UNIVERSITE DU QUEBEC

MEMOIRE PRESENTE A
L'UNIVERSITE DU QUEBEC A TROIS RIVIERES

COMME EXIGENCE PARTIELLE
DE LA MAITRISE EN MATHEMATIQUES
ET INFORMATIQUE APPLIQUEES

PAR
ISSIFOU SOULEYMANE ISSA MAIGA

EXTRACTION DE REGLES D'ASSOCIATION DE TEXTES ECRITS EN HAUSA

Septembre 2025

Résumé

L’augmentation du nombre de documents multimédias stockés sur les supports électroniques, issus de l'utilisation du Web et des réseaux sociaux, fait qu'il devient nécessaire de trouver un compromis entre l'utilisation du Web et le procédé de recherche d'information. La conception d'outils d'analyse et de traitement automatique des contenus de textes écrits en Hausa est alors nécessaire, dans le but d'assister l'utilisateur lors de la lecture et la compréhension des textes. Les outils d'exploration textuelle améliorent le processus de recherche de l'utilisateur en vue de déceler des informations pertinentes. Les informations recueillies lors de l'étape d'exploration peuvent sembler parfois peu significatives. Mais elles ont une signification en termes de recherche. Elles sont bien souvent nombreuses et très volumineuses. Malgré les adaptations apportées, elles restent très bruitées.

Cette ambiguïté constitue un obstacle majeur pour l'homme. En effet, pour ce dernier, une interprétation objective de ces connaissances extraites est très difficile dans un tel cas. Ce mémoire a pour objectif principal l’extraction des règles d’association de textes écrits en Hausa, qui repose essentiellement sur le principe de représentation vectorielle en entrée. Les informations sont présentées au moyen d'une matrice des fréquences.

Notre conception exploite principalement la force des techniques de fouille de données, pour chercher de l'information pertinente sur des données textuelles, essentiellement, les papiers ou journaux sur l’actualité mondiale, mais aussi sur les réseaux sociaux.

Abstract

The increase in the number of multimedia documents stored on electronic media, resulting from the use of the Web and social networks, means that it becomes necessary to find a compromise between the use of the Web and the information search process. . . The design of tools for the analysis and automatic processing of text content written in Hausa is then necessary, with the aim of assisting the user when reading and understanding the texts. Text mining tools enhance the user's search process to uncover relevant information. The information collected during the exploration stage may sometimes seem insignificant. But they have significance in terms of research. They are often numerous and very bulky.

Despite the adaptations, they remain very noisy. This ambiguity constitutes a major obstacle for man. Indeed, for the latter an objective interpretation of this extracted knowledge is very difficult in such a case. The main objective of this dissertation is to extract association rules from texts written in Hausa, which are essentially based on the principle of vector representation as input. The information is presented using a frequency matrix.

Our design mainly exploits the strength of data mining techniques to search for relevant information on textual data, essentially papers or newspapers on world news, but also on social networks.

Dédicace

Je dédie ce travail à :

- *Mon père Souleymane Issa MAIGA ;*
- *Ma mère Rabi BALLA ;*
- *Mon épouse Malika ;*
- *Mes deux sœurs Mariama et Halimatou ;*
- *A mes frères que la nature m'a donnés :*

  Rachid, Zoulkifli, Abdoul Djafar, Aboubacar Fallo, Chaps, Moustapha, Omar goma,

  Fayçal, toute la fada de Dan gao, toute la fada de Dakar;
- *A ma famille, ami(e)s et connaissances ;*
- *A tous mes professeur(e)s de l'UQTR.*

Remerciements

Je remercie mes parents pour tout ce qu'ils ont fait pour moi. Leurs prières, leurs encouragements m'ont permis d'avancer dans la vie.

J'exprime ma profonde gratitude à tous ceux qui, de près ou de loin, ont impacté positivement ma vie en me poussant à donner le meilleur de moi.

J'exprime ma reconnaissance envers M. Ismaïl Biskri, pour qui j'ai beaucoup d'affection et de respect pour sa disponibilité, ses conseils, son encadrement.

Je suis reconnaissant à l'égard de l'administration et du corps professoral de l'UQTR, université d'Excellence Canadienne, pour l'encadrement dont j'ai bénéficié durant ces années.

Je tiens à exprimer ma gratitude aux membres du jury pour avoir accepté d'examiner ce travail et de l'enrichir avec leurs remarques.

Merci à toutes et à tous.

Tables des matières

Liste des Figures

Liste des tableaux

Chapitre 1 : Introduction

Le monde de l'informatique change très rapidement, et la recherche ou fouille des textes est devenue cruciale dans ce domaine. Cette branche a pour vision d'extraire des informations à partir de données, pour les analyser et conclure avec des décisions importantes dans un domaine spécifique.

Dans cette logique, l'extraction des règles d'association de textes écrits en hausa, est une méthode fondamentale qui vise à connaître les liens entre les éléments d'un texte écrit en hausa en se basant sur des critères. Cette technique est essentielle dans de nombreux domaines, tels que la recherche d'informations, la veille technologique, la recommandation de produits ou de services.

L'une des méthodes les plus importantes pour la recherche de données concerne les règles d'association, le principe consistant à étudier et à analyser les bases de données, afin d'en tirer les informations les plus pertinentes. Ces informations sont présentées sous forme conditionnelle entre des antécédents et des conséquents. En d'autres termes, les règles d'association permettent de trouver une relation entre les éléments qui sont souvent utilisés ensemble.

Une règle d'association est une relation d'occurrence qui lie deux ou plusieurs items particuliers. Par exemple, si les paniers d'achats révèlent qu'un grand nombre de clients qui achètent du pain achètent également du beurre, cela indique une relation d'occurrence simultanée forte entre ces deux items.

Cette relation est appelée association et donne lieu à la règle d'association suivante: pain => beurre. Cette règle est dite classique ou traditionnelle, car les deux événements se déroulent simultanément lors de la même transaction. Lorsque l'occurrence des items est étalée selon un ordre spécifique, par exemple, les clients qui achètent du pain achètent également du beurre dans les jours qui suivent, les règles d'association entre ces items ordonnés sont qualifiées de séquentielles. L'objectif des règles d'association, qu'elles soient traditionnelles ou séquentielles, est de découvrir les associations fréquentes dans les données étudiées.

Les règles d'association sont appliquées dans divers domaines. En marketing, elles permettent d'identifier les produits ou services achetés lors d'une même transaction ou par un même client dans le temps, offrant ainsi la possibilité d'identifier des opportunités de ventes croisées. En analysant l'ordre dans lequel les internautes accèdent aux pages d'un site web, les règles d'association séquentielles permettent d'entrevoir quelles modifications rendraient le site plus convivial. Elles sont également utiles dans l'étude des comportements d'achats des consommateurs, en mettant en lumière les comportements à travers le temps. Par exemple, si les associations séquentielles montrent que les propriétaires d'une marque de voiture changent leur véhicule tous les 18 mois, le concessionnaire peut fidéliser sa clientèle en envoyant un dépliant sur les nouveaux modèles disponibles quelques semaines avant cette date.

Néanmoins, cette méthode peut générer un nombre important de règles d'associations, dont certaines sont redondantes et non pertinentes, ce qui présente un problème à résoudre, sur lequel plusieurs travaux de recherches se sont penchés. Il existe plusieurs méthodes de réduction du nombre de règles générées, cependant, elles peuvent engendrer une perte de quelques importantes associations ou même une redondance.

L’évaluation des méthodes d’algorithmes de découverte de règles d’association repose principalement sur l’optimalité, la rapidité d’exécution, l’exhaustivité et la consommation mémoire. La qualité recherchée est de pouvoir réduire le nombre de règles d'association sans perdre des règles utiles et importantes.

Notre travail dans cette recherche consiste à concevoir une approche permettant de réduire le nombre de règles d'association sans perte d'informations pertinentes dans l’extraction de textes écrits en hausa.

Rappelons que très peu de travaux existent sur le hausa, notre travail est le premier en utilisant les règles d’associations par rapport à cette langue et que l’utilisation des règles d’association est un incontournable pour analyser les données. Elle permet de trier les informations essentielles avec un effet d’explicabilité qui fournit les informations dans un format plus accessible, tout en évitant l’effet boîte noire, dont la représentation abstraite et le volume croissant des données.

Ce mémoire est structuré en sept chapitres.

Dans le deuxième chapitre, nous présentons une description non exhaustive de la langue hausa, cruciale, notamment dans les contextes de traitement automatique du langage, d’analyse linguistique ou d’applications socio-économiques. Nous présenterons en détail les principales phases adaptées à un cadre académique ou à une introduction pour un travail sur le traitement automatique du langage ou l’extraction de règles. Ce chapitre nous permettra donc d’avoir toutes les connaissances théoriques concernant le hausa, nécessaires pour la suite de notre travail.

Dans le troisième chapitre, nous présentons le Text Mining, qui est important pour plusieurs raisons, surtout dans un contexte académique, professionnel ou technique, pour clarifier le concept, car le Text Mining peut être confondu avec d'autres disciplines (NLP, analyse de données, etc.). Mais aussi pour faciliter la compréhension pour un public non expert.

Dans le quatrième chapitre, nous présentons les règles d'association, qui constituent une méthode couramment utilisée dans la fouille de données. Nous présenterons en détail les principales mesures utilisées dans l'extraction de ces règles. Par la suite, nous explorerons différents algorithmes utilisés pour l'extraction de règles (apriori…).

Dans le cinquième chapitre, nous présenterons la méthodologie pour l'extraction de règles d'association de textes écrits en hausa. Nous présenterons les différentes bibliothèques utilisées, et expliquerons les méthodes d'application de ces dernières au projet.

Au chapitre 6, nous décrirons la mise en œuvre de notre système, en détaillant le code informatique et le langage de programmation utilisés pour extraire des règles d'association à partir de textes écrits en hausa. Par la suite, dans le septième chapitre, nous exposerons les résultats puis évaluerons les résultats.

Pour enfin conclure en nous basant sur les résultats obtenus.

Chapitre 2 : Description de la langue Hausa

2.1. Historique

Dans ce chapitre, notre objectif n'est pas de présenter de manière exhaustive la langue, mais plutôt d'expliquer l'importance de la langue hausa, parlée par des millions de personnes en Afrique.

Avec le swahili, le hausa est la langue qui compte le plus de locuteurs en Afrique subsaharienne. Il est la langue première d'environ 80 millions de personnes au Nigeria, au Niger, et dans divers pays avoisinants, auquel s'ajoutent quelque 20 millions qui l'utilisent en tant que langue véhiculaire. Il est présent dans les médias et possède une presse et une littérature.

Le hausa appartient à la famille tchadique du phylum afro-asiatique. Il est donc lointainement s'apparente à l'arabe, au berbère, etc. La tradition de l'écriture et de la lecture dans les communautés musulmanes d'Afrique de l'Ouest remonte aux 9e et 10e siècles, époque à laquelle l'islam fut introduit par le commerce transsaharien. La religion fut acceptée par les rois qui l'ont propagée parmi leurs sujets dans les villes et, graduellement, dans les communautés rurales.

Les érudits fondèrent des écoles coraniques qui se sont développées en un système éducatif formel efficace pendant des siècles, particulièrement dans le nord du Nigeria, et ce, jusqu'à l'ère des conquêtes coloniales. Depuis le 15e siècle, où l'Islam devint religion d'Etat en Empire Kanem Bornou, en passant par la réforme religieuse au 19e du Califat de Sokoto, des groupes d'érudits avaient commencé à créer des centres d'apprentissage dans chaque ville et village, lançant une révolution sociale réelle orientée vers la formation intellectuelle (Yahya, 2007).

En effet, ce système de formation était une réalisation considérable et distinctive. Elle a contribué à l'émergence d'une population lettrée. Selon Aisha Lemu (Aisha, 2002), de nombreux éléments de ces populations de Nigérians du Nord et des communautés musulmanes

de l'Afrique de l'Ouest ont eu l'opportunité d'apprendre à lire et à écrire longtemps avant la venue du colonialisme.

2.2. Eléments de Phonologie

Le hausa est une langue qui possède trois tons : haut (H), bas (B) et descendant (HB), que les syllabes supportent. Quoique l'orthographe officielle ne le note pas, le ton suffit à distinguer des lexèmes : kai 'toi(masc.)' (H) vs kâi 'tête' (HB), wujà (H•B) 'cou' vs wùja (B•H) 'difficulté'.

Aussi, les grammaires notent-elles l'accent grave sur le « B », le circonflexe sur le « HB », et ne disent rien à propos du « H ». Le haoussa possède également un accent d'intensité qui se manifeste habituellement sur la première syllabe « H » avant une syllabe « B ».

Le français ignorant tons et accents lexicaux, la seule difficulté pour les hausaphones sera d'entendre et de reproduire l'accent d'intensité sur la dernière syllabe des énoncés déclaratifs, décroissants quant à la hauteur des syllabes successives en hausa.

La syllabe est canoniquement ouverte en hausa (consonne voyelle : CV). Les syllabes fermées (CVC) sont rares.

Le système vocalique se compose de cinq voyelles brèves (/i/, /u/, /e/, /o/ et /a/) ainsi que de cinq voyelles longues (/ī/, /ū/, /ē/, /ō/ et /ā/). L'opposition de longueur, lexicalement pertinente (cf. tàfi 'partir' vs. tàfī 'paume'), ici notée par un macron sur le i, n'est pas transcrite par l'orthographe officielle. Il n'y a pas de voyelle nasale. Les personnes qui parlent le hausa doivent donc apprendre les contrastes d'ouverture du français, qui ne dépendent pas de la longueur, ainsi que les voyelles arrondies hautes et moyennes /y/ (bu), /ø/ (bœufs) et /œ/ (beurre) et les voyelles nasales. Il existe deux diphtongues : /aj/ et /aw/ (graphiée).

En dehors des consonnes communes aux deux langues, le hausa possède plusieurs autres sons inconnus du français, notamment la fricative glottale [h] () ; l'occlusive glottale [ʔ] (non

transcrite initialement, représentée par une apostrophe dans un mot), ainsi que les éjectives et injectives accompagnées ou précédées d'une constriction glottale [kʔ] () et [sʔ] (). Il existe également les combinaisons [ʔb] (), [ʔd] () et [ʔj] (), la fricative bilabiale sourde [ɸ] (), les affriquées [ʧ] () et [ʤ] (), ainsi que le roulement alvéolaire [r] (). Inversement, le hausa ignore le /f/ labiodental (fille), le /ʁ/ uvulaire (Paris), la nasale palatalisée /ɲ/ (vigne), la chuintante /ʒ/ (jaune) (sauf à l'ouest du domaine hausa), la bilabiale /p/ (poule) (sauf à l'est où /ɸ/ se réalise [p]). Les occlusives /b/, /ʔb/, /k/, /kʔ/ et /g/ se palatalisent (prononciation dite « mouillée ») devant les voyelles antérieures (/i/, /e/, /ɛ/) : gidā /gj ídáː/ 'maison' ; elles se labialisent devant les voyelles postérieures (arrondies) : bùhū /bwùhúː/ 'sac'. Les apprenants devront veiller à ne pas transposer ces ajustements au français.

2.3. Eléments de grammaire

2.3.1. Le genre

Comme le français, le hausa possède deux genres grammaticaux, masculin (M) et féminin (F), inhérents pour les noms, par accord pour les adjectifs.

Pour les entités animées dont le sexe est culturellement significatif, genre grammatical et sexe biologique coïncident en général : p.ex. dālìbī 'étudiant' vs. dālìbā 'étudiante', jȁkī 'âne' vs. jȁkā 'ânesse', farī 'blanc' vs. farā 'blanche'.

Pour le reste, le genre du mot est arbitraire. Il existe une règle générale selon laquelle les mots se terminant par « a » sont féminins, tandis que les autres sont masculins. Cette règle comporte toutefois des exceptions. L'opposition de genre est neutralisée au pluriel, au profit du masculin pour l'accord.

On ne doit pas s'attendre à ce que le genre arbitraire soit le même en français et hausa :

P. ex. gīwā 'éléphant (e)' est féminin.

2.3.2. Le nombre

Le singulier (SG) (ex. jàkī 'âne') fait référence ou bien à un individu mâle de l'espèce « âne », ou bien à l'ensemble des individus de l'espèce. Le pluriel jàkunā signifie 'ânes' ou 'ânesses'.

La formation du pluriel (PL) est l'un des aspects les plus complexes de la grammaire du hausa. On distingue dix classes mettant en jeu divers procédés morphologiques : suffixation, alternance tonale, réduplication, etc. L'appartenance d'un nom à une classe est à peu près imprévisible. Deux exemples : tēbùr 'table' vs. tēbùrōrī 'tables' ; zōb 'anneau' vs. zôbbā 'anneaux'.

Le français paraît plus simple. Sa difficulté pourrait cependant venir du fait qu'à l'oral, le pluriel ne se marque pas sur les noms (sauf exception), mais sur les déterminants, tandis que la langue écrite exige des « ou » qui ne se prononcent pas.

2.3.3. Les déterminants

Il n'existe pas en hausa d'équivalent exact des articles défini et indéfini du français. Les suffixes anaphoriques (ANAPH) -n (M.SG ou M/F.PL) et -r (F.SG) indiquent que le référent du nom est censé appartenir à l'information partagée par les interlocuteurs : Gȁ kuj râr {voici chaise-ANAPH.F} 'Voici la chaise (que tu sais)'.

L'indéfini spécifique wani (M) / wata (F) / waɗansu ~ wasu (M/F.PL) précède le nom et ses modifieurs : wani yār 'un (certain) garçon (connu)'. Le nom nu exprime l'indéfini non spécifique : Gȁ kuj rā {voici chaise} 'Voici une chaise (quelconque)' Mais la même phrase se laisse aussi traduire par 'Voici la chaise', car le hausa, au contraire du français, n'oblige pas ses locuteurs à être explicites quant au caractère connu ou non de l'information.

Les démonstratifs se divisent en deux séries, proxime (PROX) et distale (DIST) selon la distance (réelle ou notionnelle) par rapport au locuteur : (a) wannàn (PROX.M/F.SG.),

wadânnân (PROX. M/F.PL) ;(b) wancàn (DIST.M.SG), waccàn (DIST.F.SG), wadàncân (DIST.M/F.PL).

A la forme longue (initiale /wan/), ils précèdent le nom : waccàn kàbēwā̀ ‘cette citrouille-là’ ; allégés de l’initiale /wan/ ils suivent le nom, alors pourvu du suffixe anaphorique -n/-r : àbincin nàn ‘cette nourriture-ci’, gōnar càn ‘cette ferme là’. Sauf insistance particulière, ces formes sont les plus courantes.

2.3.4. Les possessifs

Ils sont suffixés au nom pourvu du suffixe anaphorique (sauf à la 1re personne) et s’accordent en genre avec le Possesseur au singulier : gōnāna ‘ma ferme’ (à moi=M) / gōnāta ‘ma ferme’ (à moi=F), gōnarkà {gōna-r-kà} 'ta ferme' (à toi=M) / gōnarkì {gōna-r-kì} ‘ta ferme’ (à toi=F), gōnarsà ‘sa ferme (à lui)’, gōnartà {gōna-r-tà} ‘sa ferme (à elle)’, contrastant sur ce point avec les déterminants possessifs du français qui s'accordent en genre avec le Possessum (ton livre/ta ferme). Au pluriel, il n’y a pas d'accord en genre au pluriel : gōnarmù ‘notre ferme’, gōnarkù ‘votre ferme’, gōnarsù ‘leur ferme’.

2.3.5. Les pronoms personnels

	1 singulier	2 masculin singulier	2 féminin singulier	3 masculin singulier	3 feminin singulier	1 Pluriel	2 pluriel	3pluriel
Fort	nï	kai	kë	shï	ita	mü	kü	sü
OD	ni	ka	ki	shi	ta	mu	ku	su
OI	mini	makà	miki	shï	matà	manà	mukù	musù

OD : objet direct

OI : objet indirect

Figure 1: Liste des pronoms personnels

2.3.6. La construction génitive

Il s'agit des constructions du type de 'la nourriture du chat'. Leurs structures sont semblables en hausa et en français : {Xpm connecteur Ypr} (pm = Possessum, pr = Possesseur). En français, le connecteur (CONN) de s'appelle une préposition ; en hausa, on le nomme plutôt

« particule », na si le Possessum est masculin ou pluriel, ta s'il est féminin : àbinci na kyânwā {nourritureM CONNM chatF} 'la nourriture du chat' ou 'nourriture de chat', sānìyā ta Audù {vacheF CONNF AuduF} 'la vache d'Audu', shānū na Audù 'les bovins d'Audu'. En langue ordinaire, na se réduit à n et ta à r et ils se suffixent au Possessum : àbinci-n kyânwā, sānìya-r Audù, shānu-n Audù. Les formes pleines du connecteur servent aussi d'équivalents de 'celui/celle de' : ta Audù 'celle d'Audu (parlant de sa vache)'.

2.3.7. Les constructions relatives

La proposition relative suit son antécédent. Elle est introduite par le relateur (REL) invariable dà ou par le pronom relatif wandà (M), waddà (F), waɗàndà (PL) accordé en genre et nombre avec l'antécédent, mais ne variant pas selon la fonction de celui-ci (au contraire du français : cf. qui sujet vs. que objet). Le verbe est à l'aspect dit « relatif » (REL) :

2.4. Le verbe et le complexe verbal

Exposer en quatre pages le système verbal du hausa dans toute sa complexité est mission impossible. Nous nous contenterons d'en dégager les traits les plus caractéristiques.

La plus grande partie des verbes hausa sont dissyllabiques ou trisyllabiques, avec une nette prépondérance des premiers. 19 verbes, d'un emploi très fréquent, ont au moins une forme monosyllabique (ψι, "faire" ; χι, "manger" ; σηαα "boire", etc.). Les verbes dérivés peuvent compter jusqu'à 7 syllabes. Les verbes hausas se regroupent en 7 classes morpho-phonologiques traditionnellement appelées degrés (d°1 à d°7). Les degrés sont définis par leur suff. et leur schème tonal, par ailleurs soumis à des variations contextuelles.

Les trois premiers degrés sont dits primaires, et les autres dérivés, les monosyllabes formeront le d°0.

d°0	**ci ; sha**	**manger ; boire**		
d°1	**kaamàa**	**attraper**		
d°2	**sàyaa**	**acheter**		
d°3	**fita**	**sortir**		
d°4	**sayèe**	**tout acheter**	**-ee**	**achevement**
d°5	**sayar**	**vendre**	**-ar**	**causatif**
d°6	**sayoo**	**aller acheter et revenir**	**-oo**	**directionnel**
d°7	**sàayu**	**être (finalement) vendu/acheté**	**-u**	**passif**

Le tableau des verbes

2.4.1. Les formes

Un verbe hausa peut se manifester sous l'une ou plusieurs de sept « formes » (grades en anglais) numérotées et caractérisées chacune par un schéma tonal et une terminaison particulière exprimant diverses modulations du sens de base (incarné par les formes I-III) selon que le verbe est transitif ou intransitif. Ainsi, tārà̀ (I, H•B, -ā) 'rassembler' a pour forme IV tār (H•B, -ē) 'tout rassembler', soit 'achèvement complet', signification spécifique de IV ; sàuka (III, B•H, -a) 'descendre' a pour forme V saukar (H•H, -ar) 'faire descendre, abaisser' ; etc. C'est là une propriété typiquement afro-asiatique (cf. l'arabe).

