

UNIVERSITÉ DU QUÉBEC

MÉMOIRE PRÉSENTÉ À  
L'UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES

COMME EXIGENCE PARTIELLE  
DE LA MAÎTRISE EN MATHÉMATIQUES ET INFORMATIQUE  
APPLIQUÉES

PAR  
ACHOURI ABDELGHANI

EXTRACTION DE RELATIONS D'ASSOCIATIONS MAXIMALES DANS LES  
TEXTES : REPRÉSENTATION GRAPHIQUE

OCTOBRE 2012

Université du Québec à Trois-Rivières

Service de la bibliothèque

Avertissement

L'auteur de ce mémoire ou de cette thèse a autorisé l'Université du Québec à Trois-Rivières à diffuser, à des fins non lucratives, une copie de son mémoire ou de sa thèse.

Cette diffusion n'entraîne pas une renonciation de la part de l'auteur à ses droits de propriété intellectuelle, incluant le droit d'auteur, sur ce mémoire ou cette thèse. Notamment, la reproduction ou la publication de la totalité ou d'une partie importante de ce mémoire ou de cette thèse requiert son autorisation.

## Résumé

La croissance exponentielle des documents électroniques disponibles sur le web, suscite le besoin de nouvelles approches d'aide automatique de lecture, d'analyse et de compréhension de textes.

Les règles d'association maximales ont montré leur utilité dans le domaine de l'extraction de connaissances à partir de données (ECD).

En pratique, les algorithmes d'extractions de règles d'associations génèrent un nombre très élevé de règles, ce qui rend difficile à l'utilisateur de l'explorer.

La représentation graphique des règles d'associations s'est avérée une source d'amélioration de la performance du processus d'extraction de connaissances pertinentes à partir de grands volumes de données.

À cet effet, nous avons développé un outil de visualisation graphique et interactive, permettant à l'utilisateur de diriger sa recherche en visualisant à chaque itération une quantité limitée de règles potentiellement pertinentes.

Les résultats des expérimentations effectuées sur ensemble de corpus réels de taille moyenne (environ 20 pages) ont montré une performance considérable de notre outil dans la lecture et la compréhension de textes.

## **Abstract**

The exponential growth of the electronic documents available on the Web raises the need for new approaches based on an automatic reading, analyzing and understanding of texts.

The maximal association rules have proved their usefulness in the field of data knowledge extraction.

In practice, the association rules extraction algorithms generate a very large number of rules which makes it difficult for the user to explore.

The association rules graphic representation has been an improvement way for a relevant knowledge extraction from a large data volume.

For this reason, we have developed an interactive and visual graphic tool allowing the user to focus the search for the user by visualizing in each iteration a limited quantity of potentially relevant rules.

The results of our tests performed on real corpus set with medium size (about 20 pages) showed a considerable performance of our tool for reading and understanding texts.

## Remerciements

J'adresse tout d'abord, ma profonde reconnaissance et mes vifs remerciements à M. Ismail Biskri pour m'avoir fait l'honneur d'être mon directeur de recherche. Je le remercie pour son soutien, sa disponibilité, sa patience et ses précieux conseils durant toute la période de mon travail.

Je remercie Messieurs les membres du jury Mhamed Mesfioui et François Meunier pour avoir accepté de juger mon travail.

Je désire également remercier mon co-encadreur, M. Boucif Amar Bensaber pour ses encouragements et ses conseils.

Je dédie ce travail à toute ma famille, particulièrement mes chers parents, pour leurs amours et leurs prières, et je leurs suis très reconnaissant pour les sacrifices qu'ils ont du faire pendant mes longues années d'études.

Je remercie ma femme, qui m'a aidée à surmonter les difficultés et d'avoir accepté tant de sacrifices durant ces dernières années, ainsi mes petits enfants MOUADH et RITEJ qui suscitent en moi l'envie de me battre.

Je dédie aussi ce travail à tous les membres de ma belle famille, en particulier à ma belle-mère pour son affection et ses prières pour moi, à mon beau-père pour ses encouragements.

Enfin, j'adresse mes reconnaissances à mes amis, Lamri, Adel, Bilel, Salim, Boualem, pour leurs soutiens.

## Table des matières

Résumé .....	ii
Abstract .....	iii
Remerciment .....	iv
Table des matières.....	v
Liste des tableaux .....	xi
Liste des figures .....	xii
Chapitre 1 - Introduction .....	1
Chapitre 2 - Classification.....	5
2.1 Introduction .....	5
2.2 Type de données .....	5
2.3 Les mesures de Similarité .....	7
2.4 Qualité d'une classification .....	8
2.5 Méthodes de classification .....	9
2.5.1 Classification supervisée.....	11
2.5.1.1 Les k plus proches voisins ( $k$ -PPV) .....	11
2.5.1.2 Les Séparateurs à Vaste Marge .....	14
2.5.1.3 Bayes naïf.....	15

2.5.2	Classification non supervisée.....	16
2.5.2.1	Méthodes de classification hiérarchique .....	16
2.5.2.1.1	Classification Ascendante Hiérarchique CAH.....	17
2.5.2.1.2	Classification Descendante Hiérarchique CDH.....	17
2.5.2.2	Méthodes de classification par partitionnement.....	19
2.5.2.2.1	Méthode K-means .....	19
2.5.2.2.2	Méthodes des nuées dynamiques .....	21
2.6	Conclusion.....	24
Chapitre 3 - Règles d'association.....		25
3.1	Introduction .....	25
3.2	Définitions et notations.....	26
3.2.1	Base de données binaire.....	26
3.2.2	Itemset.....	27
3.2.3	Règle d'association .....	27
3.2.4	Support d'un itemset X .....	28
3.2.5	Support d'une règle d'association.....	28
3.2.6	Confiance .....	28
3.3	Étapes de l'extraction des règles d'association .....	29

3.3.1	Sélection et préparation des données .....	30
3.3.2	Découverte des itemsets fréquents .....	30
3.3.3	Génération des règles d'association.....	31
3.3.3.1	Support minimum (minsup) .....	32
3.3.3.2	Confiance minimum (minconf).....	32
3.3.4	Visualisation et interprétation des résultats .....	32
3.4	Algorithme de génération de règles d'association : Algorithme <i>Apriori</i> .....	32
3.5	Conclusion.....	36
Chapitre 4 - Règles d'association maximales .....		37
4.1	Introduction .....	37
4.2	Taxonomie et catégorie .....	39
4.3	Règles d'association maximale .....	40
4.3.1	M-Support d'un litemset.....	41
4.3.2	M-Support d'une règle d'association maximale .....	41
4.3.3	M-Confiance d'une règle d'association maximale .....	42
4.3.4	M-Support minimum et M-confiance minimum.....	43
4.3.5	Itemset maximal fréquent .....	43
4.3.6	M-Facteur d'une règle d'association maximale.....	44
4.4	Algorithme de calcul des règles d'association maximales.....	45



4.5 Conclusion.....	50
Chapitre 5 – Visualisation des règles d’association.....	51
5.1 Introduction .....	51
5.2 Visualisation de règles sous forme de listes textuelles.....	51
5.3 Visualisation par graphe .....	52
5.3.1 Graphe orienté.....	52
5.3.2 Graphe non orienté.....	53
5.4 Visualisation par matrice .....	56
5.5 Visualisation en mosaïque .....	60
5.6 Visualisation par métaphore .....	61
5.7 Visualisation avec les coordonnées parallèles.....	62
5.8 Conclusion.....	63
Chapitre 6 - Implémentation .....	64
6.1 Introduction .....	64
6.2 Choix de langage .....	65
6.3 Architecture du système .....	67
6.3.1 Analyse du texte .....	68
6.3.2 Classification numérique .....	70
6.3.3 Extraction des règles d’association maximales .....	73

6.3.4	Affichage graphique des résultats .....	75
6.3.4.1	Description de fonctionnement .....	75
6.3.4.2	Acquisition des données .....	75
6.3.4.3	Génération des regles d'association maximales.....	77
6.3.4.3.1	Représentation textuelle.....	78
6.3.4.3.2	Représentation Graphique.....	78
6.3.4.4	Approche dynamique et interactive.....	80
6.3.4.5	Historique d'exploration .....	81
6.3.4.6	Annulation et restauration .....	82
6.3.4.7	Recherche des segments .....	83
6.4	Conclusion.....	84
Chapitre 7 – Expérimentation et résultat.....		85
7.1	Exemple 1 : « 2011 : le printemps arabe ? ».....	86
7.2	Exemple 2 : « 8 MAI 1945 – MAI 1985, 40ème anniversaire » .....	96
7.3	Conclusion.....	105
Chapitre 8 – Conclusion.....		106
Bibliographie.....		108
Annexe A – Liste des segments du texte « 2011 : LE PRINTEMPS ARABE ?»		
	contenant la séquence des mots sélectionnés .....	113

Annexe A – Liste des segments du texte « 8 MAI 1945 – MAI 1985, 40ème  
anniversaire » contenant la séquence des mots sélectionnés ..... 124

## Liste des tableaux

Tableau 3-1	Exemple de base de données transactionnelles T .....	27
Tableau 3-2	Représentation binaire de T.....	27
Tableau 4-1	Les données de transactions ... ..	39
Tableau 5-1	Ensemble d'itemsets fréquents.....	54
Tableau 5-2	Ensemble de règles d'associations... ..	54
Tableau 6-1	Matrice contient les fréquences des N-grams dans les segments... ..	71
Tableau 7-1	Cohérence de résultat obtenu pour la prémisse " PRINTEMPS RABE" avec le texte original ... ..	88
Tableau 7-2	Liste des segments contenant les règles : $PRINTEMPS\ ARABE \xRightarrow{max} PEUPLE$ et $PRINTEMPS\ ARABE \xRightarrow{max} IMPÉRIALISME$ .....	91
Tableau 7-4	Cohérence de résultat obtenu pour la prémisse " MAI " par apport au texte original ... ..	98
Tableau 7-5	Liste des segments contenant les règles : $MAI \xRightarrow{max} NATIONAL$ et $MAI \xRightarrow{max} PEUPLE$ .....	101
Tableau 7-6	Extraction du sens général des segments .....	102

## Liste des figures

Figure 2.1	Structuration des données.....	6
Figure 2.2	Organisation des méthodes de classification .....	10
Figure 2.3	Exemple de K-PPV.....	12
Figure 2.4	Apprentissage de classificateur SVM.....	14
Figure 2.5	Exemple d'arbre hiérarchique .....	17
Figure 3.1	Etapas du processus d'extraction de règles d'association .....	29
Figure 3.2	Diagramme de Hasse représentant le treillis des itemsets obtenus a partir des données du tableau 3.1 .....	31
Figure 3.3	Algorithme Apriori (recherche des itemsets fréquents) .....	34
Figure 3.4	Algorithme d'extraction de règles .....	35
Figure 4.1	Algorithme de calcul des règles M-associations .....	46
Figure 4.2	Algorithme de la procédure M-Frequent-Set( $\hat{s}$ ) .....	48
Figure 5.1	Visualisation textuelle de règles d'association.....	52
Figure 5.2	Visualisation de règles d'association par un graphe.....	53
Figure 5.3	Graphe non orienté .....	55
Figure 5.4	Réduction des nœuds et détail d'une règle .....	55
Figure 5.5	Visualisation de règles d'association par graphe.....	56
Figure 5.6	Une matrice 3D itemset-à-itemset .....	57
Figure 5.7	Une matrice D3 Item-à-règle .....	58
Figure 5.8	Une matrice 2D Item-à-règle.....	59
Figure 5.9	Visualisation de règles d'association par matrices .....	59

Figure 5.10	Représentation en mosaïque pour les règles.....	60
Figure 5.11	Visualisation des règles d'association à l'aide d'une métaphore .....	61
Figure 5.12	Visualisation des règles d'association avec des coordonnées parallèles .....	62
Figure 6.1	combinaison de l'algèbre abstraite et du polymorphisme .....	66
Figure 6.2	Architecture du système développé.....	68
Figure 6.3	Partitionnement des segments retenus du corpus en 3 classes .....	72
Figure 6.4	Fenêtre principale du module graphique .....	75
Figure 6.5	Interface permettant de saisir le Mot-X et le nombre de cardinalité .....	76
Figure 6.6	limitation de règles d à afficher par fixation du nombre d'itemsets Y .....	77
Figure 6.7	Limitation des règles à afficher par choix d'intervalle de M- Confiance .....	77
Figure 6.8	Représentation sous forme textuelle.....	78
Figure 6.9	Représentation graphique des associations .....	79
Figure 6.10	Couleur de l'arc selon le M-confiance .....	79
Figure 6.11	Chemin parcouru dans la fouille.....	80
Figure 6.12	Historique d'exploration.....	82
Figure 6.13	sauvegarde et restauration de chemin parcouru.....	83
Figure 6.14	segments contenant <i>Mot-X</i> et <i>Mot-Y</i> .....	83
Figure 6.15	Segments contenant des mots de la séquence des mots sélectionnés.....	84
Figure 7.1	Extraction des règles pertinentes avec " <i>PRINTEMPS ARABE</i> " en prémisses .....	87
Figure 7.2	Extraction des règles pertinentes avec " PEUPLE " en rémisses.....	93
Figure 7.3	chemins parcourus durant le processus de fouille pour le texte «2011 : Printemps arabe ?» .....	94

Figure 7.4	Interprétation subjective du contenu de texte original (2011 : printemps arabe) .....	95
Figure 7.5	Extraction des règles pertinentes avec " MAI " en prémisse.....	97
Figure 7.6	chemins parcourus durant le processus de fouille pour le texte «8 MAI 1945 - MAI 1985, 40ème anniversaire » .....	103
Figure 7.7	Interprétation subjective du texte (8 MAI : 8 MAI 1985, 40ème anniversaire) .....	104

# Chapitre 1 - Introduction

Les usages récents d'Internet ont conduit à une quantité croissante d'informations textuelles disponibles. Cette masse de documents électroniques, en perpétuelle croissance, pose de façon excessive, la question de nos capacités de lecture, d'analyse et de compréhension de textes.

Il devient donc inévitable d'y mettre en place de nouvelles approches d'aide à la lecture et d'analyse de texte assistées par ordinateur.

La combinaison des deux méthodes de fouille de données : la classification textuelle et l'identification des règles d'association qui ont été mises en œuvre dans le système GRAMEXCO (outil développé pour la classification des documents multimédia (Rompré, et al. 2008)) a donné des résultats encourageants dans l'exploration et la découverte des connaissances cachées dans le texte afin d'extraire l'information pertinente. Mais les résultats obtenus sont encore sujets à amélioration, plus précisément, le mode de présentation textuelle qui ne convient pas à la grande quantité de règles souvent générées par les algorithmes d'extraction des règles d'associations.

Afin d'aider l'utilisateur à explorer cette masse d'informations produites par ces derniers, des efforts ont principalement porté sur deux axes de recherches : l'optimisation d'algorithmes d'extraction des règles d'associations et le développement de nouveaux outils de visualisations des règles. Ces efforts ont conduit à l'émergence de nombreux outils



de visualisation [29] [38], mais généralement, les représentations visuelles de règles existantes sont peu interactives.

Notre travail est une suite des travaux débutés par (Rompré, et al. 2009) dans GRAMEXCO. Pour faire face au problème cité ci-dessus, nous avons développé un outil de recherche graphique et interactif de règles d'associations maximales, pour permettre de faciliter la tâche de fouille de l'utilisateur au sein d'une grande quantité de règles, afin d'extraire les connaissances pertinentes qui peuvent aider et faciliter l'analyse et la compréhension d'un corpus peu importe sa langue.

La représentation visuelle des règles est pertinente, mais elle devient difficile à maîtriser si elle représente d'un seul coup l'ensemble des règles retenues pour un corpus donné, cela se traduit par des visualisations très grandes qui ne peuvent pas tenir sur un écran. Pour faire face à cette limitation, nous proposons une approche centrée sur l'utilisateur qui combine directement la visualisation graphique avec l'algorithme d'extraction des règles d'association, lui permettant d'explorer à chaque itération une quantité limitée de règles (les plus pertinentes), et en fonction de son intérêt pour les règles, il dirigera une suite d'explorations visuelles à la recherche de l'information pertinente.

Ce mémoire est constitué de six chapitres. Dans le **chapitre 2**, nous abordons la classification des données, nous commençons par la présentation du contexte de la technique de classification d'une manière générale, puis nous présentons les différents types de données pouvant être soumises à une technique de classification automatique, ainsi que les mesures de ressemblances entre les objets à classer les plus utilisées. Et à la fin du chapitre, nous présentons les différentes techniques de classification existantes dans la

littérature en les divisant en plusieurs familles (supervisé, non supervisé, hiérarchique, par partition...). Pour chaque technique, nous citons les algorithmes les plus utilisés et leurs avantages et inconvénients.

