qui confirme la première hypothèse de l'étude (H1) selon laquelle certains facteurs centraux dominent nettement le corpus. Les tests statistiques confirment cette lecture : le χ^2 global non significatif montre l'absence de structuration forte entre les facteurs, tandis que les χ^2 par mot, faibles avec des p-values élevées, indiquent qu'aucun terme n'est suffisamment discriminant pour signaler des oppositions claires. Le réseau lexical dense, par ses nombreuses cooccurrences transversales, révèle que les facteurs coexistent plutôt qu'ils ne s'opposent, ce qui infirme la deuxième hypothèse (H2). Le lexique reste partagé entre plusieurs thématiques, sans spécialisation marquée, et le score moyen des mots est modéré et homogène. Le χ^2 ne révèle pas de segmentation nette, suggérant que les discours ne se structurent pas en profils autonomes (humanitaire vs sécuritaire), mais forment un discours globalisant centré sur la gestion des conséquences du conflit. Cette observation confirme partiellement la troisième hypothèse (H3), car des orientations discursives existent, mais sans profils statistiquement distincts.

Enfin, le réseau lexical fortement connecté, combiné à l'absence de communautés nettement séparées et à un χ^2 global non significatif, indique que les « clusters » visuellement identifiés restent interprétatifs et ne sont pas validés statistiquement. Cela infirme la quatrième hypothèse (H4), montrant que les dynamiques post-conflit ne s'organisent pas en clusters thématiques autonomes, mais forment un système discursif fortement imbriqué et intégré. Dans l'ensemble, ces résultats contribuent à la compréhension des dynamiques post-conflit en Afrique centrale et soulignent l'importance d'approches intégrées dans les politiques de reconstruction et de consolidation de la paix, qui doivent tenir compte de l'interdépendance des enjeux humanitaires, psychosociaux et institutionnels.

Références bibliographiques

- [1]. Benzécri, J.-P., *L'analyse des données*. Dunod. 1973
- [2]. Dustin S. Stoltz & Marshall A. Taylor, *Mapping Texts : Computational Text Analysis for the Social Sciences*, Oxford PU., 2024
- [3]. Heimerl, F., Lohmann, S., Lange & Ertl, T., *Explorateur de nuages de mots : Analyse de texte basée sur les nuages de mots*. IEEE, 2014
- [4]. HRC, *Rapport de l'Équipe d'experts internationaux sur la situation au Kasaï*, 2018
- [5]. Jacomy, M., et al., *ForceAtlas2*. PLOS ONE. 2014
- [6]. Lahlou, S., *Méthodes d'analyse lexicale*. PUF. 1995
- [7]. Lebart, L., & Salem, A., *Statistique textuelle*. Dunod., 1994
- [8]. Lebart, L., & Berry, L. *Analyse Exploratoire de données Textuelles*. Kluwer. 1998.
- [9]. Léon Lebart & Salem, Analyse statistique des données textuelles, Dunod, 1988, 1994
- [10]. Polanco , J-M, *Analyse des réseaux lexicaux*. Hermès, n°15, CNRS Editions
- [11]. Saporta, G. *Probabilités, analyse des données et statistique*. Technip., 2011.