2.4.2. TPN (Temps-Personne-Nombre)

Le verbe hausa se fléchit principalement pour l'aspect : accompli (ACP) vs. inaccompli (INACP). Le verbe lui-même n'est pas modifié, mais il est précédé d'un constituant (TPN) qui cumule les valeurs d'aspect et de personne-nombre. A titre d'illustration, voici les paradigmes de l'accompli de zō 'venir' et de l'inaccompli de aikī̀ 'travailler'.

Par défaut, le premier fait référence à un événement passé ('je suis venu', etc.), le second à un événement en cours ('je suis en train de travailler', etc.) :

	1	2M	2F	3M	3F	IND
Singulier	inà aikī̀	kanà aikī̀	kinà aikī̀	yanà aikī̀	tanà aikī̀	anà aikī̀
Pluriel	munà aikī	kunà aikī̀		sunà aikī̀		

Figure 2: accompli

	1	2M	2F	3M	3F	IND
Singulier	nā zō	kā zō	kin zō	yā zō	tā zō	an zō
Pluriel	an zō	kun zō		sun zṑ		

M : Masculin ; **F : Féminin** ; **IND : indéfini**

Figure 3: inaccompli

2.5. Caractéristiques du fonctionnement lexical de la langue

Le hausa est une des langues africaines les plus parlées avec le swahili. C'est la langue tchadique la plus importante en nombre de locuteurs et compte près de cent millions de locuteurs, principalement répartis en Afrique de l'Ouest entre le Nigeria, le Niger, le Bénin, le Togo, le Ghana, etc.

Elle est également parlée en Afrique centrale dans des pays comme le Cameroun, le Tchad, la République Centrafricaine, le Gabon, etc. Il s'agit d'une langue à tons et à différence de quantité vocalique phonologique.

L'opposition phonologique au niveau vocalique est observable dans des mots comme : karoo vs kaaroo ; kishii vs kiishii ; turaawaa vs tuuraawa ; etc.

Le système vocalique serait ainsi constitué des deux séries de voyelles suivantes :

- **Voyelles brèves : a, i, u, e, o.**
- **Voyelles longues aa, ii, uu, ee, oo.**

Les mots supportent des tons hauts (H) comme ráánáá ; sáú, bárcíí, táttálíí, wáájéé ; gírmáá, des tons bas (B) comme jìkíí, jìkáá, jííkàà, bìríí, bírìì, màríí, wààƙéé, wááƙàà, rèènóó, nóónòò, et un ton modulé tombant (MHB) comme dáà, cíì, sháà.

La structure syllabique est constituée des combinaisons suivantes : cv (cii), cvv (taa-ki) et cvc (ƙarhe). Au plan morphologique, la langue observe une flexion de genre et de nombre qu'il convient de cerner sur certains aspects qui suivent :

Sur le plan du genre on identifie deux types de formations : les formations à items simples comme rami, gida, garke, gero, kunu qui sont des noms masculins, et rana, garka, moɗa, cera, turka, zumuwa qui sont des noms féminins.

Les terminaisons des noms sont déterminées comme suit :

Noms masculins singuliers : i/ii, u/uu.

Noms féminins singuliers ee, oo, a/aa.

Il semble que tous les noms féminins à quelques exceptions près finissent par –a ; et les noms masculins à quelques exceptions près finissent par i, e, o, u. Ainsi, zomo, wake, tulu et rami seraient des mots simples connus comme masculins en hausa alors que rana, dara, garka, et mota seraient des formations simples connues comme féminines en hausa.

Les formations couples' avec deux sous-types : les couples à items simples et les couples à items dérivés. Ainsi, doki et goɗiya, rago et tumkiya seraient identifiés comme des couples

composés d'items simples pour chacun des deux genres, alors que malam et malama, sarki et sarauniya, icce et itaciya seraient des couples d'items dérivés. Les 'formations composées comme macce-da-goyo, ci-ma-zamne, ɗan allau qu'observent d'autres spécificités du point de vue flexionnel.

Il y a lieu de spécifier certains aspects structurels et organisationnels qui restent sensibles dans la couverture des questions de flexion et de dérivation en hausa. Il s'agit notamment de certains principes culturels et historiques entrant en ligne de compte dans la hiérarchisation des phénomènes.

Le tableau suivant montre la hiérarchisation des phénomènes dans la formation des items simples, des couples à items simples et des couples dérivés :

Types de mots	Genre		Nombre	
	Masculin	Féminin	Pluriel spécifique	Pluriel commun
Mots à items simple	Zomo(lapin) Wake(haricot) Iko(pouvoir) Tulu(jarre) Rami(trou)	- - - - -	Zomaye Wake Ikuna Tuluna ramu	- - - - -
	- - - - -	Rana Garka Mota Kwalwa saiwa	Ranaye Garake Motoci Kwalwa saiwoyi	- - - - -
Mots couples à items simple	doki	godiya	dawaki godiyoyi	Dawakkai -
	Rago	tumkiya	Raguna tumaki	Tumakkai -
	bunsuru	akuya	Bunsura awaki	- awakkai
Mots couples à items dérivés	Rakumi Malami Icce marayi	Rakuma Malama Itaciya marayiya	- - - -	Rakumma Malamai Itace maraya

Figure 4: Hiérarchisation des phénomènes dans la formation des items

2.6. Formations lexicales et règles de formation de mots en hausa

Des aspects fondamentaux ont été dégagés dans les théories décrites. Cependant, plusieurs aspects spécifiques à la langue hausa pourraient être ignorés dans le cadre de ces méthodes et des modèles qu'elles ont engendrés.

Pour illustrer notre démarche, seuls quelques exemples ont été sélectionnés. Plusieurs auteurs ont déjà travaillé dans ce domaine comme (Roxana, 1990). Suivant notre hypothèse de travail, un modèle de morphologie lexicale traiterait les principaux niveaux qui sont le dictionnaire et les règles.

Le dictionnaire constitue la somme de tous les mots actuels en usage dans la langue et dont l'étude porterait sur la structuration de ses mots dans leurs « sous composants ». Les règles, quant à elles, entrent en ligne de compte dans cette structuration.

Ainsi, nous partageons cette assertion de (Scalise, 1984) « qu'aussi loin que le lexique puisse être concerné, on pourra suggérer que les unités du dictionnaire sont les 'mots' et les 'thèmes', et qu'aussi loin que les règles lexicales puissent être concernées, nous donnerons une représentation des règles de préfixation, des règles de suffixation, et des règles de composition, montrant comment ces règles utilisent l'information associée à un item lexical ».

Nous chercherons ainsi à trouver les conditions de bonne formation de ces trois catégories de règle dans la langue.

Des différents points de vue ci-dessus examinés, il est ressorti que la composante lexicale de la grammaire est régie par un groupe de règles, les règles de formation de mots (RFM). Il s'agit des règles de flexion (RFs), des règles de dérivation (RDs) et des règles de composition (RCs). Dans les faits, il s'agit de règles d'adjonction et que, de ce point de vue il serait difficile de faire la différence opérationnelle entre :

Dans les faits, il s'agit de règles d'adjonction et que de ce point de vue il serait difficile de faire la différence opérationnelle entre :

- [malam+i] nm / [malam+a]nf / [malam+ai]np qui sont des opérations de flexion de genre pour les deux premiers cas et de nombre pour le troisième (malami/-a/-ai = professeur /e/s) et
- [malam+tarda] v qui est une opération de dérivation ou de formation verbale traduisant le processus de formation de professeur.

Les deux se résument donc au même type d'opération X + Y où X est un mot ou une racine ou un thème selon la langue et Y un axe formateur de mot dans son acception fléchie (+i, +a, +ai), ou un opérateur d'un transfert catégoriel comme c'est le cas de '+ tarda'. Mais les linguistes restent partagés sur la question : ceux qui maintiennent que la dérivation et la flexion sont essentiellement le même type de processus comme (Halle, 1973) ou (Jackendof, 1975) et ceux qui pensent qu'il s'agit de processus différents comme (Scalise, 1984).

Notre attention va porter au second groupe pour qui les règles de flexion (RFs) sont différentes de celles de dérivation (RDs) et qu'elles s'effectuent à l'intérieur de la même catégorie. Ce groupe suppose que la flexion s'opère entièrement à l'intérieur de la composante lexicale qui se donne comme finalité de définir « le mot possible » et qu'il s'agit de règles de nature différente. Les démonstrations de Scalise sur l'italien s'appliquent au hausa comme le montrent les règles suivantes :

- Les règles de flexion (RFs) ne changent jamais de catégorie syntaxique à un mot, ce que font les règles de dérivation (RDs) malam + i > malam + a > malam + ai (la catégorie ne change pas) Sg.msc sg. fém pl com. malam + tarda (la catégorie change)

Cependant, il y a des cas où la RD ne change pas de catégorie : shedɗani > shedɗanci

- La fexion est toujours périphérique au regard de la dérivation. L'ordre des règles est Mot>Dérivation>Flexion et jamais Mot>Flexion>dérivation.

De ce fait, un axe dérivationnel ne peut pas être attaché à un mot féchi. Pour mieux saisir la portée de l'hypothèse d'Aronof en hausa, nous avons examiné quelques exemples de mots contenus dans la phrase suivante : "Wani babban mutun ya tambaye ka."

On peut partir de l'hypothèse que cette phrase est construite sur la base de deux axes :

- L'axe syntagmatique ou celui de la succession des éléments ou mots qui la constituent [{Wani} + {babban} + {mutun} + {ya} + {tambaye} + {ka}],
- Paradigmatique qui a permis une série de choix sur lesquels sont portés chacun des mots appartenant individuellement à un paradigme donné, par exemple le paradigme {Wani, wata, wasu}

Prenons par exemple 'la pile' 1 du paradigme 1 {wani, wata, wasu} : pour le locuteur natif, cette 'pile' est clôturée aux trois adjectifs indéfinis {''wani'' (un certain), ''wata'' (une certaine) et ''wasu'' (des / certains). wasu admet des variantes ''waɗansu'' et ''wa'ansu''}. On peut dire que cette catégorie est fermée, même si par ailleurs les éléments de sa 'pile' peuvent être substitués par les éléments d'une autre pile : ex. {wannan} babban mutumen ya tambaye ka.

L'adjectif indéfini 'wani' peut être substitué par l'adjectif démonstratif ''wannan'' (ce(tte) personne en question), ou par un adjectif possessif de la pile de ''nawa'' (le mien), ou par un adjectif qualificatif comme jibgege à l'acceptation de la pile de ce dernier, celles des démonstratifs et des possessifs sont fermées à l'instar de celle des indéfinis. Au stade de cette présentation, et dans la perspective du traitement particulier du hausa, on peut dire qu'une catégorie syntaxique est constituée de piles d'éléments ou paradigme : catégorie syntaxique des adjectifs {paradigme des indéfinis, paradigme des possessifs, paradigmes des qualificatifs, etc.

Catégorie syntaxique des noms {paradigme des noms propres, paradigme des noms communs subdivisés selon certains paramètres culturels, etc.},

Catégorie des adverbes {paradigme des adverbes de lieu, paradigme des adverbes de temps, paradigmes des statifs, des profusatifs, etc}. On peut également retrouver quelques mots en appliquant les règles comme :

A. Tsunts +u (oiseau masculin) >+uwa (oiseau féminin) >+aye (oiseaux pluriel)
B. Ran +a (soleil féminin) >+aye (soleils pluriel)
C. Buz +u (touareg masculin) >+uwa (touareg féminin) >+aye (touaregs pluriel)
D. Dar + e (nuit masculin) >+aye (nuits pluriel)

Au point de vue de la tonologie, le constat fait est le suivant :

A. 1.tsúntsúú HH	2. tsúntsúwáá HHH	3. tsúntsààyéé HBH
B. 1. Búúzúú HH	2. búúzúwáá HHH	3.búúzààyéé HBH
C. 1. Ø	2. ráánáá HH	3. ráánààyéé HBH
D. 1. dáréé HH	2. Ø	3. dárààyéé HBH

Il ressort qu'en hausa, la flexion est une adjonction d'un thème de flexion à une racine.

Comme dans le cas des dissyllabiques, les noms trisyllabiques sont régis par une diversité de règles allant d'un type commun ou fondamental vers des types spécifiques à des partitions données. Pour la base des noms trisyllabiques, nous avons identifié deux groupes à savoir les non dérivés (type A) et les dérivés (type B).

- HH pour le singulier masculin ;

- HHH pour le féminin singulier ;

- HBH pour le pluriel commun.

Les résultats des travaux sur ces groupes sont résumés dans les figures 5 et 6 :

Singulier	Pluriel	Type	Remarques / Schéma
Kasuwa	kasuwoyi	Simple	-a] BHH > -oCi]HHHH
dâkwara	dâkwarori	Simple	-a] HHH > -oCi]HHHH
taguwa	taguwoyi	Simple	-a] HHB > -oCi]HHHH
korama	koramomi / koramu	Trissyllabique A	-v] BHB > -oCi]HHHH / -u]BBH
dorowa	dorowoyi / doroyu / doroyi	Trissyllabique A	-v] BHB > -oCi]HHHH / -u/-i]BBH
godiya	godiyoyi / godiyu / godiyai	Trissyllabique A	-v] BHB > -oCi]HHHH / -u/-i/-ai]BBH
mummuƙe	mukamukai	Trissyllabique B	
gunguni	gunaguni	Trissyllabique B	

Figure 5: Noms trisyllabiques du type A

Partition	Masc. Singulier	Fém. Singulier	Pluriel commun	Règles / Remarques
1. ethnonymes	bahaushe	Bahausa /-(shi)ya	Hausawa	[ba-rac. -e]∈BH(H)B > [ba-pref-Rac.-a/iy∂a]∈BH(H)B > [σpref-Rac-awa]∈HH(H)H
	Bature	Baturiya	Turawa	même règle que ci-dessus
2.a. agents	Ma'aikaci	Ma'aikaciya	Ma'aikata	[ma-pref-Rac. -i] ∈HB(B)H, + {-i|HB(B)H, -iya|HH(H)BH, -a|HB(B)H}
2.b. instruments	Ma'aikaci		Ma'aikatayya	[ma-pref-Rac + {-i] ∈HH(H), -ayya] ∈BB(B)HB}
2.c. lieux		Ma'aikata	Ma'aikatu	[ma-pref-Rac. -a]∈HH(H)H, [ma-pref-Rac.-u]∈BB(B)H

Figure 6: Noms trisyllabiques du type B

2.7. Etat de l’art touchant au traitement du hausa

Jusqu’en janvier 2022, l'état de l'art concernant le traitement du hausa est encore en développement. Bien qu'il existe une reconnaissance de son importance linguistique, les travaux de recherche spécifiques sur le traitement automatique du hausa restent relativement limités par rapport à d'autres langues plus largement étudiées comme l’arabe.

- Reconnaissance de la parole et synthèse vocale : Il existe des efforts pour développer des systèmes de reconnaissance automatique de la parole et de synthèse vocale en hausa afin de faciliter l'interaction avec les technologies de l'information et de la communication pour les locuteurs de cette langue.
- Traduction automatique : La traduction automatique entre le hausa et d'autres langues est un domaine de recherche en croissance. Des travaux sont en cours pour développer des systèmes de traduction automatique capables de traiter le hausa, en particulier dans le contexte de la traduction entre le hausa et l'anglais.
- Analyse de sentiments et traitement du langage naturel : L'analyse de sentiments dans les textes hausas ainsi que le traitement du langage naturel pour diverses applications telles que la recherche d'information et l'analyse de documents sont des domaines émergents de recherche.
- Technologies éducatives : Les technologies éducatives prennent également en compte le hausa, avec des applications visant à faciliter l'apprentissage de la langue à travers des outils numériques et des ressources éducatives en ligne.
- Ressources linguistiques et corpus annotés : La création de ressources linguistiques telles que des lexiques, des corpus annotés et des outils d'annotation manuelle est essentielle pour le développement des technologies de traitement automatique du hausa.

Cependant, malgré ces développements, il existe toujours des défis à relever dans le traitement automatique du hausa, notamment le manque de ressources linguistiques annotées, la complexité de la structure linguistique et la diversité des dialectes. L'expansion de la recherche dans ce domaine est donc nécessaire pour répondre aux besoins des locuteurs du hausa dans un monde de plus en plus numérique. Ce qui va nous amener à parler du texte mining dans le prochain chapitre.

Chapitre 3 : Text Mining

L'exploration de texte a attiré une attention croissante ces dernières années, en raison des grandes quantités de données textuelles, qui sont créées dans une variété de réseaux sociaux, du Web et d'autres applications centrées sur l'information. Les données non structurées sont la forme de données la plus simple pouvant être créée dans n'importe quel scénario d'application. En conséquence, il y a eu un énorme besoin de concevoir des méthodes et des algorithmes qui peuvent traiter efficacement une grande variété d'applications de texte.

3.1 Le Data Mining

Le data Mining ou fouilles de données, est un processus d'extraction des informations utiles à partir des grandes bases de données en utilisant des techniques de statistique, de Machine Learning et de visualisations de données, pour identifier dans ces données des relations, des modèles, des tendances cachées, ces informations après peuvent être utilisées pour prendre des décisions basées sur les données.

3.2 Le Text mining

Le Text mining ou fouilles de textes ou l'exploration de données, est une sous-discipline de Data mining qui traite spécifiquement les données textuelles non structuré comme les emails, les fichiers Word, the news headline et tous les fichiers textuels, et après ça l'extraction des connaissances utiles à l'aide de spécifique techniques (comme NLP (Techniques de Traitement de Langage Naturel) par exemple). Les connaissances extraites peuvent être utilisées pour le développement des algorithmes d'apprentissage capable d'analyser des données non structurées, classer les documents, analyser les sentiments, NER (Name Entity Recognition), l'exploration de données peut être appliqué dans divers domaines tel que le Marketing, réseaux sociaux, la Santé, etc.

3.3 Le processus de Text mining

Les étapes nécessaires pour effectuer le processus de Text mining sont :

- **La sélection ou collection des données** : reçue de données textuelles non structurées à partir de différentes sources, telles que des sites web, des plateformes de médias sociaux, des articles de presse, des avis de clients, etc.

- **Prétraitement de données** : Prétraiter les données textuelles, nettoyer et normaliser les textes par la tokenisation et la lemmatisation et enlevant des informations irréalisées, telles que des mots vides, des caractères spéciaux.

- **Représentation du texte** : Convertissez les données textuelles après le prétraitement en un format numérique, tel qu'un sac de mots (bag of words) ou une matrice de term-document, pour les préparer à l'analyse.

- **Exploration du texte** : Explorez les données textuelles pour identifier des motifs, des relations et des insights à l'aide de spécifiquement techniques, telles que l'analyse de la fréquence des mots (les nuages de mots) et l'analyse de sentiment.

- **Analyse et visualisation du texte** : Appliquer des algorithmes d'apprentissage automatique tel que la classification, et visualisez les résultats pour les rendre plus faciles à comprendre et à interpréter.

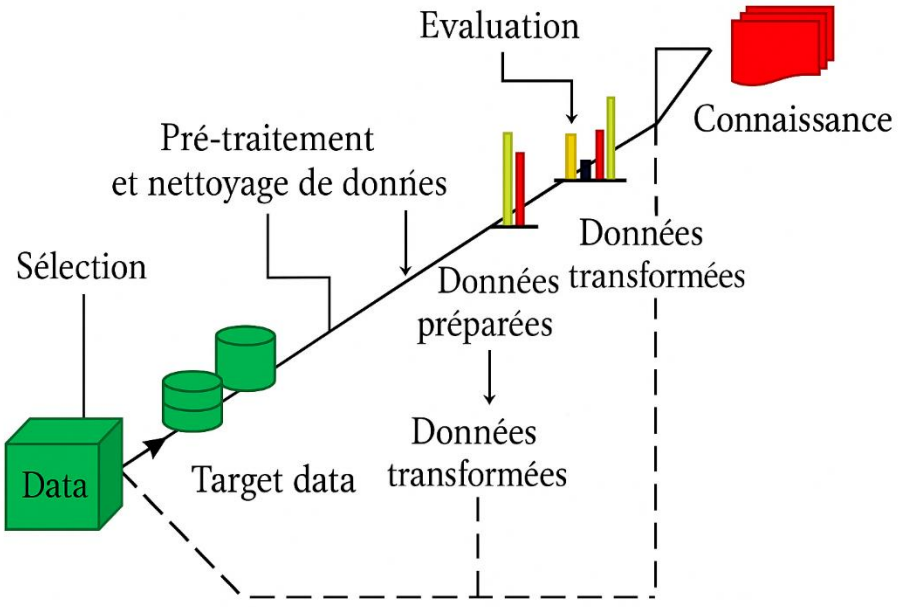

Figure 7: Processus de fouille de données

3.4 Les Techniques de Text Mining :

Dans cette partie on va présenter les différentes techniques sur le contenu des textes et vise à extraire et structurer les connaissances, et, parmi ces techniques, on trouve :

3.4.1 Techniques de Traitement de Langage Naturel :

Un traitement automatique est :

- Une suite d'actions ou calculs à faire par la machine. Le Traitement Automatique des Langues a pour objectif de traiter des données linguistiques (textes) exprimées dans une langue dite "naturelle".
- La conception de programmes capables de traiter automatiquement des données linguistiques de type : textes écrits ; dialogues écrits ou oraux ; unités linguistiques (mots, phrases, énoncés ...)

Les tâches impliquées dans cette technique peuvent inclure la tokenization, élimination des mots vides et filtrage de textes, Lemmatisation, la racinisation (ou troncature) :

- **La tokenization**

Est le processus de décomposition d'un texte en plusieurs pièces appelé tokens (jetons).

Exemple:

"Full sanctions on Russian exports would be a pivotal moment for the oil market, potentially touching off a sustained high price cycle with few precedents."

Tokenization

Tokens:| Full | sanctions | on | Russian | exports | would | be | a | pivotal | moment | for | the| oil |market | potentially | touching | off | a |sustained |high| price | cycle | with| few |precedents |

- **Le filtrage**

Le filtrage est une technique utilisée pour prétraiter les données textuelles en vue de les analyser et de les manipuler plus facilement. Il peut inclure différentes techniques, telles que :

- Convertir en minuscules ;
- Retirer les mots vides tels que "le", "la", "et", etc. ;
- Retirer les signes de ponctuation et les caractères spéciaux.