Dans le **chapitre 3**, nous étudions un outil d'extraction des connaissances à partir des données, appelé les associations. Nous commençons par présenter le formalisme permettant de définir les notions de règles d'association et les deux mesures d'intérêt les plus classiques permettant d'évaluer ces règles : le support et la confiance. Puis nous détaillons les étapes principales du processus d'extraction des règles d'associations. À la fin nous présentons l'algorithme de type A priori qui permet de déterminer ces règles.

Dans le **chapitre 4**, nous nous intéressons particulièrement aux règles d'association maximales (type de connaissances sur lesquelles nous nous appuyons dans notre travail). Ce nouvel outil, qui a une valeur particulière dans l'exploration de texte est orienté vers la découverte des associations qui sont souvent perdues lors de l'utilisation des règles d'association régulières. Nous enrichissons la présentation de la description formelle des deux mesures classiques d'évaluation de règles d'association maximales (M-Confiance, M-support et M-facteur) par des exemples concrets pour mieux comprendre le processus d'extraction des règles d'association maximales. Et à la fin nous expliquons en détail le pseudo-code de l'algorithme d'extraction des règles d'associations maximales.

Dans le **chapitre 5**, nous présentons un état de l'art des techniques de visualisation qui s'appliquent aux règles d'association, ainsi l'évaluation des ces techniques. Des logiciels existants pour la visualisation des règles d'associations seront aussi présentés.

Le **chapitre 6** montre précisément l'architecture, et le fonctionnement du système GRAMEXCO.

Nous nous focalisons dans ce chapitre sur les ajouts et les modifications que nous avons apportés dans la seconde version de GRAMEXCO :

- Le développement et l'intégration d'un module de visualisation des règles selon une approche graphique et interactive des règles permettant d'assister l'utilisateur dans son exploration du nombre exorbitant de règles.
- L'optimisation de l'algorithme d'extraction des règles d'associations en termes de temps de calcul et les ressources en mémoire utilisées.
- En plus du classificateur de neurone ART, qui est utilisé dans la version précédente de GRAMEXCO, nous avons ajouté un deuxième classificateur (K-means) le plus utilisé dans le domaine de la classification textuelle.

Dans le **chapitre 7**, nous présentons les résultats des expérimentations menées sur des corpus avec une évaluation de notre système, nous montrons aussi comment notre approche proposée peut être une aide utile dans la lecture et la compréhension de textes.

Enfin, nous concluons ce mémoire par une présentation des perspectives de recherches qu'ouvre notre travail.

## **Chapitre 2 - Classification**

### **2.1. Introduction**

La classification automatique des données a fait preuve de son utilité dans le domaine du forage des données. Elle consiste à regrouper les objets similaires dans la même classe et les objets dissimilaires dans des classes différentes en se basant sur la mesure de ressemblance entre ces objets. En fait, la classification porte à extraire de façon automatique les connaissances cachées dans les données. Autrement dit, la classification permet de déterminer les liens entre les données, dans notre cas de recherche, il s'agit de regrouper les segments similaires du corpus afin d'effectuer l'extraction des règles d'associations maximales. La technique de classification est utilisée dans de nombreux domaines d'applications. Citons par exemple : la recherche d'informations (fouille de données), Analyse spatiale de données, applications aux données du web et analyse d'ADN dans la bio-informatique.

### **2.2. Type de données**

Pour que les données puissent être soumises à une technique de classification, elles doivent être structurées souvent sous forme d'une matrice de  $n$  lignes et  $m$  colonnes, comme illustrées dans la figure 2.1.

		Variables		
		$Y_1$ ...	$Y_j$ ...	$Y_m$
individus	Y \ X			
	$X_1$	$x_{ij}$		
	$\vdots$			
	$X_2$			
	$\vdots$			
$X_n$				

Figure 2.1 Structuration des données [1].

Les lignes représentent les individus (objets, entités, instances, etc.) et les colonnes représentent les variables de ces individus (attributs, caractéristiques, etc.)[1]. Dans notre cas de recherche, la ligne représente le segment et la colonne représente l'occurrence des *N\_Grams* dans ce segment.

Dans les classifications, le type de données traitées est divers (binaires, nominales, ordinaires, continues). Autrement dit, chaque cellule de la matrice peut contenir une valeur unique, un ensemble de valeurs, un intervalle de valeurs ou une distribution sur un ensemble de valeurs [1].

Classiquement, on distingue deux types de variables : les variables *quantitatives* dites aussi *numériques* et les variables *qualitatives* dites aussi *catégorielles*. Les premières prennent des valeurs ordonnées (*binnaire, continue ou discrète*) et qui peuvent être soumises à une opération arithmétique, contrairement pour les variables qualitatives, on ne peut pas appliquer ces opérations (exemple code d'article, pays de fabrication). Les variables

qualitatives peuvent être divisées en deux types : variables *ordinales* (*exemple* : grand, moyen, petit) et variables non ordonnées dites *nominales* (*exemple* : nom du produit, commentaire) [1].

### 2.3. Les mesures de similarité

La notion de ressemblance entre objets est formalisée à l'aide d'une métrique permettant la comparaison entre paires d'objets (mesure de similarité ou de distance). Chaque domaine d'application possédant ses propres mesures de ressemblance selon la nature de données qu'il possède.

Soit  $X$  un ensemble de données défini dans un espace  $R^n$ , une fonction de distance ( $dist$ ) devrait satisfaire les propriétés suivantes [2] :

1.  $\forall X_i, X_j \in X; dist(X_i, X_j) \geq 0$  (Symétrique)
2.  $\forall X_i, X_j \in X; dist(X_i, X_j) = dist(X_j, X_i)$  (Symétrique)
3.  $\forall X_i, X_j \in X; dist(X_i, X_j) = 0 \Rightarrow X_i = X_j$
4.  $\forall X_i, X_j, X_k \in X; d(X_i, X_j) \leq d(X_i, X_k) + d(X_k, X_j)$  (inégalité triangulaire)

Dans notre cas de recherche, chaque segment est représenté par un vecteur, le calcul de la distance entre deux segments revient au calcul de la distance entre les deux vecteurs qui les représentent. Soient deux segments  $S_i$  et  $S_j$  représentés respectivement par leurs vecteurs  $\vec{S}_i = (s_{i1}, s_{i2}, \dots, s_{in})$  et  $\vec{S}_j = (s_{j1}, s_{j2}, \dots, s_{jn})$ .

Différentes mesures peuvent être utilisées pour calculer la distance entre ces deux segments [3][4] :

$$\text{Distance de Manhattan} : \text{dist}(S_i, S_j) = \|\vec{S}_i - \vec{S}_j\|_1 = \sum_{k=1}^n |s_{ik} - s_{jk}| \quad (2.1)$$

$$\bullet \text{ Distance euclidienne} : \text{dist}(S_i, S_j) = \|\vec{S}_i - \vec{S}_j\|_2 = \sqrt{\sum_{k=1}^n (s_{ik} - s_{jk})^2} \quad (2.2)$$

$$\bullet \text{ Distance de Minkowski} : \text{dist}(S_i, S_j) = \|\vec{S}_i - \vec{S}_j\|_p = \sqrt[p]{\sum_{k=1}^n |s_{ik} - s_{jk}|^p} \quad (2.3)$$

Où  $p$  est un entier positif :

- Pour  $p = 1$  : distance de Manhattan;
- Pour  $p = 2$  : distance de euclidienne;
- Pour  $p \rightarrow \infty$  : distance de Chebychev;

$$\bullet \text{ Distance maximum} : \text{dist}(S_i, S_j) = \|\vec{S}_i - \vec{S}_j\|_\infty = \max_{1 \leq k \leq m} |s_{ik} - s_{jk}| \quad (2.4)$$

Autres mesures de similarités utilisées souvent pour les données binaires :

$$\bullet \text{ Coefficient Dice} : \text{Dice}(\vec{S}_i, \vec{S}_j) = \frac{2 \sum_{k=1}^n s_{ik} s_{jk}}{\sum_{k=1}^n s_{ik}^2 + \sum_{k=1}^n s_{jk}^2} \quad (2.5)$$

$$\bullet \text{ Coefficient Jaccard} : \text{Jaccard}(\vec{S}_i, \vec{S}_j) = \frac{\sum_{k=1}^n s_{ik} s_{jk}}{\sum_{k=1}^n s_{ik}^2 + \sum_{k=1}^n s_{jk}^2 - \sum_{k=1}^n s_{ik} s_{jk}} \quad (2.6)$$

- La similarité entre  $\vec{S}_i, \vec{S}_j$  peut se calculer aussi par le cosinus de l'angle formé par les deux vecteurs  $\vec{S}_i, \vec{S}_j$  qui les représentent.

$$\bullet \text{ Cos}(\vec{S}_i, \vec{S}_j) = \frac{\vec{S}_i \cdot \vec{S}_j}{\|\vec{S}_i\| \cdot \|\vec{S}_j\|} = \frac{\sum_{k=1}^n s_{ik} s_{jk}}{\sqrt{\sum_{k=1}^n s_{ik}^2} \cdot \sqrt{\sum_{k=1}^n s_{jk}^2}} \quad (2.7)$$

#### 2.4. Qualité d'une classification

Une bonne classification produira des classes avec une grande similarité à l'intérieur de chaque classe et une petite similarité entre les différentes classes. La qualité dépend à la fois de la mesure de similarité utilisée par la méthode et de son implémentation. [3].

Une bonne méthode de classification permet de [3] :

- Utiliser un nombre minimum de paramètres à fixer au départ de l'algorithme
- Prendre en compte simultanément des variables de natures différentes
- Trouver des classes de formes quelconques (convexe ou non) et de détecter du bruit ou des points isolés.
- S'appliquer à de grande quantité de données et hautement multidimensionnelles dans une durée de traitement raisonnable.
- Ne pas être sensible à l'ordre des observations dans le fichier à analyser.
- Interpréter facilement les résultats.

### **2.5. Méthodes de classification**

Dans la littérature, divers algorithmes de classifications ont été proposés. Ils se diffèrent grâce aux facteurs : la dépendance des paramètres prédéfinis par les utilisateurs ou la connaissance a priori sur les données, le type d'attributs de données qu'ils traitent, la capacité de traiter une grande quantité de données, la complexité de l'algorithme (temps de calcul) [4].



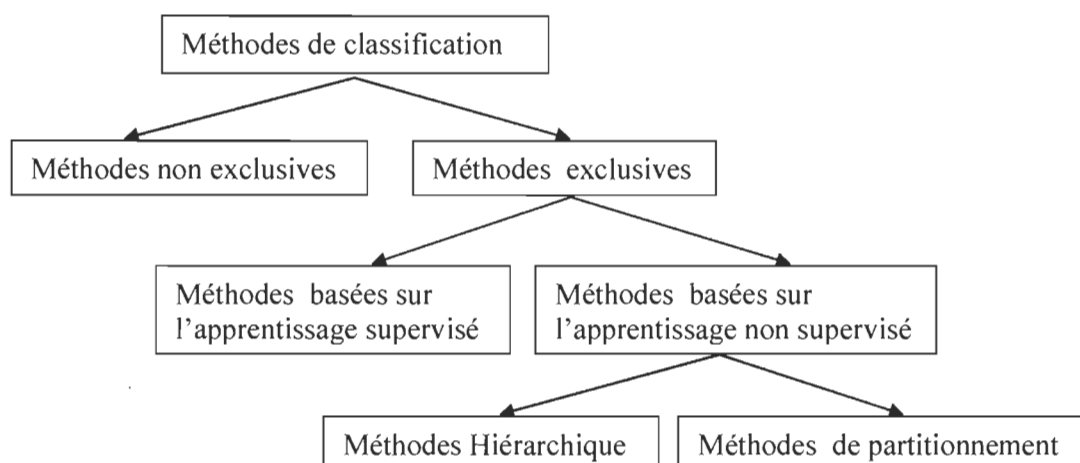


Figure 2.2 Organisation des méthodes de classification [7].

Il est difficile de proposer une subdivision de ces méthodes de classification, chaque étude de synthèse propose sa propre organisation.

La figure 2.2 illustre une organisation présentée par A. Jain et al dans [7], où les méthodes de classification sont divisées en deux groupes : méthodes *exclusives* et méthodes *non exclusives*. Dans le premier groupe, l'objet doit appartenir à une seule classe, par contre dans le deuxième groupe l'objet peut appartenir selon un degré d'appartenance à plusieurs classes en même temps (classification floue).

Les méthodes exclusives peuvent être aussi divisées en deux groupes :

- Méthodes de classification basées sur l'apprentissage supervisé (data mining) : ces méthodes ont besoin de connaître *a priori* les classes associées aux données d'entraînement. L'algorithme de classification affecte un nouvel objet à l'une de ces classes avec une erreur minimum de classement.

- Méthodes de classification basées sur l'apprentissage non supervisé (clustering) : ces méthodes n'ont pas besoin d'un superviseur ni de classes établies *a priori*. L'algorithme de classification regroupe les données en fonction de caractéristiques semblables.

Les méthodes de clustering eux-mêmes sont divisées en deux catégories principales : les *approches hiérarchiques* et les *approches par partitionnement*. Tandis que les premières construisent les classes graduellement sous une forme hiérarchique (un arbre des classes appelé dendrogramme), les deuxièmes consistent à regrouper les données dans un nombre de classes fixé *a priori*. Ces classes sont au même niveau.

### **2.5.1. Classification supervisée**

Nous présentons maintenant quelques exemples classiques de techniques de classification supervisée. Nous commençons d'abord par le classificateur bayésien naïf qui est souvent utilisé comme point de référence à cause de sa simplicité et son ancienneté, puis les machines à support vectoriel et l'algorithme des K-voisins les plus proches (K-NN), qui représentent actuellement les deux meilleurs choix en catégorisation de textes [15] [19] [16].

#### **2.5.1.1. Les k plus proches voisins (k-PPV)**

Plus connus en anglais sous le nom K-nearest neighbor (K-NN), figure parmi les plus simples algorithmes d'apprentissage automatique. Le principe général de la méthode des *K-PPV* est de déterminer pour chaque nouvel objet que l'on veut classer, la liste des *K* plus

proches voisins parmi les objets déjà classés. Puis, l'objet est affecté à la classe majoritaire parmi ces  $k$  objets trouvés. Cette méthode nécessite de choisir le nombre  $K$  et la mesure de similarité afin de comparer un nouvel objet à ceux déjà classés (habituellement, la mesure employée pour l'analyse est la distance euclidienne)[52][53].

Soit l'exemple de la figure 2.3, on doit classer le nouveau segment  $S$  (en rouge). Si on choisit  $K = 1$ ,  $S$  sera classé cercle noir. Si  $K = 6$ , le même  $S$  sera classé cercle blanc. On voit donc que le choix de  $K$  est très important dans le résultat final.

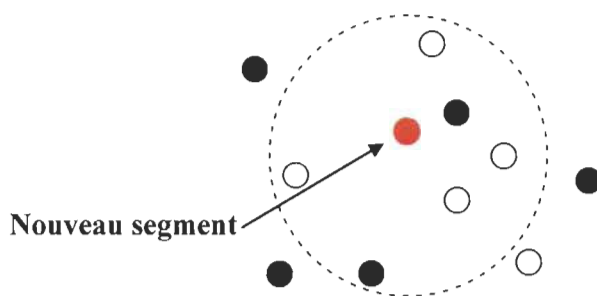


Figure 2.3 Exemple de K-PPV (K-NN).

La méthode de K-NN est souvent performante, son apprentissage est de type par cœur l'un des types les plus simples (il garde tous les exemples d'apprentissage). Cependant, l'inconvénient majeur réside dans le temps investi pour la prédiction, car il nécessite le calcul de la distance avec tous les exemples, bien qu'il existe des heuristiques pour réduire le nombre d'exemples à prendre en compte [15].

**Algorithme des k plus proches voisins [13]**

Soit : E un espace de dimension D

A un ensemble de N points dans cet espace

K un entier plus petit que N

x un point qui n'appartient pas forcément à A

La recherche des k plus proches voisins consiste à trouver quels sont les k points de A les plus proches de x. La définition complète de cet algorithme est:

Pour i allant de 1 à k

Mettre le point  $D[i]$  dans proches\_voisins.

Fin Pour

Pour i allant de k+1 à N

Si la distance entre  $D[i]$  et x est inférieure à la distance d'un des points de

proches\_voisins à x alors :

1. On supprime du proches\_voisins le point le plus éloigné de x
2. On met dans proches\_voisins le point  $D[i]$

Fin Si

Fin Pour

proches\_voisins contient les k plus proches voisins de x

### 2.5.1.2. Les Séparateurs à Vaste Marge

Les séparateurs à vaste marge (Support Vector Machines en anglais) ont été développés par *Vladimir Vapnik* (1998) à partir de la théorie statique sur l'apprentissage (*Vapnik-Chervonenki*) [54].