- **Lemmatisation :**

La lemmatisation est une technique qui réduit les mots à leur forme de base, ou lemme, par exemple, pour un verbe, ce sera son infinitif. Pour un nom, son masculin singulier. L'idée étant encore une fois de ne conserver que le sens des mots utilisés dans le corpus

Exemples :

Cherche, cherchent -> chercher

ai, as, a, -> avoir

Belles, beaux, beau -> beau

3.4.2 L'extraction d'information ("IE")

C'est la technique la plus utilisée parmi les techniques de Text Mining, qui se concentre sur l'extraction automatique d'informations structurée à partir de données textuelles non structurées. L'objectif d'IE est de convertir un texte non structuré en un format structuré qui peuvent être facilement analysé et traité par l'ordinateur.

Cela implique l'identification et l'extraction d'informations spécifiques, tel que les entités nommées (noms de personnes, noms d'organisations, emplacement, etc.), les relations entre les entités, les informations extraites sont ensuite organisées dans un format structuré, telles qu'une base de données ou un tableau, pour une analyse et un traitement ultérieur.

L'extraction d'informations est largement utilisée dans diverses applications, telles que la recherche d'informations, l'analyse des sentiments et la réponse aux questions.

3.4.3 La recherche d'informations (information retrieval "IR")

L'IR est un processus d'extraction qui fournit des informations pertinentes à l'utilisateur en fonction de sa requête et de classer les résultats en fonction de leur pertinence. Cette technique utilise différents algorithmes pour suivre et surveiller les comportements des utilisateurs et découvrir et évaluer la pertinence de chaque document et les classer sur sa pertinence. Parmi les systèmes IR les plus connus sont les moteurs de recherches Google et Yahoo.

3.4.4 La catégorisation (classification supervisé)

Dans l'apprentissage supervisé, on fournit des données d'entrée qui sont étiquetées avec les sorties souhaitées. Le but est que l'algorithme puisse "apprendre" en comparant sa sortie réelle avec les sorties enseignées pour trouver des erreurs et adapter le modèle. Ce processus permet à l'algorithme de prédire les valeurs d'étiquettes pour des données non étiquetées.

Les algorithmes d'apprentissage supervisé peuvent être utilisés pour des tâches telles que la classification de texte, l'analyse de sentiments et pour faire des prédictions.

Parmi ses méthodes :

- Les arbres de décision, les réseaux neurones, la méthode des k plus proche voisins (KNN) ou la classification bayésienne.

3.4.5 Clustering (classification non supervisée)

La classification non supervisée regroupe des points de données similaires en fonction de leurs caractéristiques. Le but du clustering est de diviser les données en groupes ou clusters homogènes, de sorte que les points de données d'un cluster soient plus similaires les uns aux autres qu'ils ne le sont avec les points de données d'autres clusters.

Les algorithmes de clustering utilisent des métriques de similarité pour déterminer les relations entre les points de données et les regrouper en clusters. Il est utilisé dans diverses applications, telles que la segmentation d'images, les études de marché et autres. Parmi les méthodes que nous utilisons dans le clustering :

- **Les règles d’association**

Les règles d'association sont des techniques d'apprentissage non supervisé qui identifient les relations entre les éléments ou les variables dans un ensemble de données. Elles révèlent des motifs, ce qui aide à prendre des décisions éclairées et à formuler des recommandations.

- **K-means**

Tout simplement, le K-means est une technique de segmentation dont l'objectif est de partitionner des données textuelles en utilisant la mesure de la distance ou de la similarité entre les observations en K clusters.

En définitive, l'exploration de texte et l'exploration de données sont des techniques puissantes pour extraire des connaissances à partir de grandes quantités de données. La classification non supervisée est une technique d'exploration de données importante qui peut être utilisée pour résoudre une variété de problèmes réels dans certains cas. L'utilisation de cette technique est devenue de plus en plus importante dans de nombreuses industries et leur potentiel pour des applications supplémentaires est vaste.

Dans le prochain chapitre, on présentera un état de l'art sur les règles d'association, ainsi que les techniques d'extraction, les techniques d'élagage, les différents algorithmes et leurs variantes.

Chapitre 4 : Les règles d’associations

Etat de l’art sur les règles d’associations

L'état de l'art des règles d'association repose sur l'apprentissage non supervisé pour découvrir des relations entre des éléments dans de grands ensembles de données. Les avancées incluent des algorithmes plus efficaces comme FP-Growth et Eclat, ainsi que l'application de mesures de filtrage plus poussées comme le Lift et la Conviction pour améliorer la qualité des règles générées. Ces techniques sont largement utilisées dans des domaines comme l'analyse de paniers d'achat pour éclairer les décisions de marketing, de vente et d'organisation des stocks.

4.1. Définition

Une règle d'association est une relation d'occurrence qui lie deux ou plusieurs items particuliers. Ceci s'illustre plus facilement à l'aide d'un exemple. Si le contenu des paniers d'achats révèle qu'un grand nombre de clients qui achètent du pain achètent également du beurre, cela signifie qu'il existe une relation d'occurrence simultanée forte entre ces deux items.

Cette relation est appelée association et de celle-ci s'induit la règle d'association suivante :

Pain => beurre. Cette règle d'association est dite classique ou traditionnelle, puisque les deux événements, l'achat du pain et l'achat du beurre se déroulent simultanément lors de la même transaction.

Lorsque l'occurrence des items est étalée selon un ordre spécifique, par exemple, les clients qui achètent du pain, achète également du beurre dans les jours qui suivent, les règles d'association entre ces items ordonnés sont qualifiées d'associations séquentielles. L'objectif des règles d'association, traditionnelles ou séquentielles, consiste à découvrir les associations dans les données étudiées.

4.2. Utilité des règles d'associations

Les règles d'association sont appliquées dans plusieurs domaines. En marketing, par exemple, elles permettent d'identifier les produits ou services qui sont achetés lors d'une même transaction ou par un même client dans le temps et offrent donc la possibilité d'identifier des opportunités de ventes croisées.

En analysant l'ordre dans lequel les internautes accèdent aux pages d'un site WEB, les règles d'association séquentielles permettent d'entrevoir quelles modifications rendraient le site plus convivial, permettant ainsi, aux internautes de trouver rapidement les informations recherchées. Les règles d'association séquentielles sont également utiles dans l'étude des comportements d'achats des consommateurs, car elles permettent de mettre en lumière les comportements d'achats à travers le temps.

Par exemple, si les associations séquentielles indiquent que les propriétaires d'une marque et d'un modèle précis de voiture changent leur véhicule aux 18 mois, le concessionnaire pourra fidéliser sa clientèle en faisant parvenir un dépliant sur les nouveaux modèles disponibles quelques semaines avant la date du changement de voiture.

4.2.1. Motif

Soit ‘’T’’ et ‘’D’’ deux ensembles et R une matrice. ’T’est un ensemble de termes et ’D’’ est un ensemble de textes tel que **:**

T= {a, b, c, d, e} et D= {d1, d2, d3, d4, d5, d6}

La matrice R représente la relation binaire qui existe entre l’ensemble T et D.

R	A	B	C	D	E
D1	1	0	1	0	1
D2	0	1	1	1	1
D3	1	1	1	0	1
D4	1	0	0	0	0

D5	0	1	1	1	1
D6	1	0	0	1	0

Tableau 1: Matrice termes-documents

On appelle un motif tous les sous-ensembles de T. Un motif ‘’t’’ est inclus dans un texte ‘di’ si $\forall t \in$ T, $(t', di) = 1$. Un motif de taille K est appelé k-motif.

Par exemple : d3 et d5 contiennent le 4-motif.

4.2.2. Motif fréquent

On dit qu'un motif est fréquenté si un support d'un motif "t" supérieur à un support minimal (qui est déterminé par l'utilisateur).

$$\mathbf{support(t) \geq min\ \_sup}$$

Tel que min_sup est le support minimal

4.2.3. Support d’une règle d'association

Le support d'une règle d’association (A C) est une mesure de la fréquence d'apparition de la Règle, il représente le pourcentage de documents qui contiennent A et C "support (A U C)" divisé par le nombre total de document :

$$\boldsymbol{Support(A => C) = \frac{support(A \cup C)}{D}}$$

Formule 1 : Formule support d'une règle d'association

D : le nombre total de documents

Le support est un indicateur de fiabilité de la règle

4.2.4. Confiance d'une règle d'association

La confiance est une mesure exprimée à l'aide de la probabilité conditionnelle d'avoir l'événement C sachant que l'événement A s'est produit, c'est une mesure descriptive qui prend ses valeurs dans l'intervalle [0, 1].

Il est défini par la formule :

$$\mathbf{Conf}(\mathbf{a} => \mathbf{c}) = \frac{\mathbf{support}(\mathbf{A} \cup \mathbf{C})}{\mathbf{support}(\mathbf{A})}$$

Formule 2 : Formule confiance règle association

4.2.5. Le Lift

Le Lift est une mesure statistique, symétrique, représente le rapport d'indépendance entre l'antécédent et le résultat de la règle. Il prend ses valeurs dans l'intervalle [0, +∞ [mais, en pratique, il est rare que le Lift dépasse 10 ou 20.

Elle est définie par la formule :

$$\mathbf{Lift}(\mathbf{X} => \mathbf{Y}) = (\mathbf{supp}(\mathbf{X}, \mathbf{Y}) / \mathbf{Nsupp}(\mathbf{X})/\mathbf{N} * \mathbf{supp}(\mathbf{Y})/\mathbf{N})$$

Formule 3 : Formule du Lift

4.2.6. Le leverage

Le "leverage" est une mesure couramment utilisée dans l'exploration de règles d'association pour évaluer le degré de corrélation entre l'occurrence simultanée d'un antécédent et d'un

conséquent dans la même transaction par rapport à ce qui serait attendu s'ils étaient indépendants. Une valeur de zéro pour le "leverage" indique une absence de corrélation, tandis qu'une valeur supérieure à zéro indique une corrélation positive.

Plus la valeur du "leverage" est élevée, plus la corrélation est forte.

La formule pour calculer la liaison (leverage) entre deux éléments X et Y est :

$$\mathbf{Leverage(X => Y) = support(X => Y) - support(X) * support(Y)}$$

Formule 4 : Formule leverage

4.3. Extraction des règles d'association

La plupart des algorithmes de recherche de règles d'association (parmi eux : apriori) adoptent une stratégie qui consiste à décomposer le problème en deux étapes :

- Extraction des ensembles d'items fréquents ;
- Génération des règles d'association.

Dont l'objectif est d'extraire toutes les règles de grande confiance à partir des ensembles d'items fréquents trouvés dans l'étape précédente.

4.4. Exemples d'Algorithmes de Génération de Règles d'association

Il existe plusieurs algorithmes de génération de règles d'association. Ils utilisent les notions de support et de confidence pour déterminer la pertinence des règles d'association. Parmi eux on cite :

4.4.1. Algorithme Apriori

Apriori est un algorithme classique de recherche de règles d'association introduit par Agrawal et Srikant en 1993.

C'est le premier algorithme destiné à la recherche de règles d'association. Apriori génère les motifs fréquents, puis les relie entre eux pour générer les règles d'association.

Il se base essentiellement sur la propriété d'anti-monotonicité existante entre les motifs. Elle est utilisée à chaque itération de l'algorithme Apriori afin de minimiser au maximum le nombre de motifs candidats à tester.

4.4.1.1. Le principe de l'algorithme Apriori

La description de l'algorithme Apriori se résume dans les étapes suivantes :

- **Trouver les 1-Itemsets :** Parcourir la base de données pour identifier les éléments individuels et collecter les ensembles d'items ayant un support supérieur ou égal à min_sup.

-Générer les (k + 1) -Itemsets : Générer des candidats pour les (k + 1) -Itemsets en combinant des k-Itemsets fréquents en utilisant la propriété d'Apriori.

-Filtrer les candidats : Vérifier le support de chaque candidat (k + 1) -Itemset et conserver uniquement ceux qui satisfont le seuil min_sup.

- **Répéter les étapes 2 et 3 :** Itérer les étapes 2 et 3 jusqu'à ce qu'aucun nouveau k-Itemset fréquent ne puisse être trouvé.

L'algorithme Apriori explore de manière itérative les ensembles d'items de taille k en se basant sur les ensembles d'items fréquents de taille k-1 déjà trouvés. Il utilise la propriété d'Apriori, qui stipule que, si un ensemble d'articles est infréquenté, tous ses ensembles supérieurs (ensembles plus larges le contenant) ne seront également pas fréquentés.

Cette propriété permet à l'algorithme de générer et de filtrer efficacement les candidats d'ensembles d'items, réduisant l'espace de recherche et améliorant l'efficacité. En répétant le processus jusqu'à ce qu'aucun nouveau k-Items et fréquent ne puisse être trouvé, l'algorithme Apriori découvre les ensembles d'items fréquents dans la base de données.

4.4.1.2. Avantages et inconvénients

L'algorithme Apriori réduit considérablement la taille d'articles candidats de plus qu'il est facile à mettre en œuvre.

Cependant, il souffre des limitations par rapport à la nécessité de nombreuses analyses de base de données ainsi que le grand nombre d'ensembles éléments candidats qui peuvent être encore générés si le nombre total d'ensembles des éléments fréquents augmente.

4.4.2. L'algorithme FP-Growth

L'algorithme FP-Growth a été introduit par Han et Al. En 2000, ils ont dit qu'il est actuellement l'une des approches les plus rapides pour l'extraction fréquente d'ensembles d'articles.

C'est une méthode différente des approches par niveaux qui permet d'extraire des ensembles d'articles fréquents sans générer de candidats, évitant ainsi les parcours et les visites répétées de la base de données.

4.4.2.1. Le principe de l'algorithme FP-Growth

La description de l'algorithme FP-Growth se résume dans les deux étapes suivantes :

-Construire l'arbre FP : Parcourir la base de données pour construire un arbre FP, qui est une structure de données compacte représentant les itemsets fréquents et leurs relations. L'arbre FP se compose d'un nœud racine et de branches conditionnelles représentant les itemsets.

-Générer les itemsets fréquents : Parcourir l'arbre FP pour trouver les itemsets fréquents en exploitant de manière récursive les motifs conditionnels. Cela implique de trouver les items individuels fréquents, de construire des arbres FP conditionnels, et de répéter le processus jusqu'à ce qu'aucun itemset fréquent supplémentaire ne puisse être trouvé.

En résumé, l'algorithme FP-Growth construit un arbre FP à partir de la base de données et extrait de manière récursive les itemsets fréquents en exploitant la structure de l'arbre et les

motifs conditionnels. Il élimine le besoin de générer des candidats itemsets, ce qui le rend plus rapide que les autres algorithmes.

4.4.2.2. Avantages et inconvénients

L'algorithme FP-Growth résout le problème de la nécessité de nombreuses analyses de base de données, vu qu'il ne fait que deux balayages de la base des transactions.

Néanmoins, cela ne garantit pas, dans le cas où la base de transactions est trop volumineuse, que toute la structure de l'arbre FP sera maintenue en mémoire centrale.

De plus, la construction de la structure FP-tree peut prendre du temps et peut consommer beaucoup de ressources système.

4.5. Utilisation des règles d'association

L'intérêt des règles d'association dans la fouille de textes est multiple, on peut citer :

4.5.1. Analyse de concepts formels :

Les règles d'association permettent d'organiser des concepts dans une structure hiérarchique à partir de laquelle il est possible d'observer des corrélations entre les individus et leurs propriétés communes. En effet, on peut hiérarchiser les concepts en utilisant la correspondance de Galois pour créer un graphe de concept muni d'une relation d'ordre entre les concepts. On appelle ce graphe un treillis de Galois. La construction d'un treillis de Galois permet de se donner une structure mathématique pour l'analyse de concepts issus d'un domaine.

4.5.2. Extraction d'information (EI) :

L'extraction de règles d'association permet de réaliser des tâches d'extraction d'information pour remplir des patrons. À ce titre, le système apprend à remplir certains attributs de patrons pour de nouveaux textes à partir de règles d'association apprises sur d'autres patrons. Dans une notice bibliographique, par exemple, un patron possède un attribut auteur inconnu mais un attribut mots-clés complet {mc1, mc2, mc3}.

4.5.3. Veille technologique et stratégique

La veille stratégique (appelée également business intelligence). C'est un processus de mise à jour périodique d'informations. Il offre une aide précieuse à la prise de décision pour les gérants d'entreprise. Les règles d'association révèlent des implications entre termes et permettent de faire des suivis scientifiques.

4.5.4. Recherche d'information (RI)

La liste de documents pertinents qui constitue une réponse relative à une requête est fondée sur le lien de cooccurrence entre les termes de la requête et leur fréquence d'apparition ensemble dans les textes. L'utilisation des motifs fermés fréquents permet, par navigation dans le treillis de Galois correspondant, de répondre à une requête par les documents constituant l'extension d'un concept.

En définitive, les règles d'association sont un outil puissant dans la fouille de textes qui nous permettent de découvrir des relations précieuses entre les éléments d'un ensemble de données. En utilisant différentes mesures et algorithmes pour évaluer les règles d'association, nous pouvons obtenir des informations qui peuvent guider les processus de prise de décision.

Dans le chapitre 5, nous présentons la méthodologie utilisé pour la mise en œuvre de ce projet de recherche. Nous aborderons entre autres les outils, les technologies et les langages de programmation qui ont servi au développement du logiciel

Chapitre 5 : Méthodologie

Comme nous l'avons mentionné dans le chapitre précédent, la fouille de règles d'association est une technique utilisée pour découvrir des modèles et des relations entre des éléments ou des variables dans un ensemble de données. Cette technique a été largement utilisée dans divers domaines, tels que l'analyse de panier d'achat, l'exploration de l'utilisation du web et la bio-informatique.

Dans cette étude, nous nous intéressons à l'extraction de règles d'association de textes écrits en hausa à partir de données textuelles, telles que des articles de presse, des actualités, etc. Pour ce faire, nous comparons deux méthodes couramment utilisées pour l'extraction de règles d'association :

CountVectorizer avec l'algorithme Apriori, et TF-IDF avec deux algorithmes différents, Apriori et FP-Growth. Nous explorons spécifiquement comment chaque approche affecte la qualité et la quantité des règles d'association extraites, y compris des métriques telles que le support, la confiance et le lift.

Pour extraire des règles d'association à partir des données textuelles, deux méthodes ont été utilisées : Countvectorizer et TF-IDF. Le but est de comparer l'efficacité des deux méthodes pour extraire des règles d'association de haute qualité.

5.1. Prétraitement de données

Convertir tous les documents en minuscules, supprimer les mots vides, tokenizer les documents et les avoir lemmatisés en utilisant un tagueur de parties du discours (POS). Cette étape vise à standardiser les données textuelles et à éliminer toute information non pertinente qui pourrait affecter les performances de l'algorithme.

5.1.2. Création une matrice terme-document

On a utilisé python Countvectorizer pour convertir les documents prétraités en une matrice de fréquence de termes. Cette méthode a compté la fréquence de chaque terme dans chaque document, créant ainsi une représentation vectorielle du document. Cette matrice est appelée matrice terme-document, qui représente la fréquence de chaque terme dans chaque document, dans laquelle chaque ligne représente un document et chaque colonne représente un terme unique trouvé dans la collection. La forme de ma matrice qui nous intéresse exprime l'apparition des termes avec une méthode binaire.

Par exemple : A est une matrice term document et α une cellule de cette matrice

α ij= {1, $si\ le\ term\ j\ est\ present\ dans\ le\ decument\ i$; 0, $sinon$}

A	Term1	Term2	Term3
D1	1	0	0
D2	0	1	0
D3	1	0	1

Tableau 2 : Matrice term document

5.1.3. Extraction des motifs fréquents

Avant d'extraire les motifs fréquents, il est nécessaire de convertir la matrice en un format Pandas Dataframe afin de la passer à l'algorithme Apriori. Après cette conversion, nous aurons une matrice qui contiendra et représentera toutes les données de notre collection.

En utilisant l'algorithme Apriori que nous avons déjà expliqué dans le chapitre précédent, nous recherchons les ensembles d'items qui apparaissent fréquemment ensemble dans les documents analysés. Le support est utilisé comme mesure pour déterminer quels itemsets sont considérés comme fréquents. Une fois que les motifs fréquents sont identifiés, ils servent de base pour la génération des règles d'association dans l'étape suivante.

5.1.4. Génération des règles d'association

Une fois que les ensembles d'itemsets fréquents sont extraits en utilisant l'algorithme Apriori, l'étape suivante consiste à générer des règles d'association à partir d'eux. Les règles d'association sont simplement des relations logiques entre deux éléments ou plus dans un ensemble de données transactionnelles.

Le processus de génération de règles d'association implique les étapes suivantes :

- Déterminez les seuils de support et de confiance minimums : Tout comme lors de l'extraction d'ensembles des motifs fréquents, définirez des seuils de support et de confiance minimums pour générer des règles d'association. Ces seuils aideront à filtrer les règles qui ne sont pas assez fortes.

- Générez toutes les règles possibles : Pour chaque ensemble d'éléments fréquents, on génère toutes les règles possibles en divisant l'ensemble en deux sous-ensembles non vides. Par exemple, si nous avons un ensemble d'itemsets fréquents de taille 3 :

  {Term1, Term2, Term3}, on peut générer les règles {Term1, Term2} {Term3}, {Term1, Term3} {Term2}, {Term2, Term3} {Term1}, {Term1} {Term2, Term3}, {Term3} {Term1, Term2} et {Term2} {Term1, Term3}.

- Calculez les mesures d'évaluation pour les règles d'association : Pour chaque règle générée, calculez son support, sa confiance, son lift, son leverage.

- Filtrez les règles faibles : Ensuite, nous pouvons filtrer les règles en fonction des seuils minimums que nous avons définis pour chacune de ces mesures d'évaluation.

- Triez et affichez les règles restantes : Triez les règles restantes par leur confiance et les afficher dans l'ordre décroissant.

Cependant, il est également possible de trier les règles en fonction d'autres mesures d'évaluation, telles que le lift ou la conviction. En fin de compte, la méthode de tri dépendra des objectifs spécifiques de l'analyse et des besoins de l'utilisateur.

5.2. TF-IDF

Dans notre projet, la deuxième méthode que nous avons utilisée s'appelle TF-IDF. Nous avons choisi ce nom car nous avons utilisé la mesure TF-IDF (Term Frequency-Inverse Document Frequency) comme composant clé de cette méthode. Nous avons appliqué cette mesure en conjonction avec deux algorithmes différents, à savoir Apriori et FP-Growth.

TF-IDF: TF-IDF signifie Term **F**requency **I**nverse **D**ocument Frequency of records. Il peut être défini comme le calcul de la pertinence d'un mot d'une série ou d'un corpus par rapport à un texte.