L'algorithme des SVM est développé à l'origine pour traiter le problème de classification "mono-classe" permettant de déterminer si un élément appartient (qualifié de positif) ou non (qualifié de négatif) à une classe. Le principe de cet algorithme est de chercher un hyperplan qui sépare au mieux les exemples positifs des exemples négatifs en garantissant que la marge entre le plus proche des positifs et des négatifs soit maximale (voir figure 2.4), cela garantit un bon niveau de généralisation, car de nouveaux exemples pourront ne pas être trop similaires à ceux utilisés pour trouver l'hyperplan, mais être situés franchement d'un côté de la frontière ou de l'autre. L'algorithme des SVM a montré son efficacité dans le domaine de la classification textuelle [17].

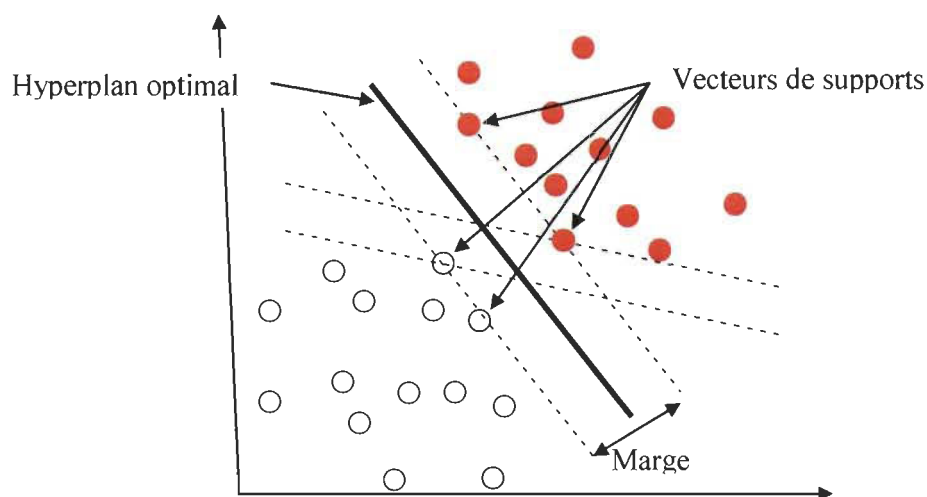


Figure 2.4 Apprentissage de classificateur SVM.

Les Systèmes à Vastes Marges ont montré de très bonnes performances dans d'autres domaines tels que la classification d'images, reconnaissance de formes. La complexité de ces algorithmes est, dans le pire des cas, en  $\mathcal{O}(N^3)$  pour la phase d'apprentissage et en  $\mathcal{O}(N)$  pour la phase de décision. [18]

Le principal inconvénient des SVM réside dans le choix de la fonction noyau adaptée aux données traitées. De plus, le classifieur a été conçu à l'origine pour la classification binaire (2 classes). Les extensions actuelles aux problèmes mutli-classes sont souvent construites sur la base d'une classe contre toutes les autres ce qui engendre un temps d'apprentissage très long. D'un autre côté, l'interprétation des résultats pour un problème particulier n'est pas toujours facile [18].

### 2.5.1.3. Bayes naïf

Le Bayes naïf est un classifieur de type probabiliste, souvent utilisé dans la catégorisation et la classification de documents: il classe les nouveaux documents selon la probabilité de ces documents à appartenir à chaque classe [17].

Soit le document  $d_j$  représenté par le vecteur  $v_j = (v_{j1}, \dots, v_{jk}, \dots, d_{jn})$ . Les variables aléatoires  $v_{j1}, \dots, v_{jk}, \dots, d_{jn}$  représentent les occurrences des mots retenues pour la classification dans le document  $d_j$ . En utilisant le théorème de Bayes (Bayes, 1763), la probabilité que le document  $d_j$  appartienne à la classe  $c_i \in C$  est [17]:

$$P(c_j | v_j) = \frac{P(c_i)P(v_j | c_j)}{P(v_j)} \quad (2.8)$$

La classe  $c_k$  d'appartenance du vecteur  $v_j$  représenté par  $v_j$  est définie comme suit [17]:

$$c_k = \operatorname{argmax} P(c_i \in C) \prod_k P(v_{jk} | c_j) \quad (2.9)$$

En d'autres termes, le classificateur Naïve de Bayes affecte au document  $d_j$  la classe la plus probable. Alors,  $P(c_i)$  est définie de la façon suivante [17]:

$$P(c_i) = \frac{\text{nombre de documents } \in c_i}{\text{nombre total de documents}} \quad (2.10)$$

En faisant l'hypothèse que les  $v_j$  sont indépendantes, la probabilité conditionnelle  $P(v_j|c_j)$  est définie ainsi [17]:

$$P(v_j|c_j) = P(v_{jk}|c_i) \quad (2.11)$$

Chaque nouvel document ajouté dans la base d'apprentissage, a besoin d'une phase d'apprentissage pour le calcul des différentes probabilités a priori et d'une phase de décision. Cette dernière phase a une complexité en  $\mathcal{O}(1)$ , c'est la phase d'apprentissage qui est plus longue, car elle nécessite d'effectuer les différents calculs de probabilité sur tous les exemples présentés par l'expert. D'un autre côté, le classificateur Naïve de Bayes a montré ses limites s'il est confronté à un trop grand nombre de mots qu'il n'a jamais rencontrés pendant la phase d'apprentissage [18].

## 2.5.2. Classification non supervisée

### 2.5.2.1. Méthodes de classification hiérarchique

Ces méthodes construisent les classes graduellement sous une forme hiérarchique. Cette représentation hiérarchique s'appelle dendrogramme (voir figure 2.5). Elles sont divisées en 2 sous-types : **Agglomération** (*Classification ascendante hiérarchique*) et **division** (*Classification descendante hiérarchique*) [4].

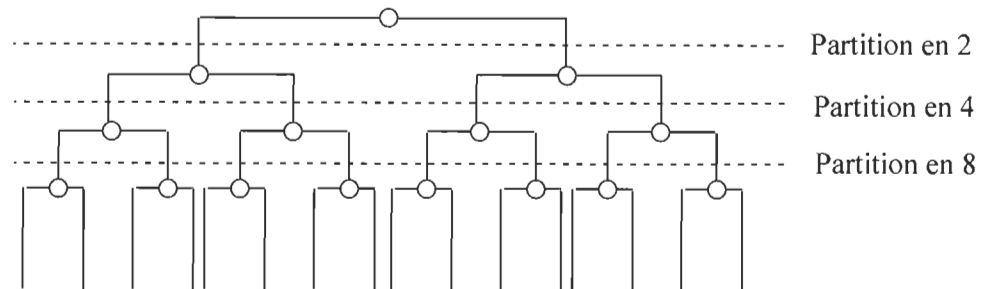


Figure 2.5 Exemple d'arbre hiérarchique.

#### 2.5.2.1.1. Classification Ascendante Hiérarchique CAH

On commence en considérant chaque objet comme une classe et en essayant de fusionner les classes les plus similaires. Le processus est itéré jusqu'à la vérification d'un critère d'arrêt ou bien jusqu'à ce que tous les objets se trouvent dans une même classe. Parmi les méthodes basées sur le Principe CAH, nous citons : Clustering Using Representatives (CURE), Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH), Robust Clustering using links (ROCK) [4].

#### 2.5.2.1.2. Classification Descendante Hiérarchique CDH

En considérant que l'ensemble des objets est regroupé dans une seule classe contenant tous les objets, en essayant de diviser la classe en deux classes dissemblables. Le processus est itéré jusqu'à la vérification d'un critère d'arrêt ou bien jusqu'à ce que chaque classe contienne un seul objet. Parmi les méthodes basées sur le principe CDH : Williams et Lambert, TSVQ (Tree Structured Vector Quantization) [4].



Afin de mesurer la similarité entre les classes, la plupart des méthodes hiérarchiques utilisent l'un des liens suivants : simple (single-link) [4]:

- Lien simple (single-link) : la distance entre 2 classes est la valeur minimum des distances entre toutes les paires d'individus, l'un de la première classe, l'autre de la deuxième.
- Lien complet (complete-link) : la distance entre 2 classes est la valeur maximum des distances entre toutes les paires d'individus
- Lien moyen : la distance entre 2 classes est la valeur moyenne des distances entre toutes les paires d'individus, l'un de la première classe, l'autre de la deuxième.

L'algorithme hiérarchique par agglomération est décrit ci-dessous. Le paramètre  $\epsilon$  représente généralement l'erreur de classification au carré.

### **Algorithme de Classification Ascendante Hiérarchique [1][3]**

Entée : Ensemble d'objets  $S = (s_1, s_2, \dots, s_n)$ ,  $\epsilon$  : critère d'arrêt,  $\beta$  : indice de similarité;

Les classes initiales sont les objets eux-mêmes (*pour  $i$  allant de 1 à  $n$ ,  $C_i = \{s_i\}$* ).

Sortie : Ensemble de classes  $C = (C_1, C_2, \dots, C_n)$ ;

1. Pour chaque paire  $(C_i, C_j)$ , calculer la similarité  $Sim(C_i, C_j)$ ;
2. Si  $Sim(C_i, C_j) \leq \beta$  Alors  $\{C_i, C_j\} \in C$  (regrouper ces deux classes dans une seule classe).
3. Répéter les étapes (1) et (2) jusqu'à ce que  $\epsilon$  Soit vrai.

Malgré que les méthodes hiérarchiques sont faciles à implémenter, mais leurs inconvénients résident dans le temps de calcul de l'algorithme quadratique  $O(n^2)$  en fonction du nombre d'objets de la base de données, car la distance entre toutes paires d'objets doit être calculée.

### 2.5.2.2. Méthodes de classification par partitionnement

Contrairement aux méthodes de classifications hiérarchiques qui construisent les classes progressivement, les algorithmes de partitionnement construisent directement une partition de l'ensemble d'individus en  $K$  classes [55]. Dans la section suivante, nous allons expliquer en détail deux algorithmes classiques qui restent très utilisés parmi les méthodes de classification non hiérarchique : méthodes de K-means et Méthodes des nuées dynamiques.

#### 2.5.2.2.1. Méthode K-means

L'algorithme *K-means* ou encore appelée *algorithme des centres mobiles* (Benzécri, 1973) est l'algorithme de clustering le plus connu et le plus utilisé, du fait de sa simplicité de mise en œuvre, il a été créé par MacQueen en 1967. Il partitionne les données en  $k$  classes. Les objets à l'intérieur de chaque classe sont aussi proches que possible les uns des autres et aussi loin que possible des objets des autres classes. Il commence d'abord par choisir aléatoirement les centres initiaux de classes, le centre de gravité  $g_i$  d'une classe est calculé à chaque itération comme suit : 
$$g_i = \frac{1}{c_i} \sum_{j=1}^{c_i} s_j^{(i)} \quad (2.12)$$

Où  $s_j^{(i)}$  représente le  $j^{\text{ème}}$  objet affecté à la classe  $C_i$ . À chaque itération les  $k$  clusters sont produits de façon à minimiser la fonction objective suivante :

$$\varepsilon = \sum_{i=1}^k \sum_{s_i \in C_i} d^2(s_i - C_i) \quad (2.13)$$

### Algorithme de classification K-means

Entée : Ensemble  $S = (s_1, s_2, \dots, s_n)$ ,  $k$  : nombre de classes;

Sortie : Ensemble de classes  $C = (C_1, C_2, \dots, C_n)$  et leurs centres de gravité  $(g_1, g_2, \dots, g_n)$ ;

1. choix d'une mesure pour le calcul des distances (euclidienne, Manhattan,...etc.);
2. Assigner aléatoirement  $k$  objets comme étant les centres initiaux;
3. pour chaque  $s_{i:1,n}$  de l'ensemble  $S$ , calculer la distance entre chaque  $s_i$  et tous les centres initiaux puis assigner  $s_i$  à la classe la plus proche;
4. mettre a jour les nouveaux centres de gravité  $g_i$  associés aux nouveaux objets ;
5. répéter les étapes (2), (3) et (4) jusqu'à ce que les centres de gravité ne changent plus.

La complexité en temps de calcul de l'algorithme du k-moyen est de l'ordre  $O(n, k, t)$ , où  $n$  est le nombre d'objets,  $k$  est nombre des classes,  $t$  est le nombre d'itérations nécessaires pour la convergence, et  $k < n$  [9].

L'avantage majeur de K-means est donc applicable pour le clustering des grandes bases de données. Cependant, malgré sa simplicité, le K-means à des inconvénients. Les résultats de la classification dépendent de la spécification a priori du nombre de classes et du choix des centres initiaux.

### 2.5.2.2.2. Méthodes des nuées dynamiques

Cette méthode peut être considérée comme une généralisation de la méthode des centres mobiles. Le principe de l'algorithme est le même, sauf que la partition n'est pas représentée par son centre mais par un noyau (ensemble de points représentatifs de partition). Le principe général des algorithmes des nuées dynamiques telles qu'il a été présenté par Diday dans [10] est le suivant [55] :

Soit  $E$  un ensemble d'objets. On cherche à constituer une division en  $k$  classes sur  $E$ . Chaque classe est représentée par son centre, également appelé noyau, constitué du petit sous-ensemble de la classe qui minimise le critère de dissemblance.

Les deux fonctions de base sur lesquelles repose l'algorithme sont définies de la manière suivante :

- On définit une application  $f$  de l'ensemble des classes d'une partition dans un ensemble de représentation (dont les éléments sont appelés les noyaux)
- On définit aussi une application  $g$  qui a une famille d'individus de l'ensemble de représentation, associe une partition.

Ainsi, si  $C = \{C_1, C_2, \dots, C_k\}$  est un ensemble des classes,  $f$  associe à cet ensemble un ensemble de noyau  $N = \{N_1, N_2, \dots, N_k\}$  par  $f(C) = N$  et  $N_i, C_j$ .

De même, pour une famille de noyau si  $N = \{N_1, N_2, \dots, N_k\}$ , l'application  $g$  associe à cette famille une partition  $C$  par  $C = g(N)$ . Où  $C_j$  est la classe associée à  $N_j$ .

Enfin, on se donne un critère (à valeurs positives) qui permet de mesurer l'adéquation d'une famille de noyaux à une partition : 
$$H(C, N) = \sum_{j=1}^k h(C_j, N_j) \quad (2.14)$$

Remarquons que  $H$ ,  $g$ ,  $h$  ne sont pas définis indépendamment :

- Pour  $C$  donné,  $f$  associe la famille de noyaux  $N$  qui minimise  $H$ .
- Pour  $N$  donné,  $g$  associe la partition  $C$  qui minimise  $H$ .

Le résultat de cet algorithme dépend du choix des paramètres initiaux. Il faut donc exécuter plusieurs fois l'algorithme et comparer les résultats de manière à découvrir les classes stables.

L'exécution de l'algorithme se termine après un nombre fixe d'itérations. Ce nombre est fixé au départ par l'utilisateur.

### **Algorithme de classification des nuées dynamiques**

Données:  $k$  le nombre maximum de classes désiré.

$i$  le nombre maximum d'itérations désiré.

$J=1$

Début

1. Donner une famille de noyaux  $N_0$
2. Dédire par  $g$  une famille de classes  $C_0$  (une partition)
3. Calculer la valeur de  $H(C_0; N_0)$
4. A l'aide de  $h$  déterminer une nouvelle famille de noyaux  $N_1$  telle que :

$$H(C_0; N_1) \leq H(C_0; N_0)$$

5. Calculer  $g(N_{01}) = C_1$  tel que :

$$H(C_1; N_1) \leq H(C_0; N_1)$$

$$j=j+1$$

Jusqu'à  $j=i$

FIN

L'avantage majeur de cette méthode est la capacité de traiter rapidement de grands ensembles de données. Par contre, les résultats finaux de la classification dépendent du choix initial du nombre de classes. En général les méthodes de classification par partitionnement permettent de traiter rapidement un grand ensemble de données, mais la spécification a priori du nombre de classes est un inconvénient puisque dans la plupart des cas, l'utilisateur ne détient pas cette information.

## 2.6. Conclusion

Ce chapitre était consacré à l'état de l'art des principales méthodes utilisées dans la résolution du problème de classification multi-classe. Plusieurs algorithmes sont proposés, ils diffèrent par les méthodes de classification, les mesures de proximité qu'ils utilisent, la nature des données qu'ils traitent. Le choix d'un algorithme approprié dépend fortement de l'application, la nature des données et les ressources disponibles.