Il se compose de 6 étapes :

- Prétraitement de données ;
- Calcul de l'importance des termes (TF-IDF) ;
- Sélection les n term ayant les valeurs d'importance les plus élevées ;
- Représentation numérique (Création une matrice term-document) ;
- Extraction des motifs fréquentes avec l'algorithme Apriori / FP-Growth ;
- Génération des règles d'association

5.2.1. Prétraitement de données

Le texte a été prétraité de la même manière que dans la première méthode de Countvectorizer.

5.2.2. Calcul de l'importance des termes (TF-IDF)

TF-IDF a été utilisé pour calculer l'importance de chaque terme dans le corpus. Cette mesure prend en compte à la fois la fréquence du terme dans le document et sa rareté dans le corpus. Le calcul implique deux étapes :

- Fréquence de terme (TF) : Cela mesure la fréquence d'un terme dans un document. Il est calculé en divisant le nombre de fois où un terme apparaît dans un document par le nombre total de termes dans ce document.

TF (t, d) = nombre d'occurrence de t dans d / nombre de mots dans d

- Fréquence inverse de document (IDF) : cela mesure la rareté d'un terme dans le corpus.

Il est calculé en divisant le nombre total de documents dans le corpus par le nombre de documents contenant le terme, puis en prenant le logarithme de ce quotient.

$$\boldsymbol{IDF(t) = log10 = \frac{N}{N(t)}}$$

Formule 5 : Formule de la Fréquence Inverse du document

N = Nombre Totale de documents

N(t) = Nombre de documents contenant le terme t

TF-IDF est ensuite calculé en multipliant la valeur TF par la valeur IDF.

TF-IDF (t, d) = TF (t, d) * IDF(t)

5.2.3. Sélection les n term ayant les valeurs d'importance les plus élevées

Après avoir calculé les valeurs de TF-IDF pour ce terme dans le corpus, nous avons sélectionné les n termes ayant les valeurs les plus élevées pour une analyse plus ciblée et précise. Dans notre application, nous avons donné à l'utilisateur le contrôle sur le choix de la valeur de n en

pourcentage. Par défaut, nous avons fixé cette valeur à 0,25 (25%). Cela signifie que le système sélectionne les premiers 25% des termes de chaque document, en se basant sur leur score TF-IDF.

Cette étape de sélection est cruciale dans notre méthode, car elle nous permet de focaliser notre analyse sur les termes les plus importants et pertinents pour chaque document. En choisissant les termes avec les scores TF-IDF les plus élevés, nous nous assurons de prendre en compte les termes qui ont le plus d'influence dans la représentation numérique et l'analyse des données textuelles.

En donnant à l'utilisateur la possibilité de définir le pourcentage, nous offrons une flexibilité pour adapter l'analyse à ses besoins spécifiques. Si l'utilisateur souhaite une analyse plus complète, il peut augmenter le pourcentage (n) pour inclure plus de termes. Cela se traduira par une couverture plus large des termes, offrant une vue plus complète de l'ensemble de données. D'autre part, si l'utilisateur souhaite une analyse plus ciblée et concise, il peut diminuer le pourcentage (n) pour se concentrer sur un sous-ensemble plus restreint de termes très importants.

5.2.4. Représentation numérique (Création une matrice terme-document)

À l'aide des termes sélectionnés, nous créons manuellement une matrice term-document qui représente numériquement les données textuelles. Dans cette méthode, chaque document est représenté par un vecteur où chaque terme sélectionné correspond à une caractéristique et la valeur associée représente son importance dans le document. Contrairement à la première méthode où nous avons utilisé CountVectorizer pour créer la matrice

5.2.5. Extraction des motifs fréquents

Dans cette étape, nous utilisons deux méthodes :

- L’algorithme Apriori pour extraire les motifs fréquents à partir de la matrice term-document créée précédemment. L’algorithme Apriori explore progressivement l’espace des itemsets possibles, recherchant les ensembles d’items qui apparaissent fréquemment ensemble dans les documents analysés.

- L’algorithme FP-Growth pour extraire les motifs fréquents à partir de la matrice term-document. L’algorithme FP-Growth utilise une structure d’arbre compacte pour exploiter les relations entre les items et extraire efficacement les motifs fréquents.

5.3. Génération des règles d’association

Dans la deuxième méthode, les règles d’association sont générées de la même manière que dans la première méthode.

Une fois que les motifs fréquents sont identifiés à l’aide des algorithmes Apriori et FP-Growth, nous utilisons les mêmes étapes de génération des règles d’association que celles décrites dans la première méthode.

En utilisant les méthodes CountVectorizer et TF-IDF, nous avons progressé vers notre objectif d’extraire des règles à partir de textes écrits en hausa. Les approches différentes nous ont permis d’acquérir une meilleure compréhension des données et d’identifier les termes importants pour extraire les règles de manière plus efficace.

Ce qui va nous conduire dans le chapitre 6, à parler de l’implémentation, dans laquelle nous proposons des pistes de solution pour évaluer la perte d'information tout en intégrant aussi les notions introduites dans ce chapitre.

Chapitre 6 : Implémentation

Dans ce chapitre, nous allons explorer les outils et packages utilisés dans le développement d'une application de génération de règles d'association à partir de documents texte. Nous examinerons également de plus près l'interface de l'application et ses différents composants, notamment ses widgets, ses boutons et sa barre d'outils, etc. En comprenant la technologie derrière l'application et son interface utilisateur, nous pouvons mieux comprendre le processus de génération des règles d'association ainsi que la flexibilité et le contrôle que l'application offre à ses utilisateurs.

6.1 Outils et langages utilisés

6.1.1 Langage de programmation : Python

Python est un langage de programmation de haut niveau, interprété, orienté objet et multiparadigme. Il a été conçu dans les années 1990 par Guido van Rossum, et est connu pour sa syntaxe claire et concise, ainsi que pour sa grande facilité d'utilisation. Python est utilisé dans une variété de domaines, tels que le développement web, l'analyse de données, l'intelligence artificielle, l'automatisation de tâches, et bien plus encore.

6.1.2 Environnement de développement intégré (Anaconda)

6.1.2.1. Présentation

Anaconda est une distribution libre et open source des langages de programmation Python et R appliqué au développement d'applications dédiées à la science des données et à l'apprentissage automatique (traitement de données à grande échelle, analyse prédictive, calcul scientifique), qui vise à simplifier la gestion des paquets et de déploiement3. Les versions de paquetages sont gérées par le système de gestion de paquets conda.

La distribution Anaconda est utilisée par plus de 6 millions d'utilisateurs. La version d'installation comprend plus de 250 paquets populaires en science des données adaptés pour

Windows, Linux et MacOS. Plus de 7 500 paquets open-source supplémentaires peuvent être installés à partir de PyPI ainsi que du gestionnaire de paquets et d'environnements virtuels conda.

Elle comprend également une interface graphique, Anaconda Navigator,'qui est une alternative graphique à l'interface de ligne de commande (CLI).

La grande différence entre Conda et le gestionnaire de paquets pip consiste dans la gestion des dépendances des paquets.

6.1.2.2. Le navigateur Anaconda

Le Navigateur Anaconda est une interface graphique (GUI) incluse dans la distribution Anaconda, et qui permet aux utilisateurs de lancer des applications, mais aussi de gérer les librairies conda, les environnements et les canaux sans utiliser la moindre ligne de commande.

Le Navigateur peut également accéder à des librairies présentes sur le Cloud Anaconda ou dans un Repository Anaconda local, afin de les installer dans un environnement, les exécuter et les mettre à jour. Il est disponible pour Windows, macOS et Linux.

Les applications suivantes sont disponibles par défaut dans le navigateur :

- **JupyterLab ;**
- **Jupyter Notebook ;**
- **QtConsole ;**
- **Spyder…**

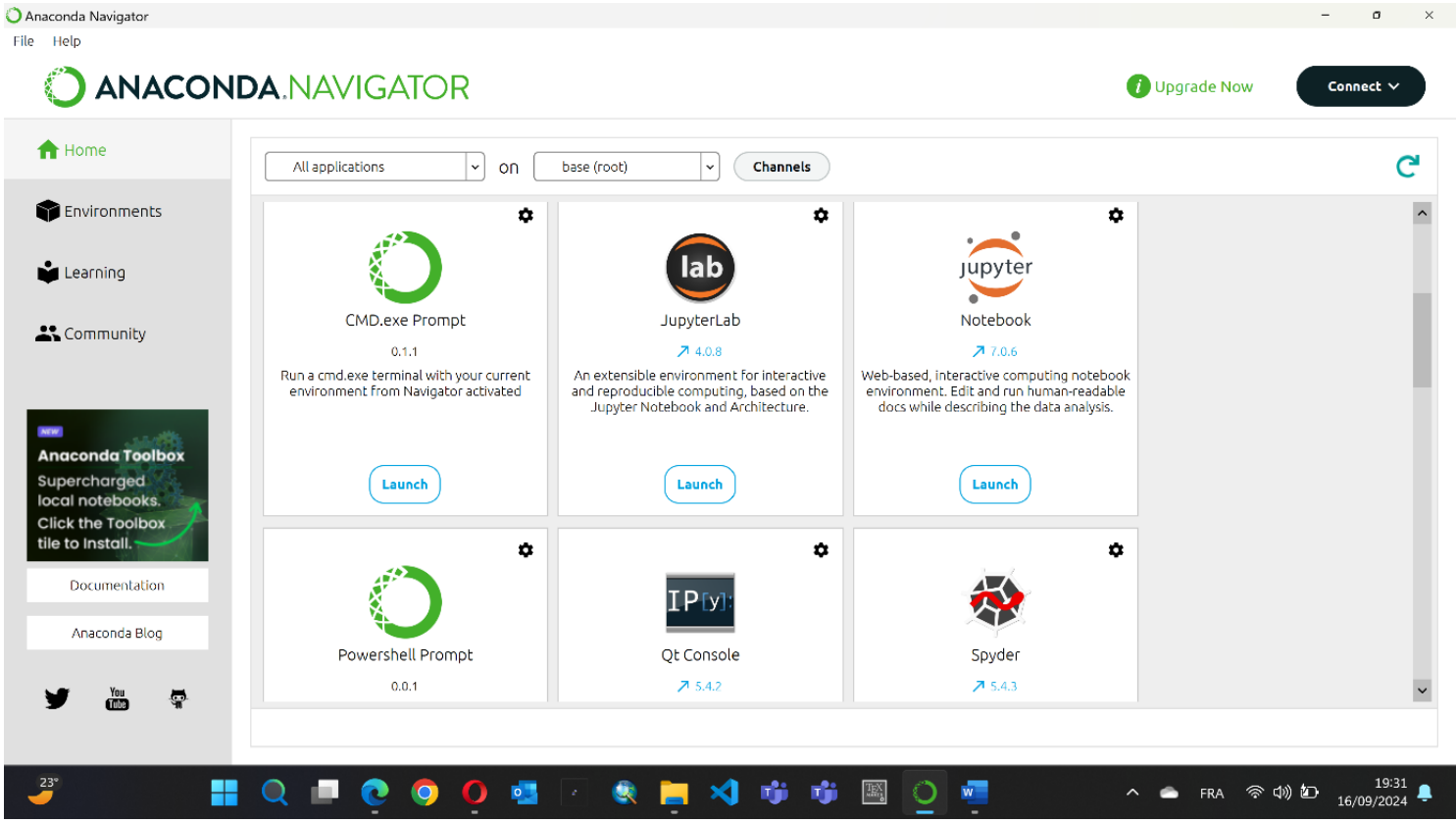

Figure 8 : Le navigateur Anaconda

Une fois le navigateur ouvert, on clique sur **Jupyter Notebook**.

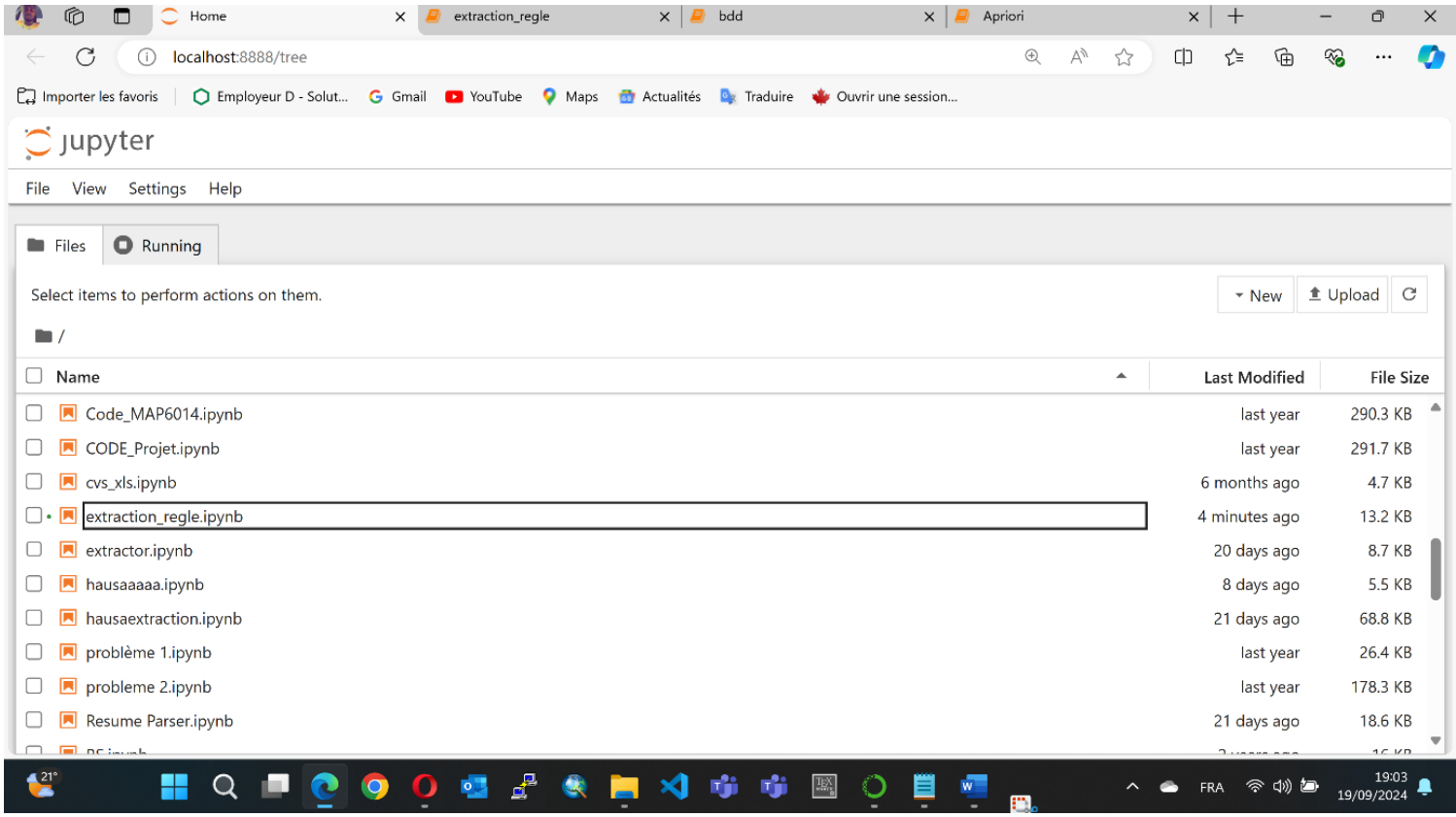

Figure 9 : Page d'accueil Jupiter Notebook

Une fois sur la page d'accueil, dans le coin droit, on clique sur **New** pour un nouveau projet et **upload** pour charger un document déjà existant.

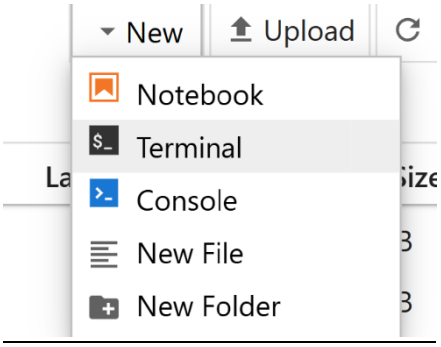

Figure 10 : Création nouveau projet

Lorsqu'on clique sur new, un menu se déroule et on choisit Notebook pour commencer notre projet.

6.2. Les packages utilisés

On a utilisé différentes bibliothèques de Python dans notre application tels que : Nltk, sklearn, mlxtend, pandas, PIL, Tkinter

6.2.1 Nltk

NLTK (Natural Language Toolkit) est une bibliothèque largement utilisée en Python pour les tâches de traitement du langage naturel (NLP) telles que le prétraitement et le nettoyage de texte. Elle offre divers outils et méthodes pour la tokenization, le stemming, la lemmatisation et plus encore. Parmi ses méthodes que nous utilisons dans notre application, nous utilisons la méthode nltk.word_tokenize pour tokenizer les documents, la méthode stopwords pour supprimer les mots courants des documents et la méthode WordNetLemmatizer pour lemmatiser les documents. NLTK est considéré comme l'un des packages les plus essentiels pour les tâches de NLP en Python en raison de sa polyvalence et de sa facilité d'utilisation.

6.2.2 Sklearn

Scikit-learn (ou sklearn) est une bibliothèque très utilisée en Python pour l'apprentissage automatique et l'exploration de données.

Dans notre application, nous avons importé la méthode CountVectorizer de sklearn, qui permet de convertir une collection de documents texte en une matrice de comptage de termes-document (terme-document matrix) pour l'analyse des fréquences des termes dans les documents. Nous avons utilisé CountVectorizer pour obtenir une matrice term-doc (terme-document) que nous avons ensuite utilisée comme entrée pour l'algorithme Apriori.

6.2.3 Mlxtend

La bibliothèque mlxtend est une bibliothèque Python pour l'apprentissage automatique et l'exploration de données. Dans notre application, nous importons l'algorithme Apriori et FP-Growth à partir de mlxtend.frequent_patterns, qui sont des algorithmes couramment utilisés

pour trouver les itemsets fréquentés. Nous importons également la méthode association rules, qui est responsable de la génération de règles d'association à partir des données traitées. La bibliothèque mlxtend est considérée comme une bibliothèque essentielle pour les tâches de fouille de données en raison de son efficacité et de sa simplicité d'utilisation.

6.2.4 Pandas

Pandas est une bibliothèque populaire en Python utilisée pour la manipulation de données. Dans notre application, nous avons utilisé pandas pour convertir la matrice terme-document en un DataFrame pandas.

Cela nous permet de mieux visualiser et manipuler les données, notamment pour la génération de règles d'association à l'aide de l'algorithme Apriori et FP-Growth de la bibliothèque mlxtend.

6.2.5.PIL

PIL (Python Imaging Library) est une bibliothèque de traitement d'images en Python qui permet de manipuler des images numériques en effectuant diverses opérations telles que l'ouverture, l'enregistrement et l'affichage d'images.[23] Nous avons utilisé PIL dans notre application pour charger et afficher des images, notamment pour afficher des icônes de fichiers et de boutons. PIL est largement utilisé pour le traitement d'images en Python et offre une grande variété de fonctions et de méthodes pour la manipulation d'images.

6.2.6 Tkinter

Tkinter est une bibliothèque GUI (Interface Utilisateur Graphique) standard de Python qui permet de créer des interfaces utilisateur fonctionnelles et attrayantes. Elle est utilisée dans notre application pour créer l'interface utilisateur, visualiser les résultats et personnaliser les paramètres de l'algorithme.

6.3. Countvectorizer

Dans notre projet, la première méthode que nous avons utilisée s'appelle CountVectorizer. Nous avons choisi ce nom, car nous avons utilisé la bibliothèque CountVectorizer de Python comme composant clé de cette méthode. CountVectorizer est une technique utilisée pour convertir des données textuelles en une représentation numérique pouvant être traitée par des algorithmes d'apprentissage automatique.

Il se compose de 4 étapes :

-Prétraitement de données ;

- Représentation numérique (Création une matrice term-document) ;

- Extraction des motifs fréquents ;

- Génération des règles d'association.

6.4. Interface graphique de notre application (G.U.I)

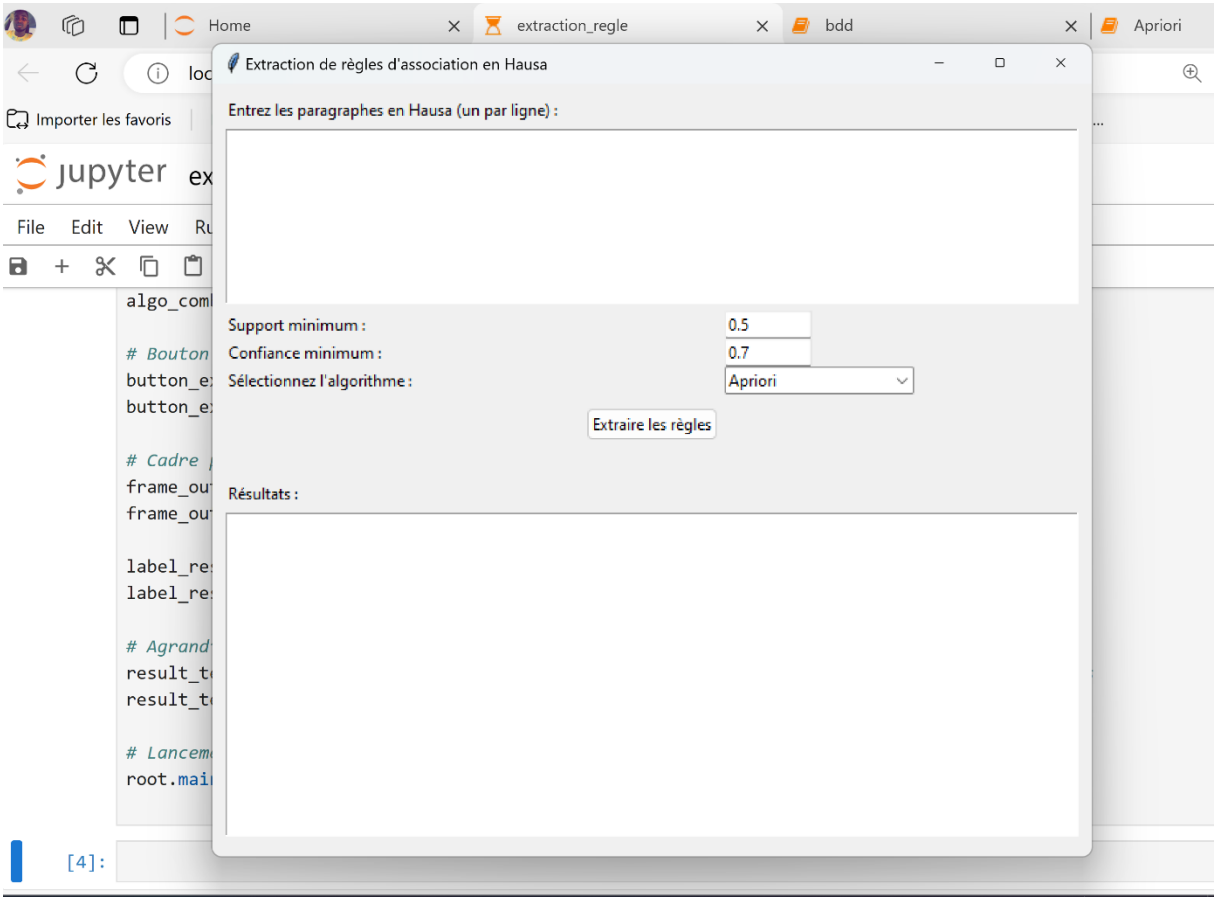

Figure 11 : Interface graphique de notre application

Notre application dispose d'une interface graphique intuitive et conviviale pour les utilisateurs.