Malgré le nombre important des méthodes de classification existantes, il n'existe pas une méthode qui peut répondre à toutes les demandes. Généralement ces méthodes connaissent plusieurs limites : complexité en temps de calcul et de performance lorsqu'il s'agit d'une grande masse de données à traiter, difficulté de fixer les paramètres en entrée des méthodes par l'utilisateur, l'utilisation d'heuristiques dans les algorithmes, difficulté de l'évaluation du résultat, difficulté pour comparer les différentes méthodes.

## Chapitre 3 - Règles d'associations

### 3.1. Introduction

Les techniques d'extraction de règles d'association ont été introduites par *R. Agrawal* en 1993 [13], dans le but de découvrir des tendances implicatives entre les données dans des bases de données dites « transactionnelles ». Plus précisément, dans une base de données qui contient un ensemble d'articles, en effet les règles d'association ont pour but d'identifier les articles ou les items qui sont fréquemment achetés ensemble [5].

Une règle d'association s'exprime de la façon suivante,  $X \rightarrow Y$ .  $X$  est appelé prémisse (aussi appelée antécédent) et  $Y$  conclusion (aussi appelée conséquent) de la règle où  $X \cap Y = \emptyset$ . on peut interpréter une règle de la façon suivante : « Lorsque j'observe la présence des événements  $X$  dans les données, alors les événements  $Y$  sont souvent observés ». Les termes « observe » et « souvent » sont quantifiés respectivement par les mesures de qualité : support et confiance [22]. Leurs définitions sont décrites plus précisément dans la section qui suit.

Les règles d'association ont été étudiées et employées dans de nombreux contextes et pour des domaines d'applications variés. On peut citer notamment : l'aide à la planification commerciale, l'aide au diagnostic et en recherche médicale, l'amélioration des processus de



télécommunications, l'organisation et l'accès aux sites Internet, l'analyse de données spatiales et de données statistiques [22].

L'extraction des règles d'association consiste à déterminer les règles dont le *support* et la *confiance* qui sont supérieurs ou égaux, respectivement, à un seuil minimal de support *minsup* et un seuil minimal de confiance *minconf*, prédéfini par l'utilisateur [5].

### 3.2. Définitions et notations

Dans cette section, nous parlerons des concepts ou termes souvent utilisés dans le domaine de l'analyse de règles d'association [11][22][23].

#### 3.2.1. Base de données binaire

Les données transactionnelles à explorer peuvent être représentées sous la forme d'un tableau binaire de dimension  $n \times p$ , croisant un ensemble  $T = \{t_1, t_2, \dots, t_n\}$  de  $n$  objets ou de transactions avec un ensemble  $I$ , de  $m$  attributs booléens, appelés *items* ( $I = \{i_1, i_2, \dots, i_m\}$ ). Chaque transaction est associée à un identifiant unique (noté  $T_{id}$ ).

Le tableau 3.1 illustre ce type de données, pour un ensemble de 6 transactions  $\{t_1, t_2, t_3, t_4, t_5, t_6\}$  décrit par 5 *items*  $\{i_1, i_2, i_3, i_4, i_5\}$ . Chaque transaction est représentée par l'ensemble des attributs observés. Dans le tableau 3.2, ces transactions sont présentées sous une forme binaire (1 pour la présence d'un d'un attribut, 0 sinon). Ces données servent de référence pour les exemples présentés dans ce chapitre.

$T_{id}$	Transaction
$t_1$	$\{i_1, i_2, i_3, i_4, i_5\}$
$t_2$	$\{i_1, i_2, i_3, i_5\}$
$t_3$	$\{i_3\}$
$t_4$	$\{i_2, i_3\}$
$t_5$	$\{i_1, i_2, i_3, i_4\}$
$t_6$	$\{i_1, i_2, i_3\}$

Tableau 3.1 Exemple de base de données transactionnelles T .

$T_{id}$	$i_1$	$i_2$	$i_3$	$i_4$	$i_5$
$t_1$	1	1	1	1	1
$t_2$	1	1	1	0	1
$t_3$	0	0	1	0	0
$t_4$	0	1	1	0	0
$t_5$	1	1	1	1	0
$t_6$	1	1	1	0	0

Tableau 3.2 Représentation binaire de T.

### 3.2.2. Itemset

L'ensemble  $S = \{i_1, i_2, \dots, i_k\} \subseteq Items$  est appelé itemset (ensemble d'articles), ou  $k$ -itemset s'il contient  $k$  éléments. Par exemple  $\{i_1, i_2, i_3\}$  est un 3-itemset.

### 3.2.3. Règle d'association

Une règle d'association est définie comme une implication entre deux itemsets. Elle est de la forme suivante  $R: X \Rightarrow Y$  telle que  $X \subseteq I$ ,  $Y \subseteq I$  et  $X \cap Y = \emptyset$ .

$X$  est appelé prémisse, antécédent ou tête de la règle et  $Y$  la conclusion, conséquent ou corps de la règle.

La règle d'association peut être quantifiée par un ensemble de mesures. Les deux mesures les plus classiques sont le support et la confiance.

### 3.2.4. Support d'un itemset $X$

C'est le rapport entre le nombre de transactions contenant  $X$  et le nombre total de transactions d'une base de données transactionnelle  $D$ .

$$\text{Support}(X) = \frac{|\{t \in D : X \subseteq t\}|}{|D|} \quad (3.1)$$

Si on prend les données listées dans le tableau 3.1, on trouve par exemple :

$$\text{Support}(i_1) = \frac{|\{t_1, t_2, t_5, t_6\}|}{6} = \frac{4}{6} = 66,66\%$$

$$\text{Support}(\{i_3, i_4\}) = \frac{|\{t_1\}|}{6} = \frac{1}{6} = 16,66\%$$

### 3.2.5. Support d'une règle d'association

On définit le support d'une règle d'association  $X \rightarrow Y$  comme étant le support de l'itemset  $X \cup Y$  (i.e. la proportion de transactions contenant à la fois  $X$  et  $Y$ ):

$$\text{Support}(X \rightarrow Y) = \text{Support}(X \cup Y) = \frac{|\{t \in D : X \subseteq t \text{ et } Y \subseteq t\}|}{|D|} \quad (3.2)$$

### 3.2.6. Confiance

La confiance d'une règle  $X \rightarrow Y$ , est défini comme étant le rapport entre le support de l'itemset  $X \cup Y$  et de celui de  $X$ . Elle exprime également la probabilité conditionnelle qu'une transaction contienne  $Y$  sachant qu'elle contient  $X$ .

$$\text{confiance}(X \rightarrow Y) = \frac{\text{support}(X \cup Y)}{\text{support}(X)} \quad (3.3)$$

Le support et la confiance sont deux mesures permettant d'évaluer la force d'une règle. Une règle d'association ayant une confiance égale à 100 % est appelée *règle d'association exacte*. Sinon, elle est appelée *règle d'association approximative ou partielle*.

### 3.3. Étapes de l'extraction des règles d'association

L'extraction de règles d'association est un processus itératif et interactif [12]. Ce processus peut être décomposé en quatre étapes qui sont représentées dans la figure 3.1 [11].

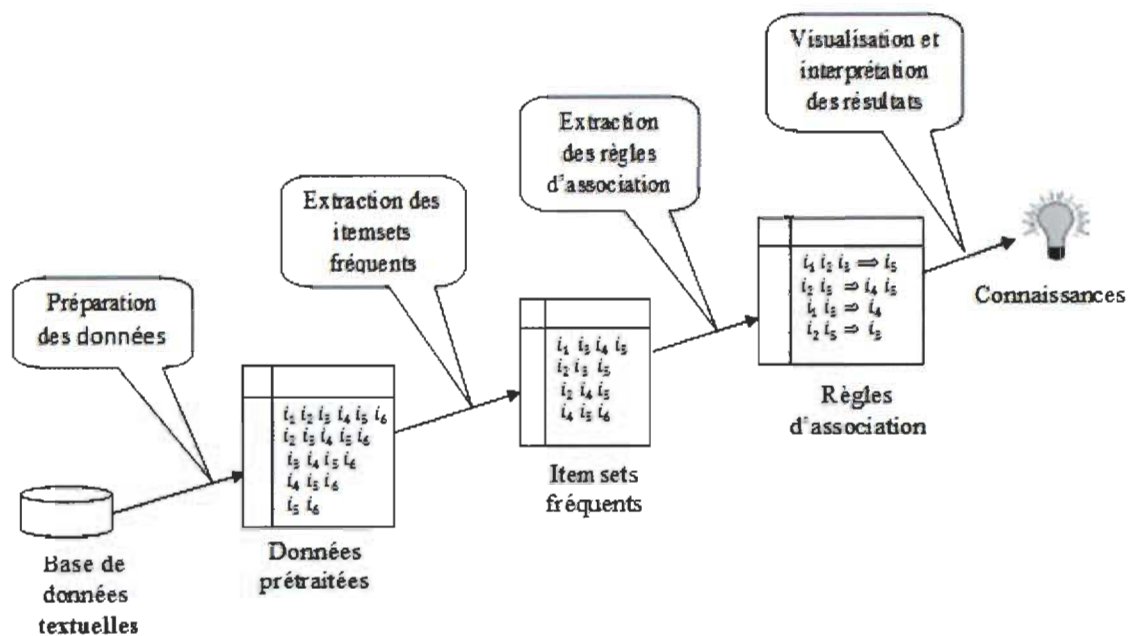


Figure 3.1 Étapes du processus d'extraction de règles d'association [11].

### 3.3.1. Sélection et préparation des données

Cette étape consiste à réduire la quantité des données en gardant seulement celles les plus pertinentes, et en transformant par la suite, ces dernières en un contexte d'extraction, c'est-à-dire une transformation en triplets constitués : d'un ensemble d'objets, d'un ensemble d'Itemsets ainsi qu'une relation binaire entre les deux [13]. Cette transformation est nécessaire afin qu'il soit possible d'appliquer les algorithmes d'extraction de règles d'association sur divers types de données (données relationnelles, transactionnelles, spatiales, multimédias, etc.) [14], et de permettre de réduire le temps d'extraction des itemsets fréquents et d'améliorer la qualité des règles d'association.

### 3.3.2. Découverte des itemsets fréquents

L'extraction des itemsets fréquents permet d'isoler, depuis une base de données, des ensembles d'attributs, qui satisfont un seuil de fréquence minimale *minsup* spécifié par l'utilisateur.

Cette étape est très coûteuse en temps d'exécution. Pour un ensemble de  $m$  items par exemple, le nombre d'Itemsets fréquents qui peut être générés est de  $2^m$ . Ces itemsets forment un treillis dont la représentation sous forme de diagramme de Hasse est présentée dans la figure 3.2 [22]. La recherche d'itemsets fréquents peut se faire selon le sens du parcours, de gauche à droite ou du haut en bas.

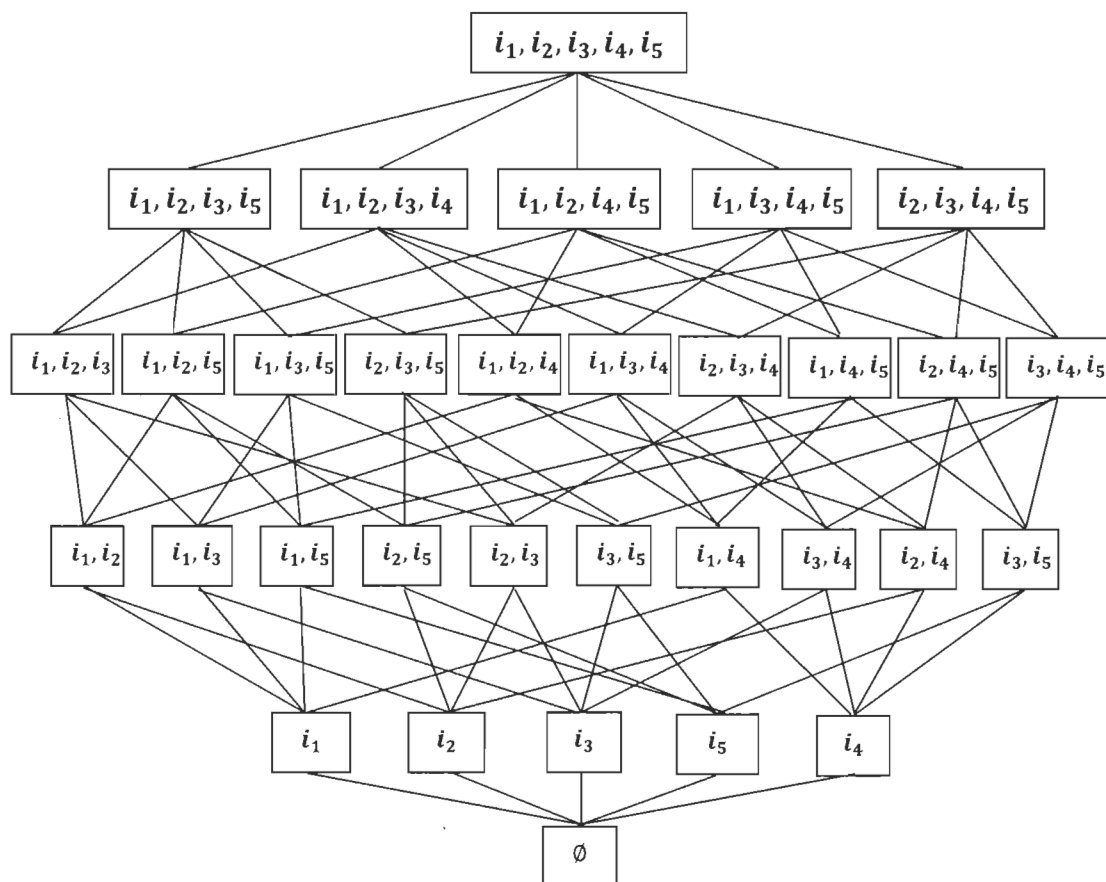


Figure 3.2 Diagramme de Hasse représentant le treillis des itemsets obtenus à partir des données du tableau 3.1 (inspiré de [48]).

### 3.3.3. Génération des règles d'association

Sur la base des itemsets fréquents extraits à l'étape précédente, il est possible de construire des règles d'associations, de forme générales  $X \rightarrow Y$  qui associent un sous-ensemble d'itemsets  $X$  avec un second sous-ensemble d'itemsets  $Y$ . Afin de limiter l'extraction aux règles d'association les plus informatives ou non redondantes entre elles, seules celles qui possèdent une confiance supérieure ou égale au seuil minimal défini par l'utilisateur sont générées [48].

**3.3.3.1 Support minimum (*minsup*) :** Le support minimum d'une règle d'association est le support minimum défini par l'utilisateur que la règle d'association doit satisfaire pour être acceptée.

**3.3.3.2. Confiance minimum (*minconf*) :** La M-confiance minimum d'une règle d'association, représente la confiance minimale définie par l'utilisateur que la règle doit satisfaire pour être acceptée.

### **3.3.4. Visualisation et interprétation des résultats**

Cette phase consiste en la visualisation par l'utilisateur des règles d'association découvertes et leur interprétation afin d'en déduire des connaissances cachées. La grande masse de règles d'association générées impose des outils de classification des règles, et de représentation de ces règles sous une forme de graphes dirigés. Cette forme peut être textuelle, graphique ou bien la combinaison des deux formes de représentation [11].

Plus de détails sur cette étape seront présentés dans le prochain chapitre.

### **3.4. Algorithme de génération de règles d'association : Algorithme *Apriori***

L'algorithme Apriori [Agrawal et al. 1993] est l'algorithme fondateur de la recherche de règles d'association. Il se divise en deux parties [51] :

– la *recherche des itemsets fréquents* : les itemsets fréquents sont les groupes d'items (attributs) dont le support est supérieur à *minsup*,

– l'*extraction des règles* parmi l'ensemble des itemsets fréquents : les règles dont la confiance est supérieure à *minconf*.

- **Première partie : Recherche des itemsets fréquents**

L'ensemble des itemsets fréquents se produit dans cette partie. L'algorithme **Apriori** (voir figure 3.3) construit des ensembles de plus en plus grands d'itemsets fréquents.

Elle consiste à parcourir la base de données de manière itérative, pour calculer les itemsets fréquents de taille  $k$  ( $k$ -itemsets noté  $L_k$ ) à partir des itemsets fréquents de taille  $k-1$  ( $k-1$ -itemsets notés  $L_{k-1}$ ). Le procédé est le suivant :

- construire un ensemble candidat  $C_k$ .
- jointure des itemsets de taille  $k$  avec ceux de taille  $k-1$ , en éliminant les itemsets de taille  $k$  qui contient un sous-itemset non fréquent. En effet tout sous-itemset d'un  $k$ -itemset fréquent est fréquent (Si  $i_1 i_2 i_3$  est fréquent, alors  $i_1, i_2, i_3, i_1 i_2, i_1 i_3, i_2 i_3$  le sont nécessairement) .
- calcul du support de chaque itemset candidat, et ne garder que ceux dont le support est supérieur à *minsup*.