Cette interface est composée de différents widgets, boutons, menus déroulants...etc. Dans cette section, nous allons examiner de plus près les éléments de l'interface utilisateur et expliquer leur fonctionnement.

- Les zones de saisie : nous avons tois cases à remplir. La première est destinée à la saisie du des paragraphes de textes hausa, la deuxième est destinée à la saisie du seuil de support minimal et la dernière a la saisie de la confiance minimale.

- La liste déroulante : qui permet à l'utilisateur de choisir l'algorithme qu'il souhaite utiliser pour la génération des règles d'association. Soit l'algorithme Apriori classique, soit s, Soit l'algorithme FP-Growth.

- Le bouton extraire les règles : comme son nom l'indique, une fois le textes saisit dans la première zone de saisie permettra d'extraire les règles d'associations.

- La zone des résultats : qui va afficher les résultats en se basant sur l'un des algorithmes choisit.

- La barre d'outils (toolbar) : elle contient deux boutons. Le premier bouton est intitulé "Description" et fournit des informations sur notre application, y compris notre approche et les algorithmes utilisés pour la génération de règles d'association. Le deuxième bouton est intitulé "about" et affiche des informations générales sur l'application, telles que le numéro de version et les détails du développeur. Ces boutons offrent aux utilisateurs un moyen rapide et pratique d'accéder aux informations de développement de l'application.

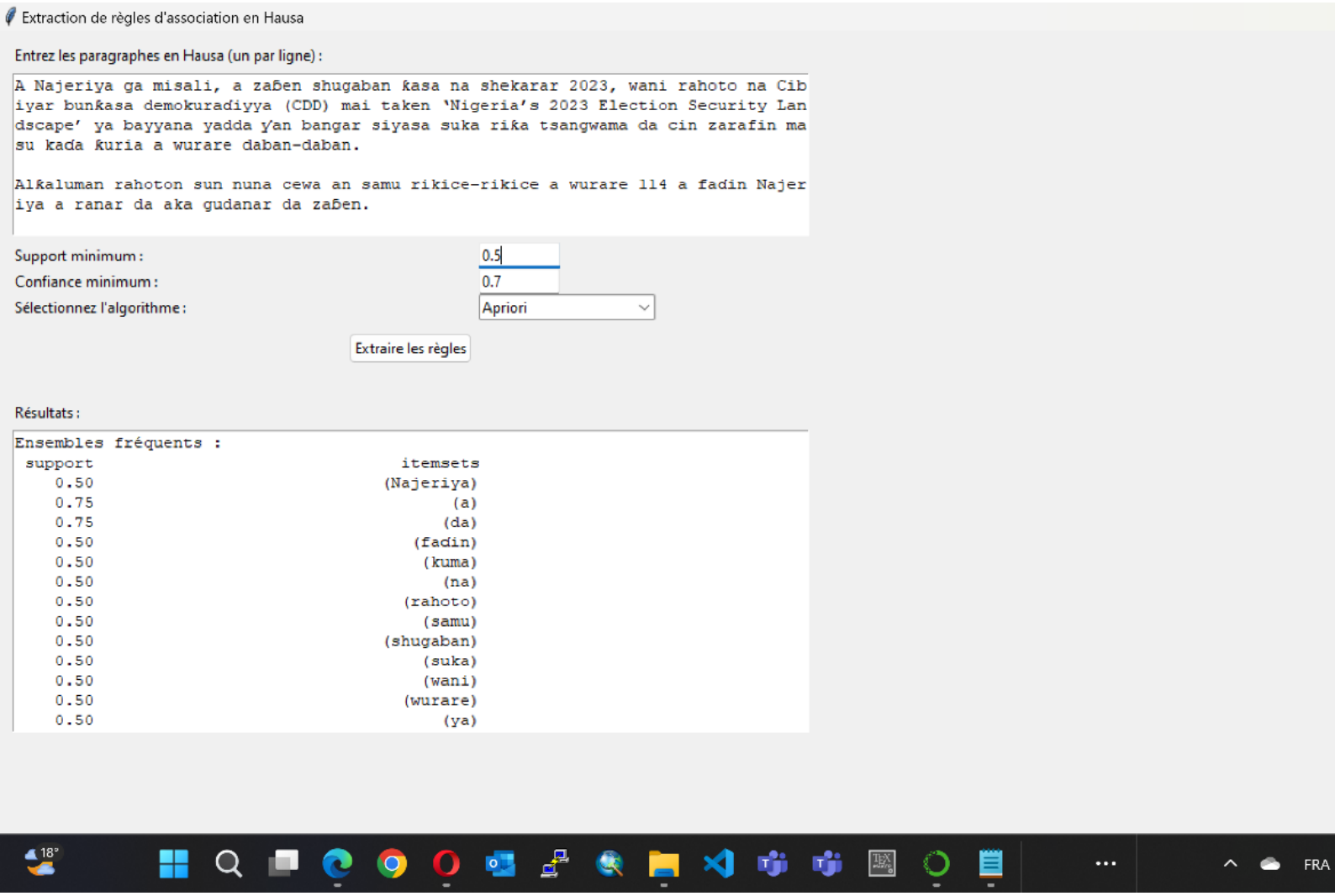

Figure 12: Interface Graphique de notre application après l'affichage

En définitive, ce chapitre a discuté des outils et des packages utilisés dans la mise en œuvre de l'application d'extraction de règles d'association de textes écrits en hausa, ainsi que les fonctionnalités de l'interface. Avec ces informations, les utilisateurs peuvent avoir une compréhension claire des options offertes de l'application et de la manière d'interagir avec elle.

Dans ce chapitre nous avons exposé le processus de traitement, l'implémentation, les différentes fonctionnalités et les paramètres de notre système. Dans le prochain chapitre, nous expérimentons notre système sur des données réelles en menant des interprétations et des discussions sur les résultats obtenus.

Chapitre 7 : Etude Expérimentale

Dans ce chapitre, nous allons inclure quelques exemples de textes sur un domaine d'étude, à savoir des informations liées à l'actualité. Nous allons étudier ces exemples et discuter des résultats obtenus à partir de l'application d'association de règles.

Texte 1 :

Niger ta kashe wani babban kwamandan dakarun juyin-juya-halin Mali Wani harin da Niger ta kai ya lalata ginin ofishin jakadancin. Mali da ke bobo babban birnin ƙasar Burkina, lamarin da ya yi sanadiyar mutuwa da kuma raunata mutane da dama, a cewar hukumomin ƙasar.

7.1. Résultat de l'ensemble 1

Support=50% (3/3), confiance=70%,

Algorithm : Apriori with Countvectorizer

Itemsets	Support
Mali	50%
Niger	50%
babban	50%
da	0.75%
ta	50%
ya	50%
ƙasar	50%
Mali, Niger	50%
ta, Mali	50%
da, ya	50%
ƙasar, da	50%
ta, Mali, Niger	50%

Tableau 3: Résultats 1

Les résultats que vous présentez montrent les ensembles fréquents générés avec l'algorithme Apriori et une implémentation utilisant CountVectorizer pour extraire les tokens (termes) des paragraphes en hausa. Voici une explication détaillée des résultats que vous obtenez :

7.1.1. Paramètres utilisés

- Support = 50% : Cela signifie que tout itemset (ensemble d'éléments) doit apparaître dans au moins **50% des transactions** (paragraphe ou phrase) pour être considéré comme fréquent.

- Confiance = 70% : Ce seuil n'est pas reflété dans les ensembles fréquents, mais serait utilisé dans la génération des règles d'association. Ici, seule la liste des ensembles fréquents est affichée.

- Algorithme : Apriori avec **CountVectorizer** : Cet algorithme est utilisé pour trouver les motifs fréquents, en transformant les paragraphes en ensembles de mots (itemsets).

7.1.2. Interprétation des résultats

- Ensembles fréquents : Voici quelques-uns des ensembles d'items (mots ou groupes de mots) avec leur support, c'est-à-dire la proportion des transactions dans lesquelles ils apparaissent. Par exemple :

- (Mali) a un support de 0.50, ce qui signifie qu'il apparaît dans **50% des transactions**.

- (da) a un support de 0.75, indiquant qu'il est présent dans **75% des transactions**.

- (Mali, Niger) a un support de 0.50, ce qui signifie que ces deux mots apparaissent ensemble dans 50% des transactions.

7.1.3. Exemples d'interprétation

Mots individuels fréquents : Les mots comme **Mali**, **Niger**, **babban**, **da**, **ta**, **ya**, et **ƙasar** apparaissent dans au moins 50% des paragraphes. Ce sont des mots importants dans le texte hausa analysé.

Paires de mots fréquents :

- (Mali, Niger) : Ces deux mots apparaissent ensemble dans 50% des transactions. Cela pourrait indiquer une forte relation entre ces deux pays dans le contexte des paragraphes.

- (ta, Mali) et **(ta, Niger)** : Cela suggère que la forme du mot "ta" (probablement un pronom ou un article) est associée à ces deux entités dans les paragraphes.

-Triplets fréquents : **(ta, Mali, Niger)** : Ce triplet montre que ces trois termes apparaissent ensemble dans 50% des transactions, suggérant peut-être une phrase ou une structure commune dans le texte.

Voici les résultats dans l'application :

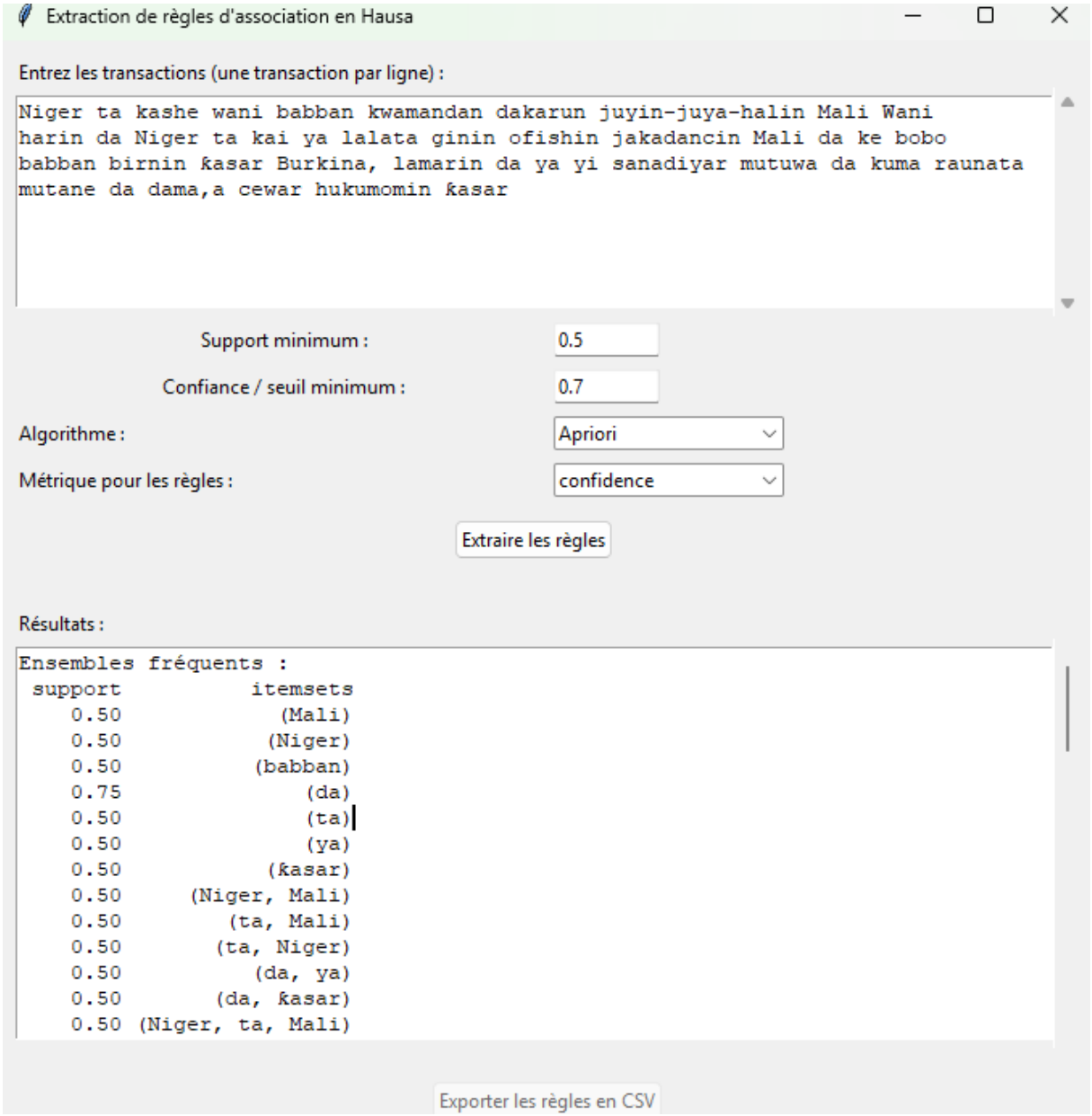

Figure 13: Résultats texte 1 dans l'application

7.1.4. Génération des règles d'association

En plus des ensembles fréquents, vous pouvez maintenant générer des **règles d'association** en appliquant le seuil de confiance (70%) sur les ensembles fréquents obtenus. Ces règles montrent la probabilité qu'un itemset B apparaisse, étant donné qu'un itemset A est déjà présent dans une transaction. Cela permet d'explorer les relations logiques entre les mots dans les paragraphes.

Les **règles d'association** qu'on a générées montrent des relations intéressantes entre les mots ou groupes de mots dans vos paragraphes en hausa. Voici une explication détaillée des différentes colonnes et de quelques exemples pour mieux comprendre ces résultats.

7.1.4.1. Colonnes des résultats

- Antécédents : L'ensemble d'items qui précède la règle. Par exemple, si (Mali) est dans les antécédents, la règle indique que lorsque "Mali" apparaît, il y a une probabilité que d'autres items (les "conséquents") apparaissent aussi.

- Conséquents : L'ensemble d'items qui suit la règle. Si (Niger) est dans les conséquents, cela signifie que lorsque "Mali" est présent (antecedent), il est probable que "Niger" apparaisse également.

- Antécédent Support : Le support de l'itemset dans les antécédents. Par exemple, pour (Mali), le support est de **0.50**, ce qui signifie que "Mali" apparaît dans **50% des transactions**.

- Conséquent Support : Le support de l'itemset dans les conséquents, similaire au support des antécédents.

- Support : Le support de la règle complète, c'est-à-dire le pourcentage de transactions où **les antécédents et les conséquents apparaissent ensemble**. Dans votre cas, toutes les règles ont

un support de **0.50**, ce qui signifie que ces combinaisons apparaissent ensemble dans 50% des transactions.

- Confidence : La **confiance** est la probabilité que les **conséquents** apparaissent dans une transaction, étant donné que les **antécédents** y apparaissent déjà. Dans votre exemple, toutes les règles ont une confiance de **1.0**, ce qui signifie qu'à chaque fois que les antécédents apparaissent, les conséquents apparaissent aussi. Cela indique une relation très forte.

- Lift : Le **lift** mesure la force de la règle en comparant la probabilité que les antécédents et les conséquents apparaissent ensemble avec la probabilité que les conséquents apparaissent indépendamment.

Un lift supérieur à 1 indique que les antécédents augmentent la probabilité que les conséquents apparaissent. Dans votre cas, toutes les règles ont un lift de **2.0**, ce qui signifie que la présence des antécédents **double la probabilité** de voir les conséquents dans les transactions.

- Leverage : mesure la différence entre le support observé de la règle et le support attendu si les antécédents et les conséquents étaient indépendants. Un leverage supérieur à 0 indique une association positive entre les items. Dans vos règles, le leverage est de **0.25**, ce qui montre une association positive.

- Conviction : est une autre mesure qui, comme le lift, évalue la force d'une règle. Une conviction infinie (**inf**) indique une certitude totale que les conséquents apparaîtront dès que les antécédents sont présents.

- Zhang's Metric : La métrique de Zhang mesure la direction et la force de l'association. Une valeur de **1.0** montre une **association positive parfaite** entre les antécédents et les conséquents.

7.1.4.2. Exemples d'interprétation

(Mali) → (Niger) :

- Confiance 1.0: Cela signifie qu'à chaque fois que "Mali" est présent, "Niger" est également présent dans 100% des cas.

- Lift 2.0: La présence de "Mali" double la probabilité de voir "Niger".

- Conviction : inf : Cela signifie qu'il est **absolument certain** que si "Mali" est présent, "Niger" apparaîtra aussi.

(ta) → (Mali) :

- La règle indique que lorsque le mot "ta" est présent, "Mali" apparaît dans toutes les transactions concernées, avec une confiance de **1.0**.

- Le lift de **2.0** montre que la probabilité que "Mali" apparaisse est doublée lorsque "ta" est présent.

(ya) → (da) :

- Confiance : 1.0 : Cela montre qu'à chaque fois que "ya" apparaît, "da" apparaît aussi.

- Lift : 1.33 : La présence de "ya" augmente la probabilité de voir "da" de 33%, ce qui est une association plus modérée que les autres.

En définitive, ces résultats montrent des associations très fortes entre certains mots dans les paragraphes en hausa que vous avez analysés. Toutes les règles montrent une **confiance parfaite** (1.0), ce qui signifie que les antécédents prédisent toujours les conséquents avec certitude, du moins dans le jeu de données utilisé.

[2]:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
0	(Mali)	(Niger)	0.5	0.50	0.5	1.0	2.0	0.25	inf	1.0
1	(Niger)	(Mali)	0.5	0.50	0.5	1.0	2.0	0.25	inf	1.0
2	(ta)	(Mali)	0.5	0.50	0.5	1.0	2.0	0.25	inf	1.0
3	(Mali)	(ta)	0.5	0.50	0.5	1.0	2.0	0.25	inf	1.0
4	(ta)	(Niger)	0.5	0.50	0.5	1.0	2.0	0.25	inf	1.0
5	(Niger)	(ta)	0.5	0.50	0.5	1.0	2.0	0.25	inf	1.0
6	(ya)	(da)	0.5	0.75	0.5	1.0	2.0	0.25	inf	0.5
7	(kasar)	(da)	0.5	0.75	0.5	1.0	2.0	0.25	inf	0.5
8	(ta, Mali)	(Niger)	0.5	0.50	0.5	1.0	2.0	0.25	inf	1.0
9	(ta, Niger)	(Mali)	0.5	0.50	0.5	1.0	2.0	0.25	inf	1.0
10	(Mali, Niger)	(ta)	0.5	0.50	0.5	1.0	2.0	0.25	inf	1.0
11	(ta)	(Mali, Niger)	0.5	0.50	0.5	1.0	2.0	0.25	inf	1.0
12	(Mali)	(ta, Niger)	0.5	0.50	0.5	1.0	2.0	0.25	inf	1.0
13	(Niger)	(ta, Mali)	0.5	0.50	0.5	1.0	2.0	0.25	inf	1.0

[ ]:

Rechercher

Figure 14: Règles d'association texte1

Les règles montrent des associations très fortes entre certains mots (ou groupes de mots) dans les paragraphes hausas. Les règles avec un lift de 2.0 et une confiance de 1.0 indiquent des co-occurrences systématiques. Cela peut aider à identifier des structures ou des relations

sémantiques régulières dans le texte. Vous pouvez ajuster ces règles en modifiant les seuils de support ou de confiance pour obtenir des résultats plus ou moins restrictifs.

7.1.5. Identification des structures ou des relations sémantiques régulières dans le texte

7.1.5.1 Relation entre les pays (Mali et Niger)

(Mali) → (Niger)

(Niger) → (Mali)

Structure sémantique : Il semble que, dans ce texte, Mali et Niger apparaissent ensemble de manière systématique. La relation entre ces deux pays est forte, avec une confiance de 100% (chaque fois que l'un apparaît, l'autre aussi) et un lift de 2.0. Cela indique que, dans le texte analysé, les deux pays sont fréquemment associés dans les mêmes contextes, peut-être dans une relation diplomatique ou géopolitique.

Implication sémantique : Cela pourrait refléter une discussion sur un conflit, une alliance ou des interactions fréquentes entre ces deux pays dans les paragraphes.

7.1.5.2. Structure des co-occurrences géopolitiques

Règle : (ta, Mali) → (Niger)

Structure sémantique : L'association entre "ta" (qui pourrait être une particule grammaticale ou un verbe) et "Mali" conduit à l'apparition de "Niger". Cela suggère que certaines structures de phrase qui incluent "ta" (peut-être une forme verbale ou un complément) sont souvent utilisées dans des discussions qui impliquent à la fois Mali et Niger.

Implication sémantique : Cela pourrait signaler une structure récurrente dans le discours qui lie ces deux pays, renforçant l'idée que le texte traite fréquemment de leur relation. L’usage de particules comme "ta" dans les phrases pourrait indiquer des événements ou actions partagés entre ces entités.

7.1.5.3. Relation entre des éléments grammaticaux ou verbes

Règles :

(ya) → (da)

(ƙasar) → (da)

Structure sémantique : Le mot "ya" (probablement un verbe ou un pronom en hausa) est toujours suivi de "da" (un mot qui pourrait être une conjonction ou une particule grammaticale). De même, "ƙasar" (qui signifie "pays" ou "nation" en hausa) est également fortement associé à "da". Cela pourrait représenter des structures grammaticales ou des phrases communes. **Implication sémantique** : Ces règles révèlent des modèles de phrases fréquents. Le mot "ƙasar" (pays) est souvent utilisé avec "da", ce qui pourrait suggérer des phrases comme "ƙasar da" (le pays et/ou avec). Ce type d'association est utile pour identifier des constructions grammaticales ou syntactiques stables dans le texte.

7.1.5.4. Interaction triangulaire (Mali, Niger, ta)

Règles :(Mali, Niger) → (ta)

(ta, Mali) → (Niger)

(ta, Niger) → (Mali)

Structure sémantique : Il y a une forte relation triangulaire entre "Mali", "Niger" et "ta". Quand "Mali" et "Niger" sont présents, "ta" apparaît aussi, et vice versa. Cela montre que ces trois éléments sont souvent utilisés ensemble dans des phrases ou paragraphes.