Notons que le temps investi pour la génération des candidats est très couteux. En effet, pour  $m$  items de départ, la taille de l'espace de recherche des  $n$ -itemsets fréquents (obtenue pour *minsup* = 0) vaut  $C_n^m$ , soit au total  $2^n$ .

La figure 3.3 présente l'algorithme Apriori de recherche des itemsets fréquents.



$C_k$ : Ensemble des k-itemsets candidats  
 $L_k$ : Ensemble des k-itemsets fréquents  
*Apriori\_Gen*: Fonction qui engendre les itemsets candidats  
**Entrée :**  $B$  : une base de données transactionnelles  
            $minsup$ : un seuil minimum de support  
**Sortie :**  $\cup_k L_k$  l'ensemble des itemsets fréquents de  $B$

**Début**

$L_1 \leftarrow \{1\text{-itemsets fréquents}\}$

**Pour** ( $k \rightarrow 2; L_{k-1} \neq \emptyset, k++$ ) **Faire**

          // *Apriori\_Gen*

          // calcul des itemsets candidats de taille  $k$  à partir de ceux de taille  $k-1$

$C_k \leftarrow \text{Apriori\_Gen}(L_{k-1})$

**Pour tout** ( $t \in B$ ) **Faire**

$C_t \leftarrow \text{Subset}(C_k, t)$  // Sélectionné les itemsets de  $C_k$  contenus dans la

  // transaction  $t$

**Pour tout** ( $c \in C_t$ ) **Faire**

$support(c) \leftarrow support(c) + 1$

**Fin Pour**

**Fin Pour**

          // Sélection des itemsets vérifiant la contrainte de support  $s_0$

$L_k \leftarrow \{c \in C_k | support(c) \geq minsup\}$

**Fin Pour**

**Fin Pour**

Retourner  $\cup_k L_k$

Figure 3.3 Algorithme Apriori (recherche des itemsets fréquents) [51].

- **Deuxième partie : extraction de règles**

L'algorithme Apriori sauvegarde à chaque passage, le support des k-itemsets fréquents. Ces derniers sont utilisés dans la deuxième partie pour calculer les confiances des k-itemsets fréquents, car la confiance est définie par une relation simple avec le support [51].

La figure 3.4 présente l'algorithme d'extraction des règles par la confiance.

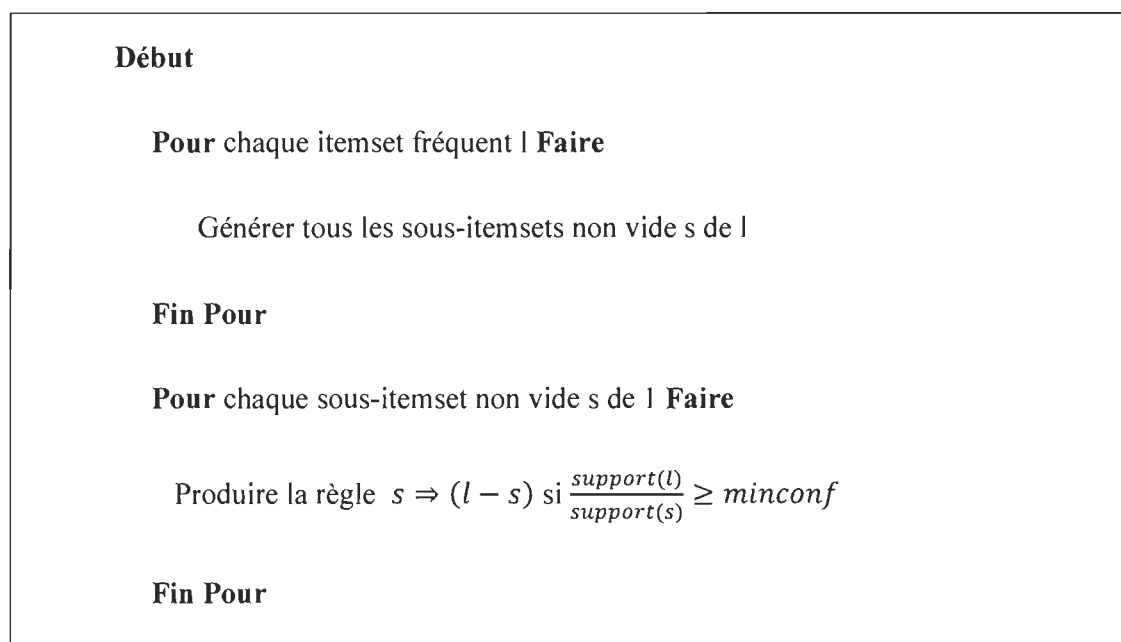


Figure 3.4 Algorithme d'extraction de règles.

### **3.5. Conclusion**

Dans ce chapitre, les itemsets, le treillis d'itemsets et des mesures d'évaluation utiles pour l'extraction des itemsets ont été présentés. Nous avons vu que le nombre d'itemsets possible dépend exponentiellement du nombre d'items manipulés. Ainsi, le seuil de support prédéfini par l'utilisateur permet d'en extraire que les itemsets fréquents. Le processus d'extraction des règles d'association et l'algorithme Apriori ont été également abordés.

## Chapitre 4 - Règles d'association maximales

### 4.1 Introduction

Les règles d'associations (Agwal et al. en 1993) sont devenues aujourd'hui un outil populaire en fouille de données. Les chercheurs ont élargi cet outil, en essayant d'extraire les informations les plus intéressantes. Plusieurs types de règles d'association ont été introduits tels que : règles d'association généralisées (en 1995, Srikant et al.), règles d'association quantitatives (Srikant, R et al.), règles d'association rares, règles d'association floues (Keith C.C. Chan et al. en 1997), et autres [24].

Bien que les règles d'association prévoient des moyens pour découvrir de nombreuses associations intéressantes, ils ne parviennent pas à découvrir d'autres associations moins intéressantes, également cachées dans les données. En 1997, Feldman et al. ont exposé un nouveau type de règle d'association, appelée règles d'association maximales, qui permet de découvrir des associations intéressantes qui sont souvent perdues lors de l'utilisation des règles d'association régulières.

Pour bien illustrer la faiblesse des règles d'associations régulières, on considère les deux exemples suivants :

**Exemple 1.** Dans une collection de documents, le mot "*HOSNI MOBAREK*" peut apparaître le plus souvent en association avec "*Egypte*", et le mot "*Windows*" le plus

souvent avec "Microsoft". Une association entre "Egypte" et "Ligue arabe" aurait moins de confiance, en raison des nombreuses transactions où "HOSNI MOBAREK" apparaît avec "Egypte" (et sans "Ligue arabe"), et aussi, l'association entre "Windows" et "Unix", serait perdu en raison du nombre important des transactions pour lesquelles "Windows" apparaît avec "Microsoft" (sans "Unix") [24].

Une règle d'association régulière  $X \Rightarrow Y$  affirme que lorsqu'on voit  $X$  il faut aussi s'attendre à voir  $Y$  (avec une certaine confiance). Cependant, la règle d'association maximale  $X \xRightarrow{\max} Y$  indique que lorsqu'on voit  $X$  tout seul on doit aussi s'attendre à voir  $Y$  seul [24].

**Exemple 2.** Soit une base de données  $D$ , qui contient plusieurs articles de presse classés par noms de pays, sujets, bourses et noms d'auteurs [24].

Supposons qu'il ya 10 articles se référant à "Pétrole" et aussi aux "USA" et "Russie". Et 20 articles concernant "Pétrole" et les pays : "Russie", "Brésil" et "Algérie".

Si on veut chercher les associations régulières avec au moins 50 % de confiance, on trouve que les règles suivantes :

$\{USA, Russie\} \Rightarrow \{Pétrole\}$  avec confiance de 66 %

$\{USA, Russie, Algérie\} \Rightarrow \{Pétrole\}$  avec confiance de 100 %

Les informations concernant la relation étroite entre "USA", "Russie" et "Pétrole" sont perdues. En substance, nous tenons à saisir la notion qu'à chaque fois que le Brésil et l'USA apparaissent seuls (sans aucun autre pays), le "Pétrole" apparaît aussi.

Les règles d'association maximales nous fournissent l'outil nécessaire qui est capable de capturer la règle :  $\{USA, Russie\} \Rightarrow \{Pétrole\}$  avec une confiance de 100 %.

Dans les sections suivantes, nous présentons formellement la notion d'associations maximales et l'algorithme d'extraction de toutes les règles, selon [25][26].

## 4.2 Taxonomie et catégorie

Soit un ensemble d'items  $I = \{i_1, i_2, \dots, i_{|A|}\}$ . Une taxonomie  $T$  de  $I$  est une partition de  $I$  en des sous-ensembles disjoints. C.-à-d.  $T = \{T_1, T_2, \dots, T_n\}$ . Chaque membre de  $T$  est appelé une catégorie. Pour un item  $i$ , on note  $T(i)$  la catégorie qui contient  $i$ . De même, si  $X$  est un ensemble d'itemsets qui sont tous d'une seule catégorie, alors on note cette catégorie par  $T(X)$ .

Soit un ensemble de données comprenant 10 transactions présentées dans le tableau

4.1 :

$T_{ID}$	Items
1	Brésil, Algérie, USA, pétrole, transport
2	Brésil, Algérie, USA, pétrole, transport
3	USA, croissance
4	USA, Emplois, prix
5	USA, Emplois, prix
6	USA, croissance, prix
7	Brésil, sucre, Blé
8	Brésil, USA, Vente, Taxe
9	Brésil, USA, Vente, Taxe
10	Brésil, USA, Croissance

Tableau 4.1. Les données de transactions.

- Deux articles se référant aux **pays** : "*Brésil, Algérie, USA*" et aux **sujets** : "*pétrole, transport*".
- Un article se référant aux "*USA*" et au "*croissance*".
- Deux articles se référant au "*USA*" et au "*emplois, prix*".
- Un article se référant au "*USA*" et au "*croissance, prix*".

- Un article se référant au "Brésil " et au "sucre, Blé".
- Deux articles se référant au "Brésil, USA" et au "commerce ", "Taxe".
- Un article se référant au "Brésil, USA" et au "croissance ".

Nous pouvons utiliser une taxinomie  $T$ , qui contient deux catégories  $T_1 = \text{"pays"}$  et  $T_2 = \text{"sujets"}$  (i.e.,  $T = \{\text{pays, sujets}\}$ ). Où  $\text{pays} = \{\text{Brésil, Algérie, USA}\}$  et  $\text{sujets} = \{\text{pétrole, transport, croissance, emplois, prix, sucre, Blé, commerce, Taxe}\}$ .

### 4.3 Règle d'association maximale

Une règle d'association maximale ou règle M-association est une règle de la forme  $X \xRightarrow{\max} Y$ , où  $X$  et  $Y$  sont des sous-ensembles de catégories distinctes, respectivement  $T(X)$  et  $T(Y)$ .

Dans une règle d'association maximale  $X \xRightarrow{\max} Y$ , on s'intéresse à capturer la notion suivante : chaque fois que  $X$  apparaît seul,  $Y$  apparaît aussi, avec une certaine confiance. Pour cela, nous devons d'abord définir la notion de "seul ". Nous le faisons par rapport aux catégories de  $T$  comme suit :

Pour une transaction  $t$ , une catégorie  $T_i$  et un itemset  $X \subseteq T_i$ , on dit que  $X$  est seul dans  $t$  si et seulement si  $t \cap T_i = X$ . c.-à-d.  $X$  est seul dans  $t$  si  $X$  est le plus grand sous-ensemble de  $T_i$  qui est dans  $t$ . Dans ce cas, on dit aussi que  $X$  est maximale dans  $t$  et que  $t$  M-supports  $X$ .

### 4.3.1 M-support d'un Itemset

Pour une base de données  $D$ , le M-support de  $X$  dans  $D$ , noté  $S_D^{max}(X)$  est le nombre de transactions  $t \in D$  qui M-supporte  $X$ .

Considérons les données du tableau 4.1, prenons par exemple les deux itemsets { Brésil, USA } et {USA}.

$S_D^{max}(X) = M - support\{Brésil, USA\} = 3$ . (Nombre de transactions dans  $D$  où l'itemset {Brésil, USA} apparait seul dans sa catégorie),  $T_{ID} = 8, 9 \text{ et } 10$ .

$S_D^{max}(Y) = M - support\{USA\} = 4$ . (Nombre de transactions dans  $D$  où l'itemset {USA} apparait seul dans sa catégorie),  $T_{ID} = 3, 4, 5 \text{ et } 6$ .

### 4.3.2. M-support d'une règle d'association

Le M-support de la règle M-association  $X \xRightarrow{max} Y$ , notée par  $S_D^{max}(X \xRightarrow{max} Y)$  est le nombre de transactions dans  $D$  qui M-supporte  $X$  et qui supporte aussi  $Y$ .

$S_D^{max}(X \xRightarrow{max} Y)$  est défini comme suit :

$$S_D^{max}(X \xRightarrow{max} Y) = |\{t: t \text{ M-supportent } X \text{ et } t \text{ supportent } Y\}| \quad (4.1)$$

Considérons les données du tableau 4.1, prenons par exemple la règle M-association  $\{Brésil, USA\} \xRightarrow{max} \{croissance\}$ .

$$S_D(Y) = S_D(\{croissance\}) = 3. \text{ (Y existe dans 3 transactions : } T_{ID} = 3, 6, 10).$$

$$S_D^{max}(X) = S_D^{max}(\{Brésil, USA\}) = 3. \text{ (les trois transactions sont : } T_{ID} = 8, 9, 10)$$



Le M-Support de cette règle d'association est le nombre de transactions qui M-Supporte l'itemset  $\{ Brésil, USA \}$  et supporte au même temps l'itemset  $\{ croissance \}$ .

Donc,  $S_D^{max} \left( \{ brésil, USA \} \xRightarrow{max} \{ croissance \} \right) = 1$ , puisque seule la transaction  $T_{ID} = 10$  vérifie les deux conditions.

### 4.3.3. M-confiance d'une règle d'association maximale

La règle M-association  $X \xRightarrow{max} Y$ , signifie que chaque fois qu'une transaction M-supporte X, alors Y apparaît aussi dans la transaction, avec une certaine probabilité.

Cependant, dans la mesure de cette probabilité, nous sommes intéressés seulement par les transactions pour lesquelles certain éléments de  $T(Y)$  (catégorie de Y) apparaissent dans la transaction. En conséquence, la confiance maximale est définie comme suit :

Soit  $D(X, T(Y))$ , sous-ensemble de la base de données D comprenant toutes les transactions qui M-supportent X et contiennent au moins un élément de  $T(Y)$ . Le M-confiance de la règle  $X \xRightarrow{max} Y$ , notée  $C_D^{max}(X \xRightarrow{max} Y)$  est définie comme suit :

$$C_D^{max} \left( X \xRightarrow{max} Y \right) = \frac{S_D^{max}(X \xRightarrow{max} Y)}{|D(X, T(Y))|}. \quad (4.2)$$

Considérons les données du tableau 4.1, prenons par exemple la règle M-association  $\{ Brésil, USA \} \xRightarrow{max} \{ croissance \}$ .

Pour trouver la M-confiance de cette règle, on doit d'abord calculer la valeur de  $D(X, T(Y))$ .

Les transactions qui M-soutiennent  $\{ Brésil, USA \}$  sont :  $T_{ID} = 8, 9, 10$ . Toutes ces trois transactions contiennent au moins un élément de  $T(Y)$  (catégories  $T_2 = "sujets"$ )

$$D(\{ Brésil, USA \}, T(\{croissance\})) = 3$$

$$\text{La M-confiance } C_D^{max}(X \xrightarrow{max} Y) = \frac{S_D^{max}(\{ Brésil, USA \} \xrightarrow{max} \{croissance\})}{|D(\{ Brésil, USA \}, T(\{croissance\}))|} =$$

$$\frac{1}{3} = 0,33 = 33\%.$$

#### 4.3.4. M-support minimum et M-confiance minimum

Comme dans les règles d'association régulières, pour extraire les règles d'association intéressantes dans un ensemble de données transactionnelles, l'utilisateur doit définir un M-support minimum (appelé *minMsup*) et un M-confiance minimum (appelée *minMconf*).

#### 4.3.5. Itemset maximal fréquent

L'itemset maximale  $X \subseteq I$  est appelé *itemset maximal fréquent* si  $S_D^{max}(X) \geq \text{minMsup}$ . Dans ce cas, un sous-ensemble d'un itemset maximal fréquent n'est pas nécessairement un itemset maximal fréquent.

#### 4.3.6. M-facteur d'une règle d'association maximale

Le M-facteur d'une règle M-association  $X \xRightarrow{\max} Y$  est le rapport entre la M-confiance de la règle M-association  $X \xRightarrow{\max} Y$  et de la confiance de la règle d'association régulière correspondante  $X \Rightarrow Y$ . Plus précisément, soit  $D'$  le sous-ensemble de transactions qui contiennent au moins un élément de  $T(Y)$ . Alors, le M-facteur de la règle M-association  $X \xRightarrow{\max} Y$  est défini comme suit :

$$M - \text{facteur} \left( X \xRightarrow{\max} Y \right) = \frac{c_D^{\max}(X \xRightarrow{\max} Y)}{c_{D'}(X \Rightarrow Y)} \quad (4.3)$$

Notez que le dénominateur est la confiance de la règle d'association régulière  $X \Rightarrow Y$  par rapport à  $D'$ . En général, pour trouver les règles les plus intéressantes, nous cherchons les règles M-associations avec de plus grandes M-facteur.

Considérons les données du tableau 4.1, prenons par exemple la règle M-association  $\{ \text{Brésil, USA} \} \xRightarrow{\max} \{ \text{croissance} \}$ .

Pour calculer le M-facteur de cette règle, on doit d'abord identifier le sous-ensemble  $D'$ .

Les transactions qui contiennent au moins un élément de  $T(Y)$  (catégorie  $T_2 =$  "sujets") sont :  $T_{ID} = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10$ . Donc  $D'$  contient toutes ces transactions.

$$\begin{aligned} c_{D'}(\{ \text{Brésil, USA} \} \Rightarrow \{ \text{croissance} \}) &= \frac{S_{D'}(\{ \text{Brésil, USA} \} \cup \{ \text{croissance} \})}{S_{D'}(\{ \text{Brésil, USA} \})} = \frac{1}{5} \\ &= 0,2 \end{aligned}$$

$$\begin{aligned}
 & M - \text{facteur} \left( \{ \text{Brésil}, \text{USA} \} \xRightarrow{\text{max}} \{ \text{croissance} \} \right) \\
 &= \frac{C_D^{\text{max}} \left( \{ \text{Brésil}, \text{USA} \} \xRightarrow{\text{max}} \{ \text{croissance} \} \right)}{C_D \left( \{ \text{Brésil}, \text{USA} \} \Rightarrow \{ \text{croissance} \} \right)} = \frac{0,33}{0,2} = 1,65
 \end{aligned}$$

Si on considère la règle  $\{ \text{Brésil}, \text{USA} \} \Rightarrow \{ \text{croissance} \}$  comme règle d'association régulière, alors : le support de la règle :  $S_D(\{ \text{Brésil}, \text{USA} \} \Rightarrow \{ \text{croissance} \}) = 1$  (le nombre de transactions qui contiennent *Brésil*, *USA* et *croissance*).

$$\begin{aligned}
 & \text{La confiance de la règle : } C_D(\{ \text{Brésil}, \text{USA} \} \Rightarrow \{ \text{croissance} \}) = \\
 & \frac{S_D(\{ \text{Brésil}, \text{USA} \} \cup \{ \text{croissance} \})}{S_D(\{ \text{Brésil}, \text{USA} \})} = \frac{1}{5} = 0,2 = 20 \%.
 \end{aligned}$$

Il est clair que la confiance de l'association régulière  $\{ \text{Brésil}, \text{USA} \} \Rightarrow \{ \text{croissance} \}$  est différente si cette règle est considérée comme association maximale.

#### 4.4. Algorithme de calcul des règles d'association maximales

Le calcul des règles d'association maximales est en général, beaucoup plus rapide que le calcul des associations régulières (à condition que le nombre de catégories ne soit pas trop grand). La raison est que, pour chaque catégorie, toute transaction M-support au plus un itemset.

Dans cette section, nous présentons une analyse détaillée du pseudo-code de l'algorithme d'extraction de tous les M-associations (figure 4.1 et figure 4.2) [24].

**Associations maximales ( $D, G, \hat{s}, \hat{c}$ )**

$D$  – Base de données,

$G$ - groupement des littéraux en des catégories,

$\hat{S}$ - M-support minimum,

$\hat{C}$ - M-confiance minimum,

1  $M \leftarrow M - \text{Frequent} - \text{Sets}(\hat{s})$

2 Pour chaque  $X \in M$  faire

3    Pour chaque  $g \in G$  faire

4        $D' \leftarrow D(X, g)$

5        $\bar{s} \leftarrow \max(\hat{s}, \hat{c}, |D'|)$

$F \leftarrow \text{Frequent} - \text{Sets}(D', \bar{s})$  (Rechercher et calculer le support de tous  
les Ensembles d'itemsets, avec  $\text{support} \geq \bar{s}$  dans  $D'$ )

6    Pour chaque  $Y \in F$  faire

7       Sortie  $X \xrightarrow{\max} Y$

8       Sortie  $M - \text{support} = S_{D'}(Y), M - \text{confiance} = \frac{S_{D'}(Y)}{|D'|}$

9       Fin pour

10    Fin pour

11 Fin pour

Figure 4.1 Algorithme de calcul des règles M-associations [24].

L'algorithme commence par la génération de tous les ensembles M-fréquents, en utilisant la procédure *M-fréquentes-Sets ()* (ligne 1).

Pour chaque ensemble de M-fréquentes  $X$  et pour chaque catégorie  $g$ , la procédure *M-fréquentes-Sets* () génère aussi des sous bases de données  $D(X, g)$ , (pour chaque transaction  $t$  M-support  $X$ ,  $D(X, g)$  comprend la partie de  $t$  qui est dans  $g$ ).

Après le calcul des ensembles M-fréquents qui constituent l'antécédent des règles, l'algorithme passe ensuite au calcul de la conséquence des règles.

Soit  $X$  un ensemble M-fréquents, et supposons  $X \xrightarrow{\max} Y$  est une M-association, où  $Y \subseteq g$ .

Soit  $D'$  sous base de données  $D(X, g)$ , seules les transactions qui appartiennent à  $D'$  peuvent probablement supporter  $X \xrightarrow{\max} Y$ .  $s_{D'}(Y)$  est le support de  $Y$  dans  $D'$ ,  $s_{D'}(Y)$  est le M-support de  $X \xrightarrow{\max} Y$  et  $\frac{s_{D'}(Y)}{|D'|}$  est le M-confiance de la règle.

Afin de trouver toutes les M-associations avec un *M-support minimum*  $\hat{s}$  et un *M-confiance minimum*  $\hat{c}$ , il suffit de chercher dans  $D'$ , tous les ensembles  $Y$  avec un support  $\bar{s} = \max(\hat{s}, \hat{c} \cdot |D'|)$  (lignes 5–6).

La génération d'ensembles fréquents dans  $D'$  peut-être effectuée en utilisant n'importe quel algorithme connu pour extraction d'ensembles fréquents.

```

M-Frequent-Set( $\hat{s}$ ) (découvrir tous les ensembles avec M-support moins de  $\hat{s}$ )
1   $Large \leftarrow \emptyset$ 
2  Pour chaque  $t \in D$  faire
3    Pour chaque  $g \in G$  faire
4       $X \leftarrow t \cap g$ 
5      Si  $X \neq \emptyset$  alors
6         $Hash(X) ++$ 
7    Fin pour
8  Fin pour
9  Pour chaque  $t \in D$  faire
10   Pour chaque  $g \in G$  faire
11      $X \leftarrow t \cap g$ 
12     Si  $X \neq \emptyset$  et  $Hash(X) \geq \hat{s}$  alors
13        $S^{max}(X) ++$ 
14       Si  $S^{max}(X) \geq \hat{s}$  alors
15          $Large \leftarrow Large \cup \{X\}$ 
16         Pour chaque  $g' \in G$  si  $g' \cap t \neq \emptyset$  alors
17            $D(X, g') \leftarrow D(X, g') \cup \{t \cap g'\}$ 
18         Fin pour
19  Fin pour
20 Retourner  $Large$ 

```

Figure 4.2 Algorithme de la procédure M-Frequent-Set( $\hat{s}$ ) [24].

La procédure *M-fréquentes-Sets* () trouve tous les ensembles de M-fréquents en utilisant deux passages dans la base de données. Le premier passage vise à réduire le nombre de candidats d'ensembles M-fréquents, tandis que le second passage génère les

ensembles M-fréquents et les sous bases de données  $D(X, g)$  correspondante. Les itemsets  $X$  sont structurés dans un tableau de hachage  $Hash(X)$

Dans le premier passage (lignes 2-8) la base de données est balayée en séquence, pour chercher les itemsets  $X$  M-supportés par chaque transaction  $t$ . Pour chaque  $X$  trouvée, on incrémente la valeur de  $Hash(X)$  par 1. À la fin de ce passage, seulement les ensembles d'itemset  $X$  pour lesquels  $Hash(X) \geq \hat{s}$  peuvent éventuellement être M-fréquent. Le nombre de candidats des ensembles de M-fréquent réduit considérablement après le premier passage (si ce n'est pas le cas, alors  $\hat{s}$  est fixé trop faible).

Dans le second passage (lignes 9-19), la base de données est de nouveau balayée en séquence Pour chaque transaction  $t$  et itemset  $X$  M-supportés par  $t$ , si  $Hash(X) < \hat{s}$  alors  $X$  est ignoré, car il n'est certainement pas M-fréquents. Sinon, il est un candidat d'ensemble M-fréquent, et son  $s^{max}(X)$  est incrémenté par 1. De plus, les parties appropriées de  $t$  sont ajoutées aux sous base de données  $D(X, g')$ . À la fin de la deuxième phase, ces itemsets avec  $s^{max}(X) \geq \hat{s}$  sont des ensembles M-frequent.

Nous notons que si la base de données est assez petite, la première phase peut être négligée.



#### **4.5. Conclusion**

Bien que les associations régulières sont basées sur la notion d'ensembles fréquents : ensembles d'attributs, qui apparaissent dans différents enregistrements.

Les associations maximales sont basées sur la notion d'ensembles maximaux fréquents. Sur le plan conceptuel, un ensemble maximal fréquent est un ensemble d'attributs, qui apparaissent seuls, ou maximaux, dans de nombreux enregistrements.

Feldman et al. [24] jugent que les associations régulières ne sont pas capables d'identifier des relations solides entre les données, et ils ont introduit les associations maximales pour améliorer la qualité d'extraction des règles d'association.

## Chapitre 5 - Visualisation des règles d'association

### 5.1. Introduction

Les algorithmes de fouille de données produisent des règles d'association en si grandes quantités. En pratique, il est très laborieux pour l'utilisateur d'extraire des connaissances intéressantes pour la prise de décisions dans un grand ensemble de règles. Face à cette masse de données, les techniques de représentation graphique de règles d'association semblent parmi les plus pertinentes. Dans ce chapitre nous présentons un état de l'art des techniques de visualisation en fouille de données, plus précisément les méthodes qui s'appliquent aux règles d'association.

### 5.2. Visualisation des règles sous forme de listes textuelles

Différentes applications ont été développées pour visualiser des ensembles de règles d'association sous forme de tableaux. Chaque ligne correspond à une règle et chaque colonne représente une caractéristique des règles, dont l'élément de la 1ère colonne représente la *prémisse*, la 2ème colonne représente la *conclusion*, les colonnes qui suivent représentent les mesures de qualités de règles.

Ces applications proposent une interface qui permet de filtrer les règles, soit par le choix du support minimum ou de confiance minimum des règles dont on veut les visualiser, soit par la spécification des items dont on veut avoir dans la règle (prémisse ou conclusion). Ces applications permettent aussi de trier les règles par ordre croissant ou décroissant de

support ou de confiance. La taille des visualisations peut être contrôlée directement en fixant le nombre de règles à visualiser simultanément.

Malgré que les explorateurs de règles textuels permettent d'afficher plusieurs mesures de qualité en même temps, la principale limite réside dans le mode de présentation textuel, qui ne convient pas à l'étude de grandes quantités de règles [30].

Il existe plusieurs logiciels qui proposent cette approche de visualisation telle que GRAMEXCO [45] et IRSetNav [46] comme illustrés dans la figure 5.1.

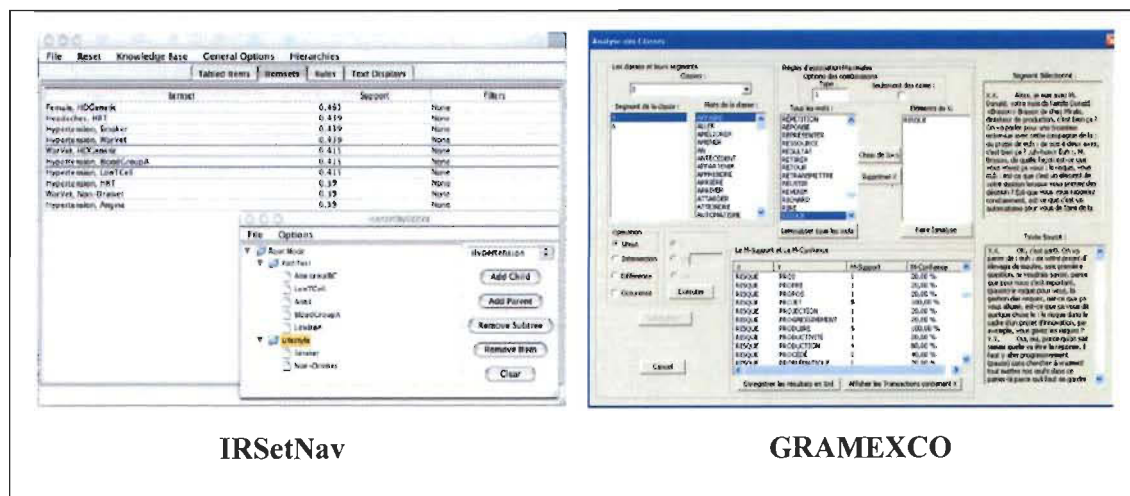


Figure 5.1 Visualisation textuelle de règles d'association .

### 5.3. Visualisation par graphe [29][36][37][41]

#### 5.3.1. Graphe orienté

Dans ce type de visualisation qui représente un ensemble des règles d'association, les nœuds symbolisent les items et les arcs symbolisent les règles, comme illustré dans la

figure 5.2 où  $i_1, i_2 \rightarrow i_3$  est une règle telle que  $i_1$  et  $i_2$  constituent la prémisse et  $i_3$  est la conclusion.

Les caractéristiques graphiques (dimensions et couleur) d'un arc représentent les mesures de qualité d'une règle (support et confiance).

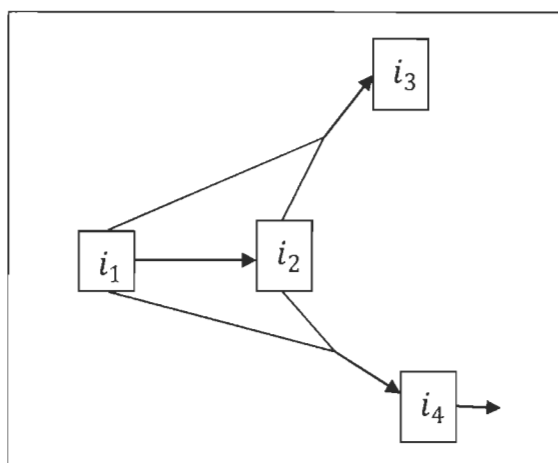


Figure 5.2 Visualisation des règles d'associations par un graphe.

Les arcs peuvent être orientés (habituellement de la prémisse vers la conclusion) ou non orientés.

### 5.3.2. Graphe non orienté

Soient les tableaux 5.1 et 5.2 montrant respectivement un ensemble d'itemsets fréquents et un ensemble de règles d'association

	<i>Itemset</i>	<i>Support</i>
$X_1$	$\{i_1\}$	9/11
$X_2$	$\{i_2\}$	7/11
$X_3$	$\{i_3\}$	7/11
$X_4$	$\{i_2, i_4\}$	6/11
$X_5$	$\{i_1, i_2\}$	6/11
$X_6$	$\{i_2, \bar{i}_5\}$	6/11
$X_7$	$\{i_3, i_5\}$	5/11
$X_8$	$\{i_1, i_3\}$	5/11
$X_9$	$\{i_1, i_2, i_4\}$	5/11
$X_{10}$	$\{i_1, i_2, \bar{i}_5\}$	5/11

Tableau 5.1 Ensemble d'itemsets fréquents.