Implication sémantique : Ce genre de structure pourrait refléter des relations complexes ou interactives dans les discussions autour du Mali et Niger, impliquant souvent des actions ou événements (représentés par "ta"). Cela pourrait indiquer des phrases narratives fréquentes ou des dialogues dans lesquels les deux pays sont impliqués dans une action commune ou opposée.

7.1.5.5 Relation répétée dans le texte

Règles similaires avec confiance et lift élevés :

(Mali) → (ta)

(Niger) → (ta)

Structure sémantique : L'apparition de "ta" avec ces deux pays suggère que des actions ou des faits impliquant Mali et Niger sont régulièrement décrits dans le texte. "Ta" pourrait représenter un verbe ou une action qui revient dans les discussions impliquant ces entités.

Implication sémantique : Le texte semble relier souvent ces deux entités dans des situations actives où quelque chose est fait par ou à ces pays. Cela pourrait être utile pour comprendre la nature des discussions (par exemple, des actions militaires, des accords, etc.).

Synthèse des relations sémantiques identifiées : Le texte semble centrer les discussions autour de la relation Mali-Niger, avec de nombreuses cooccurrences entre ces deux entités. Cela pourrait impliquer des thématiques géopolitiques, des conflits ou collaborations, souvent structurés par des actions (représentées par "ta").

Des constructions grammaticales fréquentes comme l'usage de "da" avec des éléments tels que "ya" ou "ƙasar" montrent des modèles de phrases standard qui peuvent suggérer des structures descriptives ou narratives récurrentes dans le texte.

En résumé, ces règles permettent de repérer des patterns récurrents dans les paragraphes en hausa, qui peuvent être utiles pour comprendre la sémantique du texte, que ce soit au niveau lexical (présence de certains mots dans les mêmes contextes) ou syntaxique (structure des phrases).

7.2. Résultats de l’ensemble 2

Vous trouverez le texte lié à ce paragraphe dans l’annexe 3.

7.2.1. Paramètres utilisés

- Support = 50% : Cela signifie que tout itemset (ensemble d'éléments) doit apparaître dans au moins **50% des transactions** (paragraphe ou phrase) pour être considéré comme fréquent.

- Confiance = 70% : Ce seuil n'est pas reflété dans les ensembles fréquents, mais serait utilisé dans la génération des règles d'association. Ici, seule la liste des ensembles fréquents est affichée.

- Algorithme : Apriori avec **CountVectorizer** : Cet algorithme est utilisé pour trouver les motifs fréquents, en transformant les paragraphes en ensembles de mots (itemsets).

7.2.2. Interprétation des résultats

L'analyse des ensembles fréquents permet de repérer des combinaisons de mots qui apparaissent fréquemment dans les transactions (ou documents) du corpus hausa. Voici une analyse détaillée de quelques ensembles fréquents notables basés sur le support des itemsets.

- Ensembles de mots individuels (mots fréquents seuls)

(da) – Support : 1.000

Le mot "da", qui signifie "et" ou "avec", est le plus fréquent avec un support maximal. Cela montre son usage omniprésent dans le corpus.

(a) – Support : 0.9259Le mot "a" (dans, en, à) est également extrêmement fréquent, ce qui souligne son importance dans les phrases structurelles.

(ba) – Support : 0.8519Le mot "ba", qui est souvent utilisé dans des structures négatives, a un support élevé, suggérant une présence fréquente de phrases négatives dans le corpus.

(aka) – Support : 0.8148"aka" est souvent utilisé pour indiquer une action passée, courante dans les récits ou les descriptions d'actions dans le corpus. Ces mots sont principalement des connecteurs ou des particules grammaticales, ce qui reflète leur importance structurelle dans les phrases hausa.

- Ensembles de mots à deux termes (paires fréquentes)

(da, a) – Support : 0.9259

Cette combinaison est très fréquente et montre comment "da" et "a" sont souvent utilisés ensemble pour relier des mots ou des idées.

(aka, a) – Support : 0.7407

Cette paire indique que le verbe "aka" (utilisé pour marquer des actions passées) est souvent suivi de la particule "a", renforçant l'idée d'actions localisées ou liées à un endroit spécifique.

(mai, a) – Support : 0.6667"Mai" signifie "celui qui" ou "qui a", et "a" est une préposition. Cela montre l'importance des descriptions et des relations dans le texte.

- Ensembles de mots à trois termes (triplets fréquents)

(aka, Buhari, a) – Support : 0.5185

Ce triplet montre une association fréquente du président Buhari avec des événements ou des actions passées, ce qui indique que le corpus contient probablement des discussions politiques ou des récits liés à lui.

(da, aka, a) – Support : 0.7407

Cette combinaison de trois termes indique une structure répétée dans les phrases, où "da" (et), "aka" (passé) et "a" (préposition) sont utilisés ensemble pour structurer les phrases**.**

(aka, mai, karin) – Support : 0.5556

Cette combinaison suggère que l'action passée ("aka") est souvent liée à une personne ou à une entité possédant quelque chose ("mai") en lien avec une augmentation ou un ajout ("karin"). Il pourrait s'agir de discussions sur des améliorations ou des ajouts à un processus ou une situation dans le passé.

(aka, na, karin) – Support : 0.6667

Ici, "na" peut indiquer une possession ou un lien avec "karin". Cette combinaison suggère des phrases décrivant des augmentations ou des ajouts dans un contexte passé.

(aka, ne, karin) – Support : 0.5185

Ce triplet peut indiquer que certaines actions passées sont décrites comme étant des augmentations ou des ajouts (par exemple, une action passée d'ajout ou d'amélioration pourrait être relatée).

(aka, shi, karin) et (aka, karin, su) – Support : 0.5556

Ces combinaisons suggèrent que des actions passées liées à des individus ou des groupes sont souvent en lien avec des augmentations ou des améliorations.

L'analyse des ensembles avec "karin" révèle un thème commun d'augmentation ou d'ajout, souvent dans le contexte d'actions passées. Cela pourrait refléter des discussions sur des changements ou des améliorations dans différents domaines (par exemple, des progrès politiques, économiques ou sociaux).

Les ensembles fréquents avec "kuma" qui signifient "aussi" ou "également", et il apparaît fréquemment avec "aka" (passé), "na" (possession), et d'autres termes :

(kuma, na, aka) – Support : 0.5185

Cela suggère que "kuma" est utilisé dans des phrases pour ajouter des informations à une action ou un événement passé, souvent dans le cadre d'une possession ou d'un lien avec une entité.

(suka, aka, kuma) – Support : 0.5185

Cette combinaison indique que "suka" (ils/elles) est souvent associé à des actions passées avec une connotation d'ajout ou d'extension grâce à "kuma".Les ensembles contenant "kuma" soulignent la manière dont ce mot est utilisé pour étendre ou ajouter des détails à des événements passés, renforçant ainsi le lien entre les actions passées et des informations additionnelles dans le récit.

- Tendances générales

Les mots grammaticaux comme "da", "a", "aka", "ba" forment la majorité des ensembles fréquents, ce qui indique que ces connecteurs et particules jouent un rôle essentiel dans la construction des phrases dans le corpus. Les combinaisons de mots comme (Buhari, aka) ou (Yusuf, a) révèlent que certains personnages ou entités (comme Buhari et Yusuf) sont fréquemment mentionnés dans des actions ou des événements.

- Implications

Politique et narration : L'apparition fréquente de noms comme Buhari et Yusuf dans des combinaisons indique probablement des discussions autour de la politique ou de personnages influents, suggérant que le corpus traite de récits liés à ces individus.

Répétition de structures : Les combinaisons comme (da, a) et (aka, a) montrent des schémas récurrents dans la façon dont les phrases sont construites, ce qui pourrait indiquer une certaine formalité ou uniformité dans la rédaction. En résumé, l'analyse des ensembles fréquents montre

que le corpus est dominé par des connecteurs et des éléments grammaticaux, avec des mentions fréquentes de figures politiques.

La répétition de ces ensembles indique des structures linguistiques stables et des thèmes récurrents, probablement autour de la politique et de la narration.

7.2.3. Génération des règles d'association

L'analyse des règles d'association se fait à travers plusieurs métriques essentielles telles que le support, la confiance, le lift, et d'autres indicateurs comme la métrique de Zhang. Voici une explication des règles fournies à partir des données du texte 2.

- Support

Il mesure la fréquence d'apparition des éléments (antécédents et conséquents) dans les transactions. Par exemple, la règle (Buhari) -> (da) a un support de 0.592593, ce qui signifie que cette règle apparaît dans environ 59% des transactions. Confiance (confidence) :

Elle indique la probabilité que le conséquent (ce qui suit "->") soit présent lorsque l'antécédent est présent. C'est une mesure de précision. Par exemple, la règle (Buhari) -> (da) a une confiance de 1.0, ce qui signifie que chaque fois que "Buhari" est présent, "da" est toujours présent.

- Lift

Il mesure l'importance d'une règle. Un lift supérieur à 1 indique que les antécédents et conséquents sont corrélés positivement (ils ont tendance à apparaître ensemble plus souvent qu'on ne s'y attendrait).Par exemple, la règle (Buhari) -> (aka) a un lift de 1.150568, ce qui signifie que la présence de "Buhari" augmente la probabilité de trouver "aka" d'environ 15%.

- Leverage

Il calcule la différence entre le support attendu si les antécédents et conséquents étaient indépendants et le support observé. Une valeur proche de 0 signifie une indépendance des événements.

La règle (Buhari) -> (aka) a un leverage de 0.072702, ce qui indique une corrélation modérée.

Conviction

La conviction mesure à quel point les règles sont fiables, en particulier lorsqu'elles ne se vérifient pas (inversé de la confiance). Une valeur plus élevée signifie une forte conviction. Par exemple, la règle (Buhari) -> (aka) a une conviction de 2.962963, ce qui signifie que cette règle a de fortes chances d'être vérifiée.

Métrique de Zhang

Elle mesure le degré de dépendance asymétrique entre les éléments d'une règle. Une valeur positive signifie que les antécédents augmentent la probabilité d'observer le conséquent. Pour la règle (Buhari) -> (aka), la métrique de Zhang est de 0.321212, indiquant que "Buhari" augmente notablement la probabilité de voir "aka".

Exemples de règles intéressantes :

(Buhari) -> (aka) : Support : 0.555556, Confiance : 0.937500, Lift : 1.150568.

Cette règle montre une forte corrélation avec "aka" quand "Buhari" est présent. Le lift > 1 indique une association positive. (Yusuf) -> (aka) : Support : 0.555556, Confiance : 1.0, Lift : 1.227273. Cette règle est très forte avec une confiance de 100% et un lift supérieur à 1, montrant une association forte entre "Yusuf" et "aka".

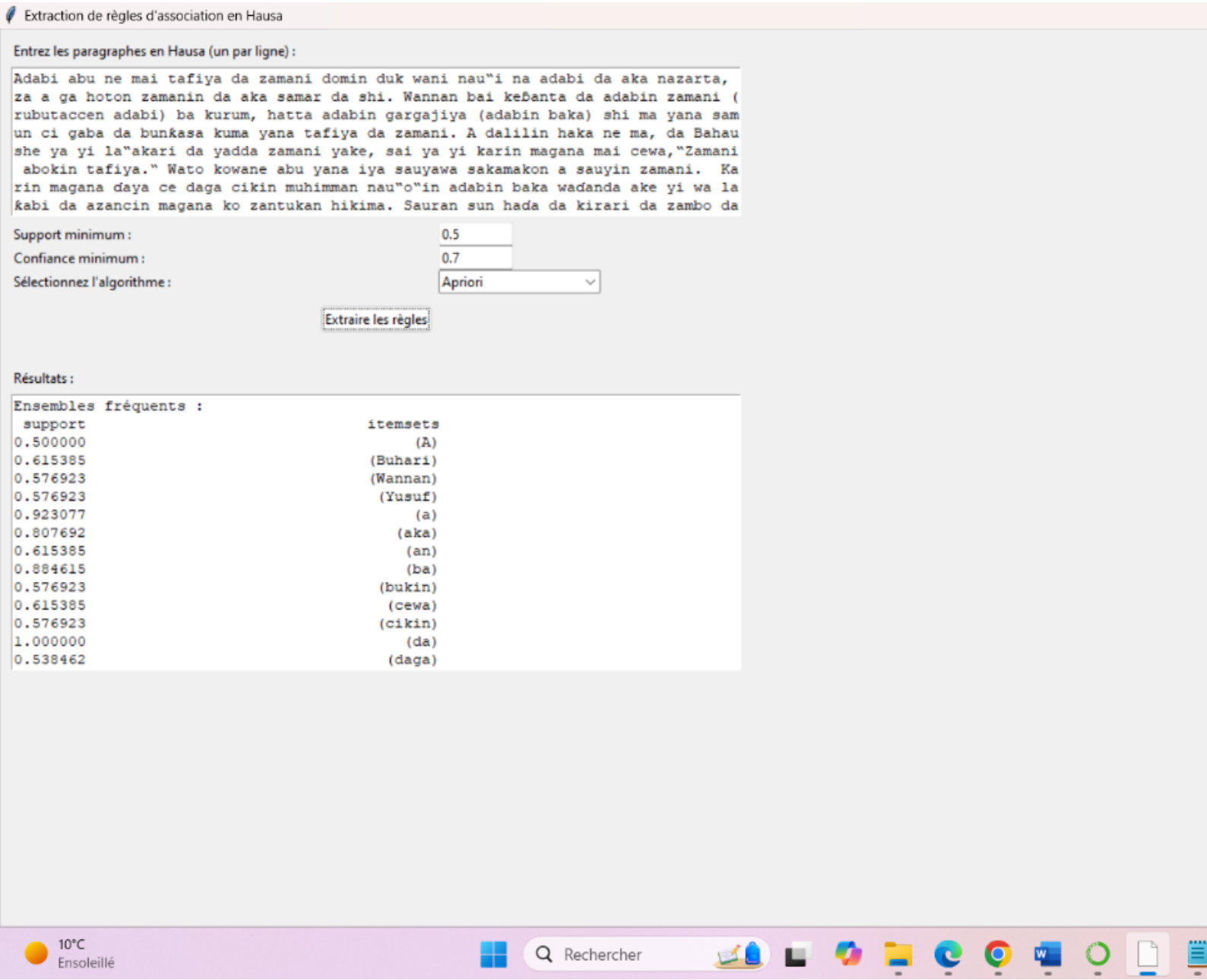

Figure 15: Ensembles fréquents texte 2

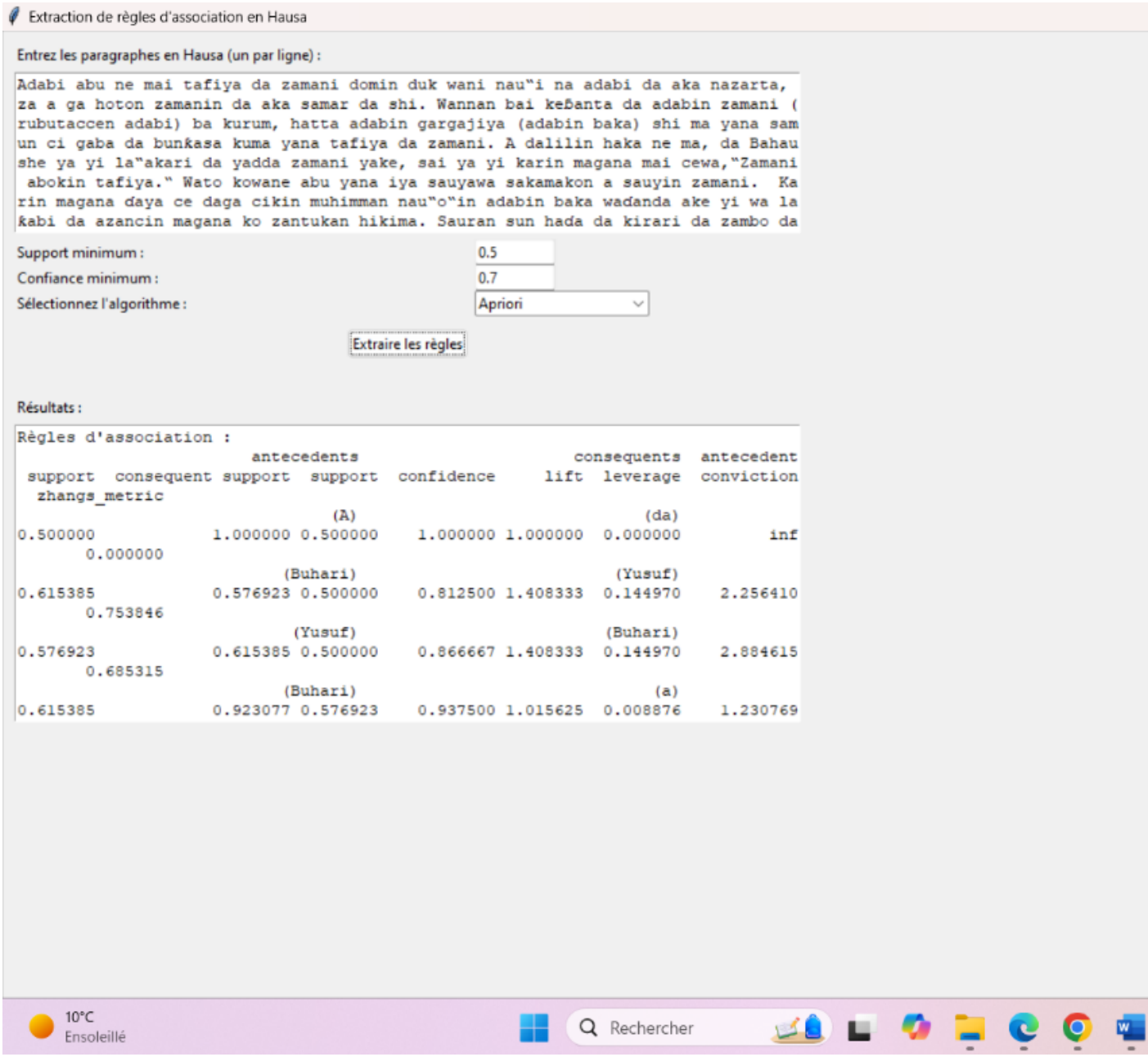

Figure 16: Règles d'association texte 2

Catégorie	Elément(s)	Détails/Interprétatio,
Paramètres d'analyse	Support = 50% Confiance = 70% Algorithme : Apriori	Seuils utilisés pour extraire les itemsets fréquents et générer des règles d'association
Mots fréquents (1 mot)	da (1.000) a (0.926) ba (0.852) aka (0.815)	Particules grammaticales structurelles, omniprésentes dans le corpus
Paires fréquentes (2 mots)	(da, a) – 0.926 (aka, a) – 0.741 (mai, a) – 0.667	Révèlent des constructions verbales fréquentes
Triplets fréquents (3 mots)	(aka, Buhari, a) – 0.519 (aka, na, karin) – 0.667 (da, aka, a) – 0.741	Témoignent de structures narratives liées à des actions ou descriptions
Mot-clé thématique	**karin** ("ajout", "augmentation")	Présent dans plusieurs triplets fréquents thème de progression ou amélioration
Mot-clé connecteur	**kuma** ("aussi", "également")	Sert à ajouter ou étendre des informations dans les récits
Personnalités citées	Buhari, Yusuf	Fréquemment associés à des verbes d'action (aka) thématiques politiques dominantes
Règles intéressantes	(Buhari → aka) : Support 0.556, Confiance 0.94, Lift 1.15 (Yusuf → aka) : Support 0.556, Confiance 1.00, Lift 1.23	Indiquent des associations fortes entre noms de personnes et verbes d'action
Tendances générales	Dominance des particules grammaticales Structures répétitives- Thèmes politiques et narratifs	Le corpus présente une structure régulière, avec des schémas linguistiques constants
Indications linguistiques	Structures stables Usage élevé des connecteurs	Style formel ou narratif régulier dans les documents analysés
Implication thématique	Politique, narration d'événements passés	Les noms propres et verbes d'action récurrents traduisent des récits d'événements liés à des figures politiques

Tableau 4: Tableau récapitulatif

Conclusion et perspectives

En définitive, ce travail s'est concentré sur l'extraction des règles d'association de textes écrits en hausa, en utilisant l'algorithme Apriori et une combinaison de TF-IDF avec l'algorithme Apriori et l'algorithme FP-Growth. Ce mémoire a permis d'obtenir des informations précieuses sur l'efficacité de ces méthodes dans différents scénarios. L'algorithme d'Apriori sans TF-IDF s'est avéré efficace lorsque tous les documents contenaient les mêmes mots-clés ou les mêmes termes. Cependant, lorsque l'on traite de différents facteurs ou aspects dans les documents, l'intégration de TF-IDF avec les algorithmes d'extraction de règles d'association a donné des résultats prometteurs.

Malheureusement, le manque de documents, de données ou encore de datasets hausa ont limité la possibilité d'explorer et d'étudier plus en profondeur d'autres approches et algorithmes.

Dans le domaine du texte, nous sommes confrontés à des corpus en constante modification. Chaque nouvelle page peut possiblement contenir des informations que le système peut ne jamais avoir rencontrées. Cependant, il existe des perspectives potentielles pour de futures recherches dans ce domaine.

Celles-ci consisteront à mettre en place des datasets hausa qui seront à porter de main, mais aussi de La combinaison des autres techniques d'exploration de texte, telles que wordembeddings ou topic modeling, avec l'extraction de règles d'association, y compris l'extraction des règles avec d'autres algorithmes tels que ECLAT, afin d'améliorer l'extraction de règles significatives. Enfin, utiliser les règles d'association comme étiquettes dans une classification supervisée pour former nos modèles. Cette approche permettrait d'exploiter pleinement les informations contenues dans les règles d'association extraites, en les utilisant comme des indicateurs pour l'entraînement des modèles de classification. Cela pourrait conduire à une meilleure performance des modèles en exploitant les relations et les dépendances entre les termes dans les données textuelles.

Références

[1]. Yacubu M.A. (2022). Algaita. In revue. Journal of current research in hausa studies. Vol. 15, No. 1. Kano, Nigeria.

[2]. Adamou M.R. (2014). Le haoussa. Dans le Dictionnaire bilingue Haoussa. Institut National de Documentation de Recherches et Animations Pédagogiques. Niamey, Niger.

[3]. Charif K, Soualem C. (2023). Extraction des règles d'association à partir du texte. Mémoire de master en Informatique. Bordj Bou. Arréridj; Algérie .

[4].Caron B, Bonivi E, Busuttil J, Peyraube A. Haoussa. (2011). Dictionnaire des Langues, Presses Universitaires de France, pp.263-269. ffhalshs-00643960.

[5]. Meunier J-G, Biskri I, Nault G, Nyongwa M. (1997). Exploration de classifieurs Connexionnistes pour l'analyse de textes assistée par ordinateur. In conference proceedings. Montréal (Québec), Canada.

[6]. Mijinguin A; Naroua H. (2012). Règles de formation des noms en haoussa. In revue. Agence Nigérienne des Langues et du Livre (ANLL), Niamey, Niger.

[7]. Ahmed N. (2009). Adaptation des écritures et de la lecture des langues étrangères au pays Haoussa de l'Afrique de l'Ouest. In revue. Synergies Algérie n° 6- 2009 pp. 61-69. Région du Sunyani Brong-Ahafo, Ghana.

[8]. Caron, Bernard. (2011). La grammaticalisation de l'enfance en haoussa. In revue. Afrika und Übersee, 2008, 88 (1-2), pp.53-62. ffhalshs-00644321 Paris, France.

[9]. Tassaou B. (2013). Formes des interférences linguistiques dans l'enseignement bilingue : Cas de l'école bilingue de Dogondoutchi [rapport de recherche]. Niamey, Niger.

[10]. Mackey W. (1976). Initiation à la linguistique : bilinguisme et contact de langues. Paris, France.

[11]. Gaoh Z.B. (2008). Grammaire comparée hausa-français : référence pour l'enseignement Bilingue Niamey [rapport de recherche]. Institut National de Documentation, de Recherche et d'Animation Pédagogiques (INDRAP). Niamey, Niger.

[12]. Gaoh Z.B. (2008). Karatu da rubutu Aji na 2, littafin dan makaranta [rapport de recherche]. Institut National de Documentation, de Recherche et d'Animation Pédagogiques (INDRAP). Niamey, Niger.

[13]. Bedhouch L. (2023). Vers une approche de réduction de nombre de règles d'associations Mémoire de maitrise. Trois Rivières, Québec, Canada.

[14]. Newman M.R. (1990). An English-Hausa Dictionary ;New Haven. Yale University Press, pp 327.

[15]. Scalise S. (1984). Generative Morphology. Dordrecht, Netherlands.

[16]. Ray J. (1975). Morphological and Semantic Regularities in the Lexicon. Linguistic Societyof America. Washington, DC, USA.

[17]. Meunier J.G, Biskri I, Jouis C, Le Priol F, Descles J.P. et Mustafa E. (1997). « Outil d'aide à la fouille documentaire : approche hybride numérique linguistique ». Théories et Outils pour le Traitement Automatique des Langues, Revue Annuelle BULAG. Besançon, France(Hors-S), pp. 35-43.

[18]. Ait El H.A. (2019). Etat de l'art sur les règles d'associations séquentielles : proposition d'une nouvelle solution PSBI. Mémoire de master en Informatique. Universite Mouloud Mammeri de Tizi Ouzou.

[19]. Agrawal R., Amielinski T, Swami A. (1993). Mining association rule between sets of items in large databases. Proceedings of the ACM SIGMOD International Conference on Management of Data, (pp. 207-216). Washington,DC.

[20]. Agrawal R, Srikant R. (1994). Fast algorithms for mining association rule. Proceedings of the 20th International Conference on Very Large Data Bases (pp.487-499), Santiago, Chile.

[21]. Ashrafi M Z, Taniar D, Smith, K. (2007). Redundant association rules reduction techniques. International Journal of Business Intelligence and Data Mining, 2, 29-63.

[22]. Liu B, Hsu W, Ma Y. (1998). Integrating classification and association rule mining. In Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining (KDD98), 80-86. AAAI Press.

[23]. Turčínek P, Turčínková J. (2015). Exploring consumer behavior: Use of association rules. Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis, 63, 1031-1042.

[24]. Bokhabrine A, Biskri I, Ghazzali N. (2020). Textual Clustering: Towards a More Efficient Descriptors of Texts. In International Conference on Computational Collective Intelligence, 801-810. Springer.

[25]. Zeng Y, Yin S, Liu J, Zhang M. (2015). Research of improved FP-Growth algorithm in association rules mining. Scientific Programming.

[26]. Agrawal R, Srikant R. (1994). Fast algorithms for mining association rules. In Proc. 20th int. conf. very large data bases, VLDB, 487-499. Citeseer.

[27]. Kaur P, Singh R. (2019). An efficient Apriori-TID algorithm for mining frequent itemsets in transactional databases. International Journal of Grid and High Performance Computing, 11(1), 40-53.

[28]. Halle M. (1973). The accentuation of Russian word. In revue. Linguistic Society of America. Language, Vol. 49, No. 2.

[29].Yahya H. (2007). L'Atlas de la création. Editions Global. Istambul. Vol 1 772 p.

[30]. Lemu A. (2002). Women in Da'wah. A working paper presented to a trustees meeting of the International Council for Islamic Information. Markfield Da'wah Centre. Leicester, U.K.

Annexe 1 code sans interface graphique

```
import pandas as pd

from mlxtend.preprocessing import TransactionEncoder

from mlxtend.frequent_patterns import apriori, association_rules

# Exemple de liste de paragraphes en hausa

paragraphes = [

    "Niger ta kashe wani babban kwamandan dakarun juyin-juya-halin Mali Wani harin da Niger ta kai ya lalata ginin ofishin jakadancin."

   " Mali da ke bobo babban birnin ƙasar Burkina,lamarin da ya yi sanadiyar mutuwa da kuma raunata mutane da dama, a cewar hukumomin ƙasar.",


# 1. Prétraitement : Tokenisation (chaque paragraphe devient une transaction de mots uniques)

transactions = [set(paragraph.split()) for paragraph in paragraphes]

# 2. Transformation en format transactionnel pour Apriori

te = TransactionEncoder()

te_ary = te.fit(transactions).transform(transactions)

df = pd.DataFrame(te_ary, columns=te.columns_)


# 3. Application de l'algorithme Apriori

frequent_itemsets = apriori(df, min_support=0.5, use_colnames=True)

# 4. Génération des règles d'association


rules = association_rules(frequent_itemsets, metric="confidence", min_threshold=0.7)

# Affichage des ensembles fréquents et des règles

print("Ensembles fréquents :\n", frequent_itemsets)

print("\nRègles d'association :\n", rules)print("\nRègles d'association :\n", rules)
```

Annexe 2 : Code avec interface graphique (1/4)

```
import tkinter as tk
from tkinter import ttk, messagebox
import pandas as pd
from mlxtend.preprocessing import TransactionEncoder
from mlxtend.frequent_patterns import apriori, fpgrowth, association_rules

# Fonction pour l'extraction des règles d'association
def extract_rules():
    # Récupérer les entrées utilisateur
    paragraphes_input = text_paragraphs.get("1.0", "end").strip().split("\n")
    min_support = float(entry_support.get())
    min_confidence = float(entry_confidence.get())
    selected_algo = algo_combobox.get()

    # Prétraitement des données
    transactions = [set(paragraph.split()) for paragraph in paragraphes_input if paragraph.strip()]

    if not transactions:
        messagebox.showwarning("Erreur", "Veuillez entrer des paragraphes valides.")
        return

    # Transformation en format transactionnel
    te = TransactionEncoder()
    te_ary = te.fit(transactions).transform(transactions)
    df = pd.DataFrame(te_ary, columns=te.columns_)

        return
```

Annexe 2 : Code avec interface graphique (2/4)

```
# Application de l'algorithme sélectionné
    if selected_algo == "Apriori":
        frequent_itemsets = apriori(df, min_support=min_support, use_colnames=True)
    elif selected_algo == "FP-Growth":
        frequent_itemsets = fpgrowth(df, min_support=min_support, use_colnames=True)
    else:
        messagebox.showerror("Erreur", "Algorithme non supporté.")
        return

    # Vérifier si des ensembles fréquents ont été trouvés
    if frequent_itemsets.empty:
        messagebox.showinfo("Résultat", "Aucun ensemble fréquent trouvé avec le support fourni.")
        return

    # Génération des règles d'association
    rules = association_rules(frequent_itemsets, metric="confidence", min_threshold=min_confidence)

    if rules.empty:
        messagebox.showinfo("Résultat", "Aucune règle d'association trouvée avec la confiance fournie.")
        return
# Affichage des résultats dans la zone de texte de sortie
    result_text.delete("1.0", "end")
    result_text.insert("end", "Ensembles fréquents :\n")
    result_text.insert("end", frequent_itemsets.to_string(index=False))
    result_text.insert("end", "\n\nRègles d'association :\n")
    result_text.insert("end", rules.to_string(index=False))
```

Annexe 2 : Code avec interface graphique (3/4)

```
# Création de la fenêtre principale
root = tk.Tk()
root.title("Extraction de règles d'association en Hausa")

# Cadre pour l'entrée des paragraphes
frame_input = ttk.Frame(root, padding="10")
frame_input.grid(row=0, column=0, sticky="ew")

label_paragraphs = ttk.Label(frame_input, text="Entrez les paragraphes en Hausa (un par ligne) :")
label_paragraphs.grid(row=0, column=0, sticky="w")

# Agrandissement de la zone de saisie des paragraphes
text_paragraphs = tk.Text(frame_input, height=8, width=80)  # Ajustement de la taille pour plus de lisibilité
text_paragraphs.grid(row=1, column=0, columnspan=2, pady=5)

# Entrée pour le support minimum
label_support = ttk.Label(frame_input, text="Support minimum :")
label_support.grid(row=2, column=0, sticky="w")
entry_support = ttk.Entry(frame_input, width=10)
entry_support.grid(row=2, column=1, sticky="w")
entry_support.insert(0, "0.5")  # Valeur par défaut
# Entrée pour la confiance minimum
label_confidence = ttk.Label(frame_input, text="Confiance minimum :")
label_confidence.grid(row=3, column=0, sticky="w")
entry_confidence = ttk.Entry(frame_input, width=10)
entry_confidence.grid(row=3, column=1, sticky="w")
entry_confidence.insert(0, "0.7")  # Valeur par défaut
```

Annexe 2 : Code avec interface graphique (4/4)

```
# Sélection de l'algorithme (Apriori ou FP-Growth)
label_algo = ttk.Label(frame_input, text="Sélectionnez l'algorithme :")
label_algo.grid(row=4, column=0, sticky="w")
algo_combobox = ttk.Combobox(frame_input, values=["Apriori", "FP-Growth"], state="readonly")
algo_combobox.grid(row=4, column=1, sticky="w")
algo_combobox.current(0)  # Par défaut, l'algorithme est "Apriori"

# Bouton pour lancer l'extraction des règles
button_extract = ttk.Button(frame_input, text="Extraire les règles", command=extract_rules)
button_extract.grid(row=5, column=0, columnspan=2, pady=10)

# Cadre pour l'affichage des résultats
frame_output = ttk.Frame(root, padding="10")
frame_output.grid(row=1, column=0, sticky="ew")

label_result = ttk.Label(frame_output, text="Résultats :")
label_result.grid(row=0, column=0, sticky="w")

# Agrandissement de la zone de texte des résultats
result_text = tk.Text(frame_output, height=15, width=80)  # Taille ajustée pour les résultats
result_text.grid(row=1, column=0, pady=5)

# Lancement de la fenêtre
root.mainloop()
```

Annexe 3 : Texte 2 de l'étude expérimentale

Texte 2 :

Adabi abu ne mai tafiya da zamani domin duk wani nau"i na adabi da aka nazarta, za a ga hoton zamanin da aka samar da shi. Wannan bai keɓanta da adabin zamani (rubutaccen adabi) ba kurum, hatta adabin gargajiya (adabin baka) shi ma yana samun ci gaba da bunƙasa kuma yana tafiya da zamani. A dalilin haka ne ma, da Bahaushe ya yi la"akari da yadda zamani yake, sai ya yi karin magana mai cewa,"Zamani abokin tafiya." Wato kowane abu yana iya sauyawa sakamakon a sauyin zamani. Karin magana ɗaya ce daga cikin muhimman nau"o"in adabin baka waɗanda ake yi wa laƙabi da azancin magana ko zantukan hikima. Sauran sun haɗa da kirari da zambo da habaici dabaƙar magana da gatse da sauran ire-irensu. Idan aka dubi karin maganar Hausa ta fuskar zamani, za a tarar akwai hoton zamuna daban-daban da Bahaushe ya ratso a cikinsu. Bahaushe ya rayu a lokacin maguzanci sannan daga baya ya karɓi addinin Musulunci. Ana nan ana tafiya kuma sai Turawa suka shigo. Duk waɗannan zamunan sai da karin magana ta taskace su. Alal misali, akwai karin maganar da ake cewa, „Dodo ɗaya ake yi wa tsafi" sannan a wata karin maganar kuma cewa ake „Me na ci na asham da zan yi ramuwar sallah? Akwai wata karin maganar kuma wadda take cewa, „Aiki da hankali ya fi aiki da agogo." Waɗannan karin maganganu suna wakiltar zamunan da aka ambata a baya ne kamar yadda aka jero su. A zamanin da muke ciki a ƙarni na ashirin da ɗaya, an samu ci gaba ta fasahar ƙirƙire-ƙirƙire wanda ya haifar da yawaitar shafukan sada zumunta na zamani. Daga cikin waɗannan shafukan akwai Fesbuk (Facebook) da Was"af (WhatsApp) da sauran makamantansu. A cikin waɗannan shafukan, akan yi amfani da nau"o"in adabi da dama musamman ma karin magana. Irin kare-karen maganar da ake amfani da su a waɗannan shafukan suna da tasirin zamani matuƙa kasancewar yawanci matasa ne suke amfani da shafukan.

Wannan muƙala ta nazarci irin karin maganar da aka yi amfani da su a shafukan Fesbuk da Was‟af waɗanda suka danganci bukin ɗan shugaban ƙasa, wato Yusuf Buhari da xiyar Sarkin Bichi wanda aka yi a watan Agusta, 2021. Karin Magana: Ma‟ana, Ire-ire da kuma Hanyoyin Nazarinsa .A Kamusun Hausa (2006:235), an bayyana karin magana da cewa magana ce ta musamman wadda sai an yi tsokaci ake gane ta.A yayin da Tudun Wada (2006), ya bayyana karin magana da wata dunƙulalliyar magana ta hikima wadda kan ƙunshi faffaɗar ma‟ana musamman idan aka tashi bayanin manufarta. Finnegan (1970) ta yi nuni da cewa, karin magana ɗaya ce daga cikin nau‟o‟in zantukan hikima da yawancin al‟ummomin Afirka suna da ita. Ana iya samun dangantaka a tsakanin karin magana. Ita wannan dangantakar tana iya kasancewa ta fuskar tsarin karin maganar ko ma‟anarta ko ma duka biyun (Tadi, 2005:103). Dangane da ire-iren karin magana kuwa, akwai hanyoyi daban-daban da masana suka bi wajen karkasa ta. Alal misali, wasu sun yi la‟akari ne da tsarin jumloli, wasu kuma kalmomin da suke ƙunshe a cikin karin maganar sannan wasu kuma saƙon suka duba. A taƙaice, Tudun Wada (2006) da kuma Malumfashi da Nahuce (2014) sun kawo waɗannan nau‟o‟in: Karin magana mai ƙwar ɗaya, misali: Duk kanwar ja ce. Karin magana mai ƙwar biyu, misali: Ba saban ba, ɓera ya je zance. Karin magana mai labari, misali: Kamun gafiyar Ɓaidu. Karin magana mai „inji‟, misali: „Ban sa a ka ba‟ inji ɓarawon tagiya. Karin magana mai „sai ;, misali: Sai uwa ta ƙoshi, ɗanta ke malolo. Karin magana mai „an ce‟, misali: „Mutum ba ya ƙin ta mutane‟ an ce da ɓarawo ya gudu. Karin magana mai „daga‟, misali: „Daga baya‟, wai sadaka da bazawara Karin magana mai tambaya, misali: „Kai kuma a su wa‟ kare da gudun layya ? Karin magana mai „ko‟, misali: Ko yanzu ruwa na maganin dauɗa. Karin magana mai „dole‟, misali: Dole ƙanwar „na ƙi‟. Karin magana mai „akan‟, misali: A bakin wawa akan ji magana. Karin magana mai „ta‟ misali: Ta faru ta ƙare, an yi wa mai dami ɗaya sata. Karin magana mai „tun‟, misali: Tun kafin a yi ɗaran, aka yi kwandi. Karin magana mai wanda, misali: Wanda ya ƙi ji, ba zai ƙi gani ba. Karin magana mai „da‟, misali: Da babu gara ba dadi. Karin magana mai „a‟, misali: A sa a baka ya fi a rataya. Karin magana mai „ba‟, misali: Ba girin-girin ba, ta yi mai. Karin magana mai tasirin gargajiya, misali: Fankam-fankam ba shi ne kilishi ba, tsomi. Karin magana mai tasirin Musulunci, misali: Kyauta daga Allah, gwauro da „yan tagwaye. Karin magana mai tasirin zuwan Turawa, misali: Aiki da hankali ya fi aiki da agogo. Skinner (1977) yana mai ra‟ayin cewa an fi nazarin karin magana ta fuska biyu: jigo da tsari. Amma kuma a cewarsa, akwai wasu hanyoyin nazarin karin magana ; misali salo.

Irin waɗannan hanyoyin su masana da manazarta da dama suka bi wajen nazarin karin maganganun Hausa. Daga cikinsu akwai Tudun Wada (2006) da Junaidu da „Yar‟aduwa (2007) da Ɗanhausa (2012) da Gwammaja (2013) da sauransu. Amma ban da su akwai wasu da dama waɗanda suka yi amfani da karin magana domin nazarin falsafar al‟umma ko kuma wani abu na al‟adunta. Daga cikinsu akwai KirkGreene (1973) da Amin (2002) da Bugaje (2014) da kuma Shede (2014). Tasirin Zamani a Karin Magana. Akwai ayyuka daidai gwargwado da suka shafi nazarin sababbin kare-karen magana. Alal misali, aikin Bugaje (2017) ya nazarci kare-karen magana na zamani ne musamman waɗanda ake amfani da su a kafafen sa da zumunta. Nazarin ya bayyana cewa yawancin sababbin nau‟o‟in karin maganar da ake amfani da su a kafafen sa da zumunta suna ƙunshe da kalmomin aro waɗanda ba a fassara su ko baddala su cikin harshen Hausa ba. Nazarin ya ta‟allaƙa dalilin faruwar hakan da kasancewar matasa ne suka fi mu‟amala da kafafen na sa da zumunta.Domin haka muƙalar ta yi kira ga masana da manazarta da su yunƙura wajen ganin an ƙirƙiro makwafinsu a Hausa. Daga cikin misalan irie-iren waxannan karin maganar da aka kawo akwai: - An yi ba a yi ba, an bar facebook an koma WhatsApp - Wanda ya riga ka log-in zai riga ka log-out,da sauransu. Aikin Mohammed (2018) shi ma ya taɓo kirari da kuma karin magana da suka riƙa yawo a kafar sada zumunta ta Was‟af a shekara ta 2017. Daga cikin misalan karin magana da aka kawo akwai: „Haƙuri dole, wai ɗan Shi‟a ya ga Buhari a Villa. Wannan karin magana ta samu ne bayan da aka samu rashin jituwa tsakanin ƙungiyar Shi‟a da gwamnatin Buhari bayan hatsaniyar „yan ƙungiyar da sojoji a Zariya. Wannan na daga cikin sababbin kare-karen magana da aka samu a zamanin nan. A ƙarshe, aikin Mohammed (2018) ya jaddada cewa, tabbas harshen Hausa da adabinsa da ma al‟adun Hausawa duk suna tafiya da zamani. Shi ma Yakubu (2019:66) ya tattaro wasu nau‟o‟in karin magana da dama kuma da dukkan alamu, sun taɓo mabambanta zamuna. Ciki kuwa akwai waɗanda suke da tushe iri guda amma masu ciko iri daban-daban dangane da zamanin da ake ciki, misali: Abin nema ya samu, matar falke ta haifi jaki. Abin nema ya samu, an yi wa mayunwaci bisimilla. Abin nema ya samu, matar bakanike ta haifi sifana. Abin nema ya samu, ɗan siyasa ya yi takara ya samu. Idan aka lura da kare-karen maganar da aka kawo a sama, za a ga cewa tushensu iri guda ne amma kuma cikonsu ya bambanta. Kowanne daga cikin cikon yana nuni da wani lokaci na musamman a tarihin ƙasar Hausa. Ciko na farko yana nuni da lokacin gargajiyar Bahaushe, wato lokacin da ake safara da jaki da sauran hanyoyin sufuri na gargajiya. A karin magana ta biyu kuwa, cikon yana nuni da cewa Bahaushe ya karɓi addinin Musulunci har ya san cewa idan za a ci abinci, akan fara da bisimilla.

A karin magana ta uku kuma, har Turawa sun shigo ƙasar Hausa an sami motoci da sauran abubuwan hawa waɗanda ake amfani da sifana wajen kwance su da kuma ɗaure su. A karin magana ta ƙarshe kuma, har tafiya irin ta zamani ta fara nisa, an shigo zamanin siyasa wadda a cikinta, takara ake yi, wanda ya yi nasara, shi zai yi mulki.