	<i>Prémisse</i>	<i>Conclusion</i>	<i>Support</i>	<i>Confiance</i>
$r_1$	$\{i_4\}$	$\{i_2\}$	6/11	1
$r_2$	$\{i_2\}$	$\{i_4\}$	6/11	6/7
$r_3$	$\{i_5\}$	$\{i_3\}$	5/11	1
$r_4$	$\{\bar{i}_5\}$	$\{i_1\}$	6/11	1
$r_5$	$\{i_2\}$	$\{i_1\}$	6/11	6/7
$r_6$	$\{i_4\}$	$\{i_1\}$	5/11	5/6
$r_7$	$\{i_2, i_1\}$	$\{i_4\}$	5/11	5/6
$r_8$	$\{\bar{i}_5\}$	$\{i_2\}$	5/11	5/6
$r_9$	$\{i_2, i_1\}$	$\{\bar{i}_5\}$	5/11	5/6

Tableau 5.2: Ensemble de règles d'association.

Dans le cas d'un graphe non orienté, des indicateurs sont utilisés pour différencier le corps et la tête des règles. Dario Bruzese et Paolo Buono [42] ont utilisés deux couleurs pour distinguer le corps et la tête des règles. Si un itemset est en même temps la prémisse d'une règle et la conclusion d'une autre règle, alors son nœud porte les deux couleurs. Trois autres couleurs sont utilisées pour exprimer le support des règles sur les arcs : une couleur claire pour les supports faibles, une couleur foncée pour les supports élevés et une couleur noire pour les supports de 1. La confiance d'une règle est représentée par la longueur de l'arc qui lie son corps et sa tête. Plus la confiance est élevée, plus l'arc est long [41]. La figure 5.3 montre ce type de visualisation.

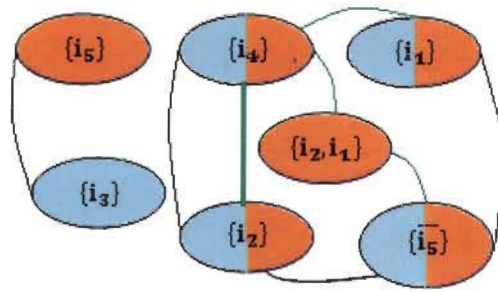


Figure 5.3 Graphe non orienté.

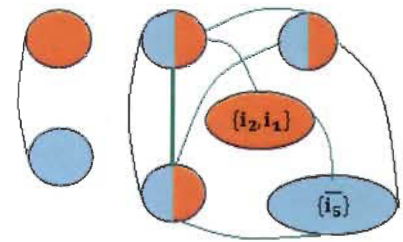


Figure 5.4 Réduction des nœuds et détail d'une règle.

Dans le cas d'une grande quantité de règles, les nœuds vont se transformer en petits cercles comme illustrés sur la figure 5.4. Le positionnement de la souris sur un arc fait apparaître le détail de la règle. La taille de la visualisation ne peut pas être contrôlée. Pour une exploration plus efficace des règles, Bruzzese et Buono proposent de sélectionner un sous-ensemble de règles qui partagent la même prémisse [41].

En général, les visualisations basées sur les graphes montrent vite leurs limites quand le nombre de règles est important : les arcs s'entrecroisent très souvent et l'abondance des nœuds et des arcs entraîne une occlusion. D'autre part, les visualisations ne donnent pas une vue globale des ensembles de motifs, car toutes les faces du graphe n'apparaissent pas simultanément. [41].

Il existe également plusieurs logiciels qui utilisent cette approche de visualisation telle que **DBMiner** (Han *et al.*, 1997)[31] et **DB2 Intelligent Miner Visualization**[35] comme illustrés dans la figure 5.5.

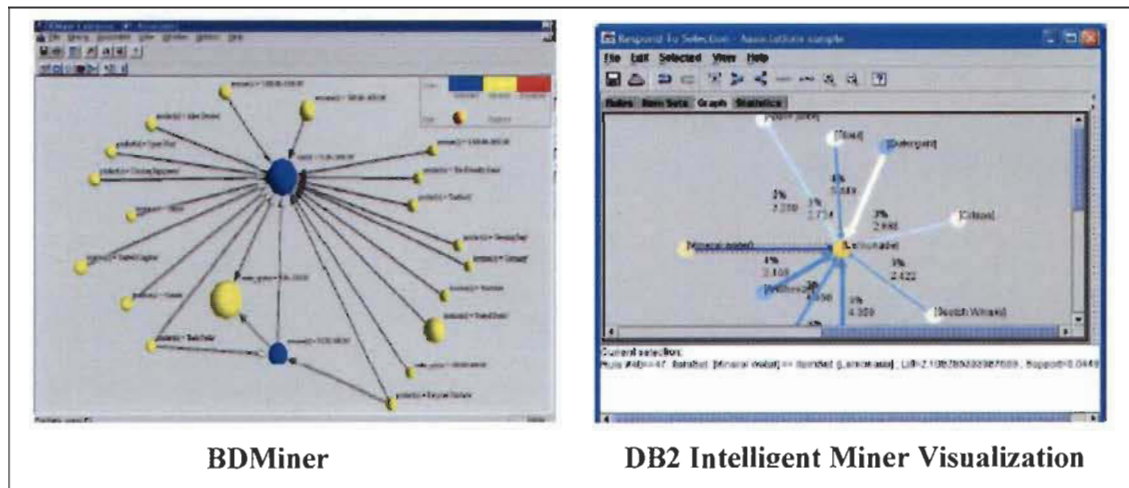


Figure 5.5 Visualisation de règles d'association par graphe.

#### 5.4. Visualisation par matrice

Comme son nom l'indique, c'est une méthode de visualisation sous forme de matrice « itemset-à-itemset », illustrée dans la figure 5.6, où chaque ligne correspond à un itemset en conclusion et chaque colonne à un itemset en prémisse. Une règle est illustrée dans la case d'intersection par un objet 2D ou 3D indiquant le support et la confiance de cette règle [43].

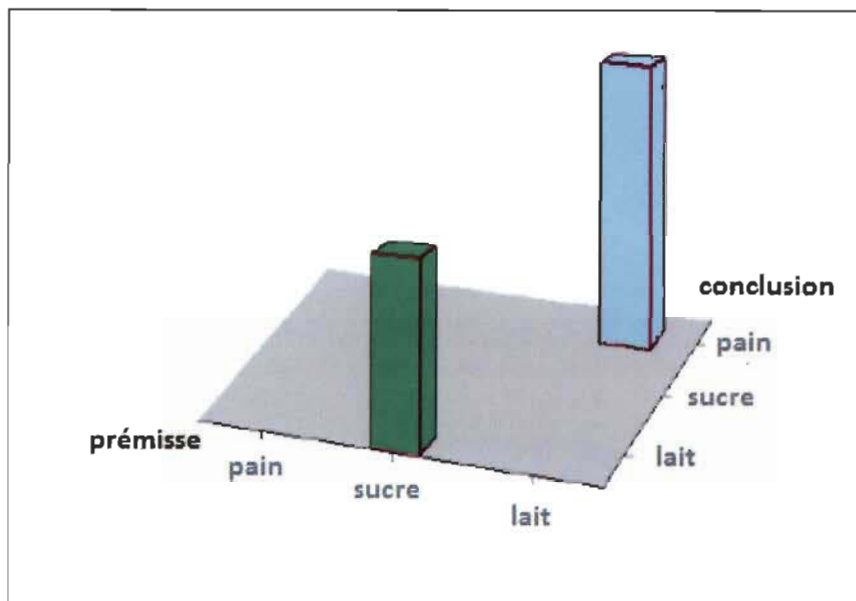


Figure 5.6 Une matrice 3D itemset-à-itemset [29].

Cette technique de visualisation a été améliorée en matrices item-à-règle (Wong *et al.* 1999) [27], comme indiqué dans la figure 5.7. Où chaque ligne correspond à un item en conclusion et chaque colonne à un item en prémisse. Une règle entre deux items est symbolisée dans la cellule à l'intersection par un objet 2D ou 3D qui indique la présence ou non de l'item dans la règle. La couleur de ces objets indique la présence de l'item dans la partie gauche ou la partie droite de la règle. La matrice est complétée par deux lignes qui indiquent le support et la confiance de chaque règle par la hauteur de barres dessinées en trois dimensions.



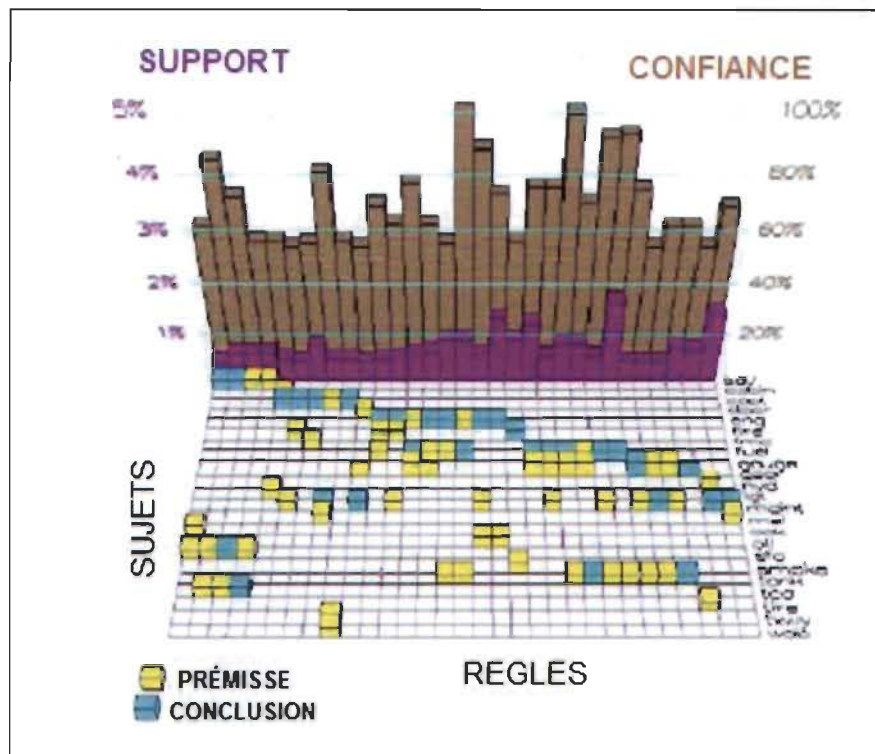


Figure 5.7 Une matrice D3 Item-à-règle [58].

Boulicaut [43] et ses collègues ont proposé une représentation matricielle 2D d'ensembles de règles d'association. Les têtes des règles sont représentées sur la ligne et les corps sur les colonnes de la matrice ou inversement, la cellule d'intersection correspond à une règle.

Un couple de valeurs qui représentent son support et sa confiance sont affichées comme sur la figure 5.8

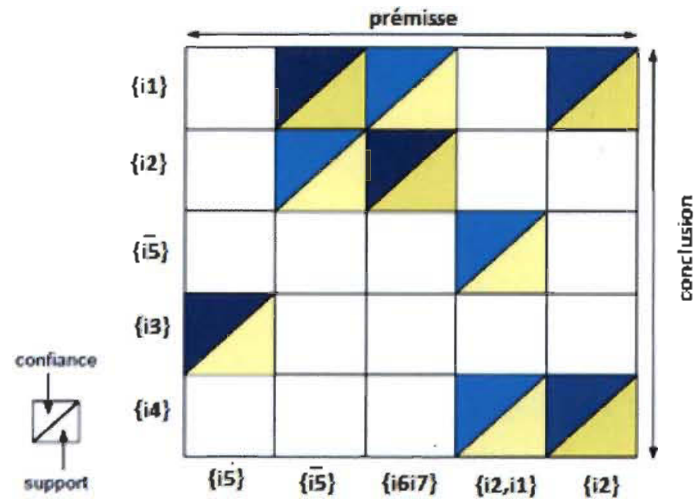


Figure 5.8 Une matrice 2D Item-à-règle [15].

L'inconvénient principal de ces types de représentations matricielles est qu'elles atteignent des tailles considérables dans le cas de grands ensembles de règles portant sur de nombreux items [29].

Il existe plusieurs logiciels qui utilisent la technique de visualisation par matrice comme **BDMiner** et **Mineset** (Brunk et Kohavi, 1997) [32] illustrés respectivement dans la figure 5.9.

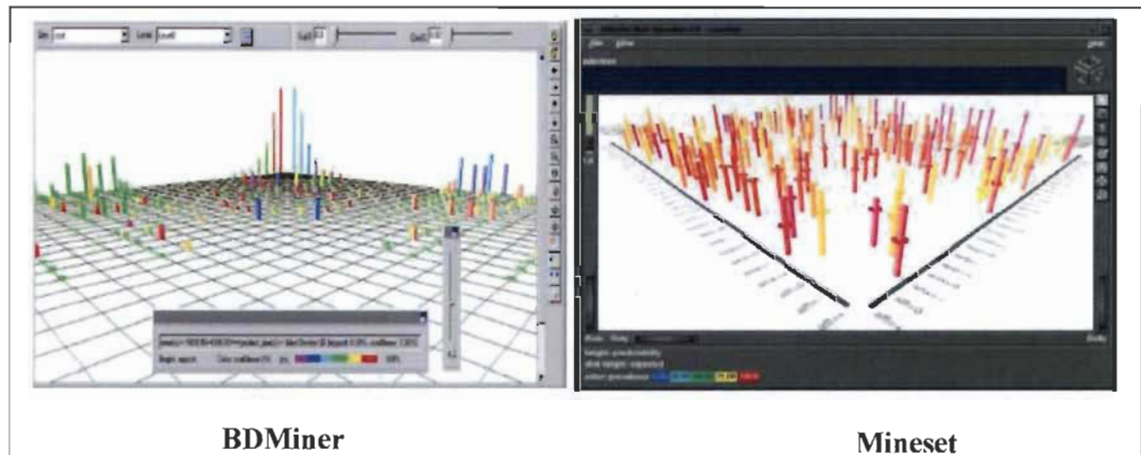


Figure 5.9 Visualisation de règles d'association par matrices [32].

### 5.5. Visualisation en mosaïque

Le diagramme en mosaïque pour visualiser des tables de contingence (mosaic display) a été adapté aux règles d'association dans [38], chaque règle étant représentée par un rectangle dont l'aire est le support et la hauteur est la confiance (voir figure 5.10).

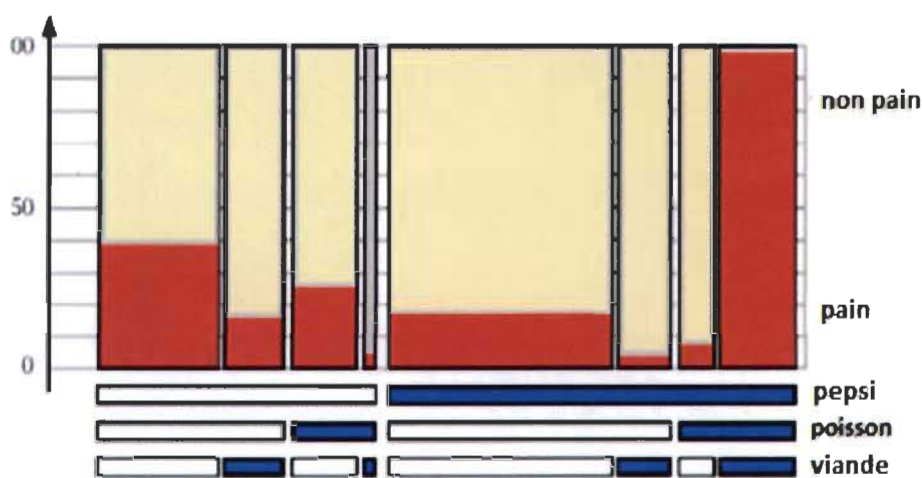


Figure 5.10 Représentation en mosaïque pour les règles [38].

Les règles représentées portent sur les items *viande*, *poisson*, et *pomme* en prémisse, et *pain* en conclusion. La prémisse et la conclusion des règles se lisent respectivement sur l'axe horizontal et sur l'axe vertical. Le premier rectangle à gauche indique que la règle  $(pepsi = 0 \wedge poisson = 0 \wedge viande = 0) \rightarrow (pain = 1)$  possède une confiance d'environ 40% (partie inférieure du rectangle), tandis que la règle contraire

$(viande = 0 \wedge poisson = 0 \wedge pepsi = 0) \rightarrow (pain = 0)$  possède une confiance d'environ 60% (partie supérieure du rectangle). Le dernier rectangle à droite indique quant à lui une règle beaucoup plus forte :  $(viande = 1 \wedge poisson = 1 \wedge pepsi = 1) \rightarrow$

(pain = 1) avec une confiance d'environ 100%. La règle (viande = 1  $\wedge$  poisson = 1  $\wedge$  pepsi = 1)  $\rightarrow$  (pain = 0) possède une confiance d'environ 0% [38][29].