Kare-karen Maganar da Suka Bijiro a Sakamakon Auren Yusuf Buhari.Wannan ɓangare na wannan muƙala zai karkata ne a kan kawo tare da yin fashin baƙi a kan karekaren maganganun da suka cika kafofin sada zumunta na zamani musamman ma Was‟af da na Fesbuk a kwanakin da aka ɗaura auren Yusuf Buhari, a cikin watan Agusta 2021. Waɗannan karekaren maganganu wasu daga cikin su sababbi ne fil, wasu kuma sun tasirantu da zamani ne kawai amma tushensu daɗaɗɗe ne. Amma kafin a kawo kare-karen maganar da kuma bayanai a kansu, ya kamata a tuna cewa shi dai Yusuf Buhari ɗa ne ga Shugaban Ƙasa Muhammadu Buhari kuma ya auri ɗiyar Sarkin Bichi ne na yanzu. Domin haka bukin ya tara manya masu-faɗa-a-ji daga ko‟ina a cikin faɗin ƙasar nan da ma wasu manyan baƙi daga qasashen waje. Ga dai karin maganganun an kawo su ɗaya bayan ɗaya tare da sharhi a kan kowannensu.Da yake bukin ɗaurin auren Yusuf Buhari na manya ne (daga gwamnoni sai ministoci sai „yan majalisa da sauran manyan „yan siyasa sannan sai kuma sarakunan gargajiya), idan aka ga talaka a wajen akan ɗauka cewa ba gayyatarsa aka yi ba, ya zo ne kawai ba tare da gayyata ba. Toda yake shi talaka ne kuma ana ɗaura auren ɗan shugaban ƙasa da ɗiyar Sarki mai daraja ta ɗaya, to bai isa a gayyace shi ba. Domin haka wannan wuri ba muhallinsa ba ne, kamar yadda masu iya magana kan ce, “Qwarya ta bi qwarya…” Wannan ya ƙara fitowa ƙarara a cikin takwarar wannan karin maganar inda ake cewa: Kai a su wa, talaka a bukin Yusuf Buhari?Wannan karin maganar ta kasance tana sassauyawa daga wannan zamanin zuwa wancan. An faro ne da cewa, „Kai a su wa, kare da gudun layya?“. Daga baya da Turawa suka shigo ƙasar Hausa sai aka sauya ɓangarenta na biyu ta koma Kai a su wa, ɗanwake a otal?. A duka biyun, abin da ake ƙoƙarin bayyanawa shi ne, gazawa ta wanda ake zance da shi. Domin haka ita ma wannan karin maganar irin wannan saƙon ne ta ƙunsa. Akwai kuma wata takwararta wannan karin maganar da aka riƙa amfani da ita a shafukan fesbuk da was‟af bayan auren Yusuf Buhari inda ake cewa:Banza a banza, talaka a xaurin auren Yusuf Buhari. Wannan karin maganar ma takwarar waɗanda suka gabace ta ne ta fuskar ma‟ana. Abin nufi dai shi ne, shi talaka da shi da banza duk ɗaya suke a wajen bukin domin ba ta shi ake yi ba. Hasali ma, ba ya cikin waɗanda aka gayyata. Shi gayyar soxi ne ! A gaba kuma za mu ga wata karin maganar sabuwa fil wadda ta riƙa yawo a kafofin sada zumunta cikin kwanakin da aka yi bukin Yusuf Buhari inda ake cewa:Ashe da rabon mu gana? Talaka

Buhari a matsayinsa na shugaban ƙasa ba kowa zai samu damar ganinsa ba cikin sauƙi ballantana ma talaka. Da yawa talakawa rabon su da ganinsa tun kafin ya zama shugaban ƙasa, lokacin da yake yawan neman a zave shi. A wannan lokacin sai ga Buhari cikin sauƙi,a Bichi ya zo xaurin auren xansa. Wannan ya zamo wani abu ba sabon ba ga mutanen Bichi musamman ma talakawa. A gaba kuma sai wata karin maganar mai cewa :Mulki da sarauta, auren Yusuf Buhari. Wannan karin maganar na bayyana matsayin shi kansa Yusuf Buhari da kuma na waɗanda suka halarci bukin ɗaurin auren nasa. A ɓangaren shi Yusuf Buhari, da farko dai ɗa ne ga shugaban ƙasar Nijeriya, domin haka yana taƙama da mulki. Na biyu kuma ana saura kwanaki kaɗan a ɗaura masa aure aka ba shi sarauta a Daura. Sannan kuma ya auri „yar Sarki sukutum kuma jikar Sarki ! Saboda 4 haka aure ne na taƙama da sarauta da mulki. A ɓangaren mahalarta taron ɗaurin auren kuwa, yawancinsu ko dai masu muƙamai na siyasa ne (mulki) ko kuma na sarautar gargajiya (sarauta). Domin haka da wanne za a ji, ko kuma wanne ne ba wanne ba? Ga mulki, ga sarauta. Wannan ne ma ya sa aka riƙa raba kyaututtuka ga waɗanda suka halarci bukin wanda har hakan ya haifar da wata karin maganar da ke cewa: Abun rabo ne, samun kyautar wayar hannu a bukin ɗan Buhari.Abu ne sababbe musamman a wannan zamani a rarraba tsaraba ga mahalarta bukin ɗaurin aure ko suna (har ma da mutuwa), domin su riƙa tunawa da wannan bukin duk lokacin da suka zo amfani da wannan abu. Wasu kan raba jakunkuna, wasu kofuna da sauran makamantansu. Ya danganta ga ƙarfin hali da zuciyar mai buki, kamar yadda masu iya magana suka ce, „Fadawa mai zuciya buki ba mai dukiya ba“. A bukin Yusuf Buhari, wayar hannu mai tsada wadda ake ce wa iphone aka riƙa rabawa. Ba a ce shi ko iyayensa ne suka raba wayoyin ba, amma „yan“uwa da abokan arziƙi da sauran masu neman gyara miyarsu su suka riƙa yin hakan domin gwanintarsu ta fito fili. Ban da kyauta ga mahalarta bukin, ita ma amarya an yi mata gara sha tara ta arziki na kayan da za a kai ta gidan miji da shi. Wannan shi ma ya haifar da karin magana mai cewa :Nan gani nan bari, kwastam sun ga foreign rice a garar ɗan Buhari.Lokacin da Shugaba Buhari ya zama shugaban ƙasa, ya rufe hanyoyin shigo da abinci daga ƙasashen waje domin ƙarfafa wa „yan ƙasa guiwa su noma abinci isasshe. Wannan ya sa idan jami“an kwastam suka kama mutum da shinkafar waje za su yi masa hukunci mai tsanani. Amma a cikin kayan gara da aka kai gidan Yusuf Buhari, akwai shinkafa „yar waje. Wannan ya sa aka yi ta maganganu wanda har daga ƙarshe aka samar da wannan karin maganar. Ba hana shigo da abinci ne kaɗai sauyin da gwamnatin Buhari ta kawo a Nijeriya ba, akwai wasu sauye-sauyen kamar tsuke bakin lalitar gwamnati da hana almubazzaranci da kuɗaɗen gwamnati waɗanda suka haifar da ƙarancin kuɗi a hannun jama“a. Wannan shi ya haifar da karin magana mai cewa:Da ma da kuɗi?“

Wai talaka ya ga ana rabon iphone a bukin Yusuf Buhari.Raba wayoyin iphone da aka yi a bukin Yusuf Buhari ya jawo hankalin jama‟‟a matuƙa saboda abin ya zo ne a lokacin da ake kukan rashin kuɗi sakamakon koma-bayan tattalin arziƙi. „Yan adawa sai suka samu hanyar sukar gwamnati mai ci domin su gamsar da talakawa game da abubuwan da suke zargin gwamnati da su. Amma su talakawan sun gamsu da wannan kuwa, musamman ta la‟‟akari da karin maganar nan da aka samar duk a dalilin auren na Yusuf Buharin? Karin maganar cewa take:Ashe haka suke? Deliget (delegate)ya ga „yan siyasa a Bichi Takwararta kuma cewa take:Ashe haka suke? Talaka ya ga manyan „yan APC da PDP na tafa hannu a bukin Yusuf Buhari.Deliget wakilai ne na jama‟‟a waɗanda suke zaɓen ɗan takarar da zai tsaya a kowace jam‟‟iyya. „Yan siyasa na matuƙar ji da su a lokacin zaɓen fitar da gwani amma da zarar sun ci zaɓe, sai su yi watsi da su. Su ma talakawa hakan take kasancewa da su lokaci da bayan babban zaɓe. Lokacin da „yan takara ke fafatawa, za ka yi tsammanin cewa idan suka haɗu kaurewa da faɗa za su yi. Wannan ne 5 ya sa da deliget da kuma talakawa suka ga „yan siyasa suna tafawa a bukin Yusuf Buhari, babu bambancin jam‟‟iyya sai abin ya yake ba su al‟‟ajabi. Wannan abu sai ya bayar da mamaki matuqa, saboda yadda talakawa ke gaba saboda su. Amma fa akwai wasu kaɗan cikin „yan siyasar da ba su zo bukin ba. Wannan ma ya haifar da wata karin magana mai cewa : Da sauran manya, Ba a ga Kwankwaso a Bichi ba A cikin manyan „yan siyasa a Nijeriya ta yau, Injiniya Rabiu Musa Kwankwaso sananne ne. Musamman rawar da ya taka lokacin da yake Gwamnan Kano, ga shi kuma ana babban buki a cikin jahar. Daxin daxawa; yanzu yana takarar shugaban qasa.Rashin ganinsa a Bichi duk da kasancewar garin a jiharsa ta Kano ya jawo ka-ce-na-ce a tsakanin jama‟‟a. Ana ganin wannan hujja ce mai tabbatar da raɗe-raɗin cewa Kwankwanson ba ya jituwa da Buhari. Kasancewar duk wanda ya je bukin ya ci, ya sha yadda yakamata, shi ya sa aka samar da karin magana mai cewa:Ma sha Allah!‟‟ Wai ustaz ya kai loma a dinar ɗan Buhari.Da yake an ci an sha kowa ya ƙoshi kuma abinci ne haɗaɗɗe dahuwar manya, jama‟‟a sai suka ƙirƙiri wannan karin maganar domin tunawa da wannan al‟‟amari. To amma me ya sa aka danganta batun da ustaz? Wannan ba zai rasa nasaba da kalmar „Ma Sha Allah‟‟ da aka yi amfani da ita ba, domin yawancin Hausawa musulmi ne kuma shi musulmi idan abu ya yi masa daɗi sai ya ce, „Ma Sha Allah‟‟. To amma da yake ba kowa yake tunawa ya faɗi hakan a kowane lokaci ba, sannan kuma idan mutum ya cika ambaton irin waɗannan kalmomin sai ka ji ana zolayarsa ta hanyar kiran sa da suna „ustaz’. Wannan shi ya sa aka yi amfani da ita kalmar a nan. Ba iyakar kare-karen maganar da wannan buki na Yusuf Buhari ya haifar ba ke nan, domin akwai wata karin maganar da ke cewa:Bana kowa ya ji daɗin buki bai kai na Bichi ba.

Lallai kam, tun da an ci an sha kuma an yi buki lafiya an tashi lafiya ga kuma tsaraba har da wayoyin iphone, ai wannan batu babu tantama. Ga kuma halartar manyan mutane na qasa da baqi. Ai babu wani jin dadi da ya wuce wadannan da aka ambata sannan kuma ba kowane buki zai dace da hada dukkan wadannan abubuwa ba. To amma rashin zuwan bukin nan laifi ne, musamman ma ga maanya ? Bari mu ji abin da karin magana ta gaba ke cewa. Daga ban je Bichi ba… Cire ministan noma da damina. Daya daga cikin manyan jami"an gwamnati da ba su samu zuwa wajen daurin auren Yusuf Buhari ba shi ne ministan aikin gona,wato Alhaji Muhammad Sabo Nanono. Kwanaki kadan bayan auren sai aka cire shi daga kan mukaminsa na ministа saboda wasu dalilai. Kasancewar bai je daurin auren ba duk da cewa a jiharsa aka yi, shi ya sa wasu suka rika danganta hakan da cire shi a mukamin. Tun a baya an riga an fada cewa bukin Yusuf Buhari fa na manya ne. Wannan ya sake haifar da wata karin maganar mai cewa: Wa ya aike ka? An kade xan talaka a Bichi.Abin da ake fada na kadewa ba lallai ba ne ya faru da gaske, amma an yi amfani da shi ne a babin kaddarawa. Da yake idan ka ga talaka a wajen to shi ya kai kansa, ba gayyatarsa aka yi ba, shi ya sa aka kawo zancen cewa ko an kade shi, shi ya jawowa kansa. Watakila kwadayin ciye-ciye da tandetande da tsaraba ya kai shi, kamar yadda direban abokan ango ya more inda har zancensa ya haifar da karin magana mai cewa:Allah ya maimaita, inji direban abokan Yusuf Buhari. A bayanan da aka rika yayatawa, direban da ya dauko abokanan ango (Yusuf Buhari) daga filin jirgin sama zuwa Bichi, sannan kuma ya mayar da su filin jirgin bayan daurin auren ya sha kyaututtuka. An ce sai da suka hada masa fiye naira dubu dari uku (N300,000). Wannan ya sa aka kirkiri wannan karin maganar aka kuma dangata ta da zancensa na fatar maimaituwar irin wannan bukin. Ita ma garar da aka kai amarya da ita ta jawo hankalin duk wanda ya gani kamar yadda aka kirkiri wannan karin maganar mai cewa: Shegiya aboki! Inyamuri ya ga garar Yusuf Buhari . Bahaushe yana kiran mutumin Ibo da Inyamuri saboda wani dalili na tarihi. An ce lokacin da mutanen Ibo suka zo Arewa ba su iya Hausa ba, sai yarensu. A yaren nasu suna kiran ruwa da muri.Da kishi ya kama wani daga cikinsu sai ya rika cewa, inyamuri wato zan sha ruwa. To daga nan aka lakaba musu wannan sunan. Su kuma sukan kira Bahaushe da suna aboki, musamman Hausawanmu da suke zaune da su a can Kudu. A al"adar al"ummar Ibo babu wani abu wai shi gara. Domin haka da mutumin Ibo ya ga tirela cike da kayan abinci da sunan gara, wai sai ya ce shegiya aboki. Shi ma yawan sadakin da aka biya ya haddasa samuwar wata karin magana mai cewa:

Dalar Amurka an ji jiki, a bukin Yusuf Buhari.An ce kuɗaɗen ƙasashen waje musamman ma daloli su aka riƙa liƙe da su a bukin Yusuf Buhari. An yi rawa a kan dala ana tattaka ta. Wannan ne ya sa aka ce dalar ta sha wuya. Da ma a baya an ambaci cewa a Bichi aka ɗaura auren na Yusuf Buhari. Wannan ya sa aka samar da karin magana mai cewa:Bana Bichi an shiga tarihi…bukin Yusuf Buhari.Shi dai Bichi ba wani gari ba ne babba sosai, domin haka wasu manyan ma ba su taɓa zuwa garin ba sai ranar ɗaurin auren. Amma tun daga wannan ranar garin ya shiga tarihi saboda irin manyan mutanenda suka taru a wajen. Irin shinkafar da aka dafa a bukin ma ya sa an samar da wata karin maganar mai cewa:

Ashe da rabonmu, talaka ya kai lomar shinkafar waje a bukin xan Buhari

Da kuma tawararta mai cewa:Jiya ba yau ba, talaka ya kai lomar shinkafar waje a bukin xan Buhari.A baya an yi bayanin cewa gwamnatin Buhari ta hana shigowa da shinkafa „yar waje, amma sai ga shi ita aka dafa a bukin Yusuf. Kasancewar talaka ya daɗe bai ci irin wannan shinkafar ba, shi ya haifar da waɗannan tagwayen kare-karen magana da aka ambata a sama. Haɗuwar masu mulki na gargajiya da „yan siyasa ma akwai abin da ya haska wa jama‟a masu lura. Wannan ya sa aka samar da wata karin maganar mai cewa:

Kanku ɗaya! Talaka ya ga manyan yan siyasa da masu mulki a Bichi Akwai lokuta da dama da za ka ji maganganu daga ɓangarori daban-daban a tsakanin masu mulki da kuma „yan siyasa sai ka zaci kamar ba sa ga-maciji-da-juna. Amma haɗuwarsu a Bichi ya gwada wa talaka cewa kansu a haɗe yake, savanin magoya bayansu da ke gaba da junansu babu gaira babu dalili. Su ma ƙawayen amarya ba a bar su a baya ba domin zancensu ya haifar da wata karin magana mai cewa: Mun gwangwaje! Inji qawayen amaryar Yusuf Buhari a Bila.Da yawa daga cikin jama‟ar Nijeriya ciki har da ƙawayen amaryar Yusuf Buhari ba su taɓa shiga fadar shugaban ƙasa (Villa) ba. Amma a dalilin auren Yusuf Buhari sai ga ƙawayen amarya a can, aka shiga aka kashe ƙwarƙwatar ido, baya ga ladar ƙawance wanda ba a rasa ba. Kammalawa.Wannan muƙalar ta nazarci karin maganganu har ashirin (20) waɗanda suka bijiro kuma suka yi ta yawo a kafofin sa da zumunta a sakamakon bukin auren Yusuf Buhari wanda aka yi a cikin watan Agustan shekarar 2021. Wani abin lura da kuma ban sha‟awa game da waxannan karin maganganu shi ne, samuwarsu na ƙara jaddada cewa harshen Hausa harshe ne mai ci gaba da bunƙasa. Sannan kuma harshe ne mai tafiya da zamani saboda yadda yake ci gaba da samun karɓuwa ta hanyar yawan masu amfani da shi a kafofin sa da zumunta na zamani. Waɗannan karin maganganu guda ashirin da aka nazarta an tsakuro su ne kawai don wannan nazari ba wai su kaɗai ke nan ba. Ta iya yiwuwa akwai wasu da yawa da nazarin nan bai ci karo da su ba.

Haka kuma harshe ne muhimmin abu ga al‘‘umma, domin da shi ake gudanar da harkokin da suka shafi al‘‘ada. Sai an yi amfani da harshe za a gudanar da ciniki. Ilimi ma sai an yi amfani da harshe yake samuwa. Hatta addini ma sai an yi amfani da harshe (Ɗantumbushi, 2008:20). Ma'ana harshe yana da dangantaka da duk al'amuran rayuwar al'umma. Wajen aiwatar da al'adu Kamar na haihuwa da reno da aure da sauransu, duka ana amfani da harshe wajen aiwatar da su. Haka ma a wajen ciniki ana amfani da harshe wajen isar da saƙo tsakanin mai saya da mai sayarwa.Sannan harkokin ilimi ma ana amfani da harshe wajen aiwatar da su.Misali, tsakanin mai koyo da mai koyarwa. A ɓangaren addini ma harshe yana taka rawa. Misali, wajen aiwatar da addu'o'i da karatukan da suka shafi addini, da harshe ake gudanar da su. Don haka, ashe harshe yana da matuƙar muhimmanci a duk al'amuran rayuwar al"umma. Siyasa da harshe suna da dangantaka mai ƙarfi. Dalili kuwa, babu wani tsarin siyasa ko mulki, kowane iri, wanda ba ya amfani da hanyar sadarwa tabaka ko rubutu wajen jawo hankalin jama‘‘a ko fahimtar dasu wani abu. Saboda haka, matsayin harshe a harkokin mulki da na siyasa muhimmai ne ƙwarai da gaske (Yakasai, 2012:51).Wato akwai dangantaka mai ƙarfi tsakanin harshe da siyasa.'Yan siyasa suna amfani da harshensu wajen bayyana ra'ayoynsu ga al'umma har su fahimce su. Wannan takarda an shirya ta ne da nufin nazarin jumloli masu harshen damo a cikin kalaman 'yan siyasa na jihar Kano.Waɗannan 'yan siyasa suna taka muhimmiyar rawa wajen bunƙasa harshen Hausa. An yiamfani da wasushirye-shiryengidanrediyo da ake aiwatarwa a kafafen yaɗa labarai na jihar Kano. A cikin zantukansu da suke amfani da su wajen bayyana ra'ayoynsu sukan yi amfani da harshen damo.Masana Ilmin Kimiyyar Harshe sun bayyana ra'ayoyinsu game da ma'anar jumla a nahawu. Wato waɗannan masa sun kawo ma'nonin jumla a fahimtarsu. Wasu daga cikinsu su ne kamar haka: Ita jumla Balarabiyar kalma ce aka Hausantar da ita. Ma'anarta ɗaya ce a dukkan harsunan guda biyu; tana nufin jeren kalmomi waɗanda suke bayar da cikakkiyar ma'ana ga mai sauraro (Jinju, 1981)". Abin nufi a nan shi ne, ita kalmar "Jumla" an aro ta ne daga harshen Larabci, wato "alJumlah" Wadda take nufin tarin kalmomi masu ɗauke da ma'ana a cikin magana. Hausawa ma da suka ari wannan kalma ba su sauya mata ma'ana ba. Wato tana nan a ma'anarta ta Larabci. Yahaya, da wasu (1992 :19), sun bayyana ma'anar jumla da cewa, "Jumla yanki ce na magana mafi tsayi. Ma'ana a duk yankuna na magana, Kamar gaɓa, ɗafi ko kalma duk jumla tafi tsawo". Duba da wannan ma'ana, za mu fahimci cewa ita jumla ta ƙunshi duk wani nau'i na magana mafi tsawo Wanda ya ƙunshi gaɓa ko ɗafi ko kuma kalma da za su ba da ma'ana a cikin zance.

Zarruƙ, (2001 :31) cewa ya yi, "Magana ce wadatacciya wadda ba ta buƙatar ciko ko ƙari". Abin nufi a nan shi ne, duk wata magana cikakkiya da za ta ba da ma'ana ita ce jumla. Idan ba ta ba da cikakkiyar ma'ana ba, to ba za a kirawo ta jumla a tsarin nahawu ba. Don haka, idan muka duba wannan ma'ana ta jumla, za mu fahimci cewa jumla ita ce duk wani zance da zai ba da cikakkiyar ma'ana a harshe. Ita jumla a Hausa iri-iri ce, ya danganta da yadda tazo a cikin zance. Don haka, masana Ilimin Kimiyyar Harshe sun karkasa ta Kamar haka: akwai sassauƙar jumla da harɗaɗɗiyar jumla da sarƙaƙƙiyar jumla da sauransu. A ma'ana ta asali harshe na nufin " tsokar da ke lallagi cikin baki, wadda ake lasar abu da ita, kuma idan babu ita ba a iya magana (CNHN, 2006)" Haka ma damo na nufin "wata ƙaramar dabbar daji mai siffar kada ko guza, mai zama a rami, yana da yawan haƙuri (CNHN, 2006)" Haka kuma wannan dabba tana da harshe guda biyu. Ke nan harshen damo na nufin harshen wata ƙaramar dabbar daji mai siffar kada ko guza, mai zama a rami, yana da yawan haƙuri kuma mai harasa biyu. A ilmance, harshen damo na nufin samuwar ma'anoni fiye da guda ɗaya. Kalma ko bayani wanda za a fahimta ta fuskoki fiye da guda ɗaya (Odford, 2011). Haka kuma, masana irin su Wurma (2008), sun yi bayanin kalmomi ɗai-ɗai ma'anoni da yawa. Wato kalmomi masu harshen damo. Jumloli masu harshen damo sun ƙunshi jumloli waɗanda suke ɗauke da ma'anoni fiye da guda ɗaya (Bello, 2014).Kasancewar akwai nau'o'in jumloli da ake li'irabinsu amma masu harshen damo ba a sami hanya mai sauƙi ta li'irabinsu ba, sai daga baya wani masanin Kimiyyar harshe mai suna Noam Chomsky ya samar da hanya sassauƙa ta li'irabin wannan jumla mai harshen damo (Chomsky, 1962).Wannan takarda ta ƙunshi irin gudummawar a 'yan siyasa suke bayarwa wajen bunƙasar harshen Hausa, ta hanyar amfani da hikimominsu wajen sarrafa harshensu na siyasa. An kalli Jumloli masu harshen damo a cikin kalamansu. A wannan aiki an bayyana ma'anar jumla da ire-irenta. An yi cikakken bayanin harshen damo a cikin jumla. Wato samuwar ma'anoni fiye da guda ɗaya a cikin jumla. Sannan an kawo jumloli masu harshen damo da 'yan siyasa suke amfani da su a cikin kalamansu na siyasa, tare da bayanansu. Don haka, 'yan siyasa ba a bar su a baya ba suna ba da gudummawar wajen bunƙasar harshen Hausa.