Cette méthode de visualisation ne permet de représenter simultanément qu'une partie des ensembles de règles, c.-à-d. les règles ayant la même conclusion et partageant les mêmes attributs dans la prémisse.

### 5.6. Visualisation par métaphore

Blanchard et ses collègues proposent une métaphore visuelle pour représenter des sous-ensembles de règles d'association dans un espace 3D [29]. La représentation graphique choisie est le suivant (voir figure 5.11) :

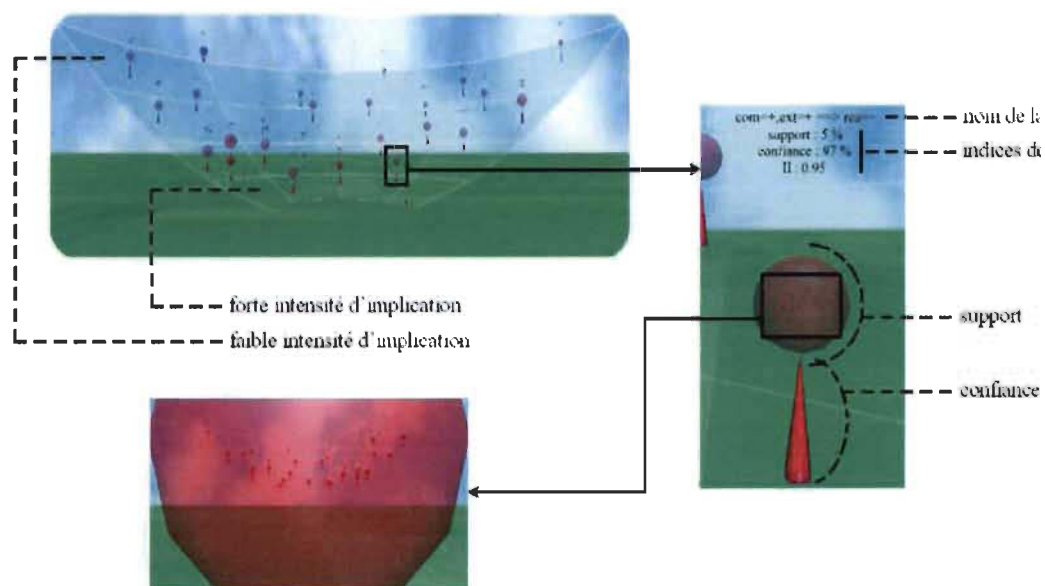


Figure 5.11 Visualisation des règles d'association à l'aide d'une métaphore [29].

Si une règle a le support, la confiance et l'intensité d'implication élevés, elle est représentée par une grande sphère rouge posée sur un haut cône à l'avant d'une arène sur les marches basses, tandis qu'une règle dont les trois mesures sont faibles, elle est représentée par une petite sphère bleue, posée sur un cône bas au fond d'une arène sur les marches hautes. Des étiquettes textuelles complémentaires sont affichées au-dessus de chaque itemset pour contenant le nom et les trois mesures de qualité de la règle [29].

### 5.7. Visualisation avec les coordonnées parallèles

Des techniques inspirées des coordonnées parallèles sont également utilisées pour visualiser des règles d'association [47]. Les variables sont représentées par des axes parallèles sur lesquels sont répartis les items, et chaque règle est symbolisée par une ligne brisée qui coupe les axes parallèles au niveau des items qu'elle contient. Les indices de règle peuvent être indiqués par l'épaisseur ou la couleur de la ligne, ou bien par des axes supplémentaires [29] [41]. La figure 5.12 illustre ce type de visualisation.

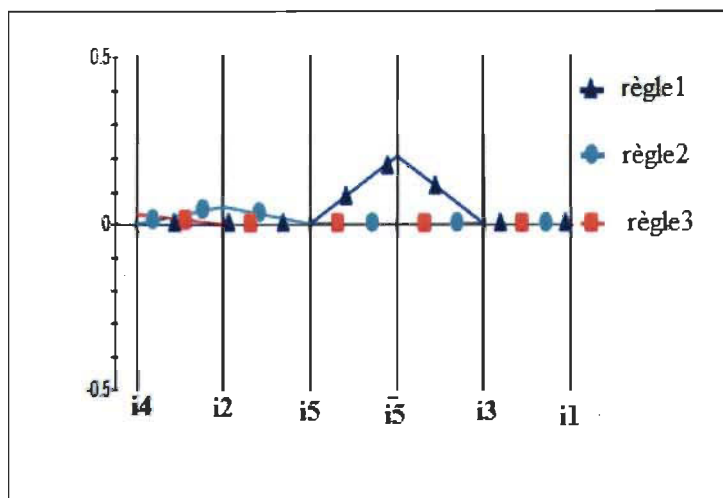


Figure 5.12 Visualisation des règles d'association avec des coordonnées parallèles [41]

## 5.8 Conclusion

Dans ce chapitre nous avons présenté diverses méthodes de visualisation qui s'appliquent aux règles d'association. Chaque méthode s'appuie sur un type de support différent tels que: tableau, graphe, matrice et cube. Bien que ces techniques visualisations règles d'association s'avèrent efficace lorsque la quantité des règles à visualiser est petite, les expériences réelles montrent qu'ils sont limitées lorsqu'il s'agit de représenter un nombre de règles trop important.

## Chapitre 6 - Implémentation

### 6.1 Introduction

Le mode de visualisation textuel adopté dans GRAMEXCO a montré ses limites quand le nombre de règles est important. Ainsi, l'état de l'art sur les visualisations graphiques présentées dans le chapitre précédent a montré aussi ses insuffisances pour représenter une grande quantité de règles.

Face à cette problématique, nous avons présenté une solution basée sur l'interactivité de l'utilisateur avec la représentation graphique des règles.

Le module de visualisation graphique des règles doit avoir les fonctionnalités suivantes :

- Pour chaque itération, le support visuel présente un nombre limité de règles.
- Les actions de l'utilisateur sur le support visuel des règles déclenchent l'exécution de l'algorithme d'extraction des règles, qui en réponse, provoque la réactualisation de la visualisation.
- Mémoriser l'historique d'exploration, pour donner à l'utilisateur la possibilité de retrouver et visualiser les règles qu'il a jugées intéressantes lors de ses explorations passées.
- Assurer un temps raisonnable de découverte de règles d'association maximales.

Dans ce chapitre, nous allons présenter l'architecture du système GRAMEXCO, le choix de langage, les modules qui le constituent et à la fin, nous allons nous focaliser sur le fonctionnement du module *représentation graphique* que nous avons développée.

## 6.2 Choix de langage

Notre système a été implémenté par CodeGear C++ builder 2009. Nous avons opté pour le langage C++ pour sa rapidité d'exécution.

L'extraction de règles d'association maximales dans le texte ainsi que la visualisation graphique du résultat nécessitent de la performance, à cause de la forte complexité calculatoire résultant de l'augmentation de combinaisons possibles entre les mots du corpus (items). L'efficacité doit être équivalente à celle d'un langage machine où le temps de réponse des algorithmes n'est pas soumis à des contraintes de compilation. En d'autres termes, la compilation doit être la plus optimale possible.

Ainsi, la forte demande de mémoire lors de calcul et d'affichage graphique, oblige le logiciel à bien gérer sa propre utilisation des ressources en mémoire dynamique afin d'optimiser l'utilisation et pour éviter d'éventuels craches du système.

La technique de polymorphisme et de l'algèbre abstraite nous permettrait de réduire la mémoire vive utilisée, en supposant la symétrie entre (voir Figure 6.1) :

- Le groupe des mots et le groupe de leurs références en mémoire (adresse physique).
- Le groupe des opérations internes (opérations) sur les mots et le groupe de leurs adresses en mémoire.



- Le groupe des opérations externes (fonctions) sur les mots et le groupe de leurs adresses en mémoire.

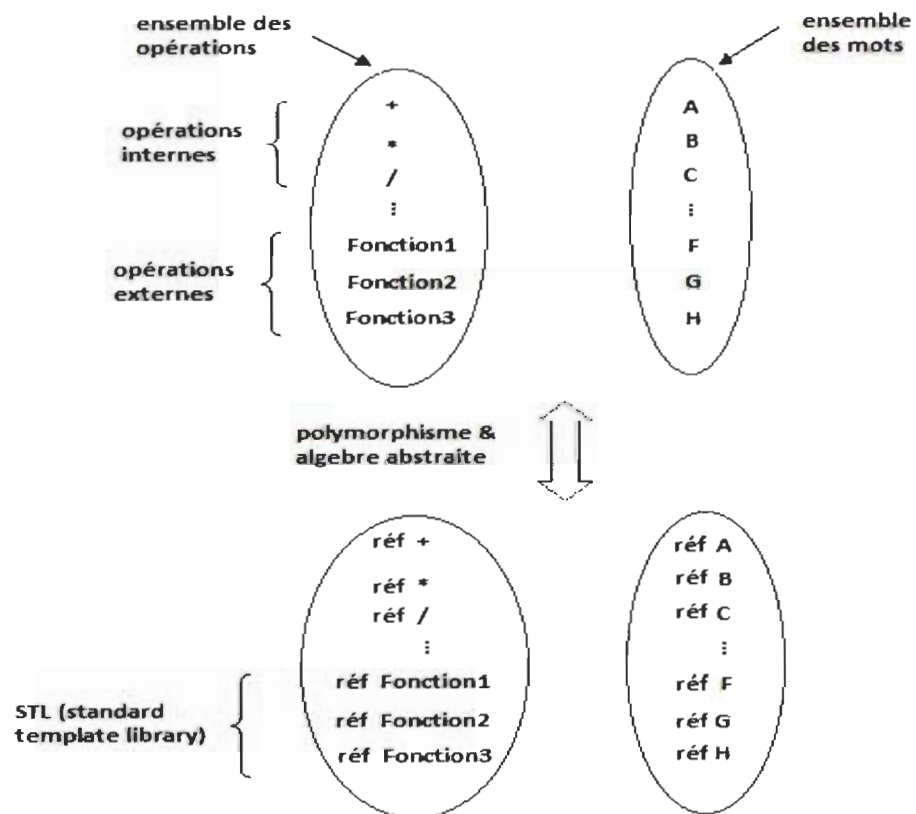


Figure 6.1 Combinaison de l'algèbre abstraite et du polymorphisme.

Si on prend par exemple, le mot *INFORMATIQUE*, il occupe en UNICODE  $12 \times 4$  octets = 48 octets de mémoire. Si ce mot est présent 10 fois dans un corpus, au total, on doit réserver  $10 \times 48 = 480$  octets de mémoire.

Cette approche nous permet de réserver uniquement 46 octets de mémoire (42 octets pour le mot *INFORMATIQUE* et 4 octets (référence ou pointeur a ce mot).

Soit un corpus de  $N$  mots. La vitesse d'un algorithme naïve de calcul des règles d'association maximales est exponentielle en fonction du nombre mots  $N$  : ordre de  $N^4$ .

La vitesse d'un algorithme exploitant la puissance du langage C++ (La *Standard Template Library* -STL) est de l'ordre de  $N^3 \text{Log}(N)$ . Cependant, la vitesse de notre algorithme est de l'ordre de  $N * \text{Log}(\text{Log}(N))$ .

### 6.3 Architecture du système

L'architecture de notre système repose sur un découpage modulaire. En effet, cette particularité rend possible la modification d'un module sans que cela n'entraîne de modifications des autres.

Ces modules sont : analyse du texte, classification numérique, calcul des règles d'association maximales et le module de visualisation graphique (voir figure 6.2). Notons que les trois premiers modules sont développés par d'autres membres de notre laboratoire. Dans le but de l'optimisation, nous avons reprogrammé les deux modules : "*classification numérique*" et "*calcule des règles d'association maximales*".

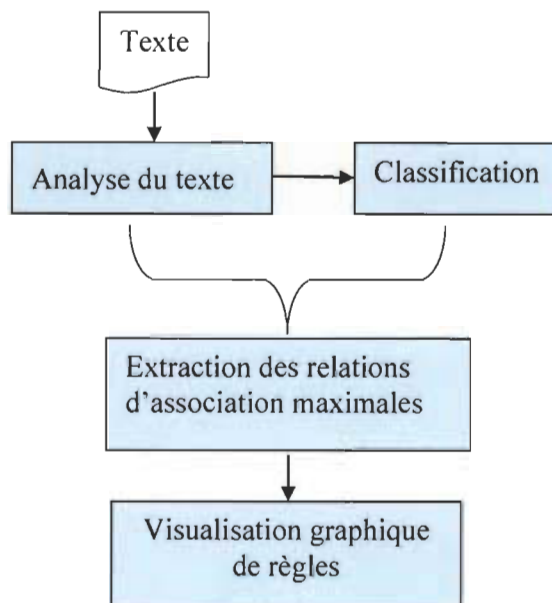


Figure 6.2 Architecture du système développé.

Dans ce qui suit, nous allons donner plus de détails et expliquer le fonctionnement de chaque module de notre système.

### 6.3.1. Analyse du texte

Ce module permet de découper le corpus en unités (N-Grams) et domaines d'information (segments de textes dont on veut comparer la similarité) pour faciliter la classification.

Les unités d'information peuvent prendre la forme de bi-grams, de tri-grams, de quadri-grams, etc. [49].

On définira un n-gram de caractères par une suite de n caractères : bi-grams pour n=2, tri-grams pour n=3, quadri-grams pour n=4, etc. Par exemple pour le mot *classification* les tri-grams sont : *cla, las, ass, ssi, sij, ifi, fic, ica, cat, ati, tio, ion* [6].

La segmentation d'un texte peut prendre plusieurs formes. Ainsi, on peut diviser un document en mots, paragraphes, phrases, ou selon de textes délimités par un caractère ou une chaîne de caractères.

Les deux opérations (découpage du texte en segments et N-grams) se faisant simultanément, nous récupérons en sortie une matrice où sont répertoriées les fréquences d'apparitions de chaque N-Gram dans chaque segment.

Le choix des paramètres ci-dessous est fait par l'utilisateur en fonction de ses propres objectifs :

- La taille du N-gram.
- Le type du segment.
- La suppression des N-grams fonctionnels, des N-grams avec espace ou des N-grams dont la fréquence est inférieure ou égale à un certain intervalle choisi par l'utilisateur.
- La suppression des N-grams dont la fréquence totale dans le corpus se situe dans un intervalle choisi par l'utilisateur (nettoyage par fréquence totale).
- La suppression des N-grams dont la fréquence totale dans chaque segment se situe dans un intervalle choisi par l'utilisateur (nettoyage par fréquence relative).

### 6.3.2. Classification numérique

Ce module prend en entrée la matrice générée par le module "Analyse de texte", afin d'organiser les segments en classes (*clusters*), dont les membres ont en commun la propriété suivante : les segments qui contiennent les mêmes N-Grams avec des occurrences qui sont presque identiques seront regroupés dans la même classe. Cependant, le lexique de ces segments forme le vocabulaire des classes auxquelles ils appartiennent.

Nous avons utilisé le classifieur non-supervisé *K-means*, ainsi généralement traduit par *K-moyennes*. Sa simplicité de mise en œuvre et sa capacité de traiter un grand ensemble de données font la popularité de cette approche, c'est l'algorithme le plus connu et le plus utilisé dans les applications scientifiques et industrielles de clustering. Il partitionne les segments en  $K$  classes. L'algorithme renvoie une partition des données, dans lequel les segments à l'intérieur de chaque classe sont aussi proches que possible les uns des autres et aussi loin que possible des segments des autres classes. La rapidité de *K-means* est un avantage notable : il ne compare pas tous les objets entre eux, mais par rapport aux centres des classes.

Le principal problème de l'approche *K-means*, est l'influence de la partition initiale (le choix du paramètre  $K$  qui n'est pas toujours évident et qui est souvent choisi de façon aléatoire).

Nous donnons ci-dessous une description sommaire de l'algorithme *K-means* utilisé dans notre système :

On fournit en entrée une matrice des fréquences de  $d$  N-grams pour l'ensemble de  $L$  segments, comme illustré dans le tableau 6.1.

$S/N$	$N_1$	...	$N_j$	...	$N_d$
$S_1$	$N_{11}$	...	$N_{1j}$	...	$N_{1d}$
$\vdots$	...	...	...	...	...
$S_i$	$N_{i1}$	...	$N_{ij}$	...	$N_{id}$
$\vdots$	...	...	...	...	...
$S_L$	$N_{L1}$	...	$N_{Lj}$	...	$N_{Ld}$

Tableau 6.1 Matrice contenant les fréquences des N-grams dans les segments.

La case  $N_{ij}$  représente la fréquence du N-gram  $N_j$  dans le segment  $S_i$ , chaque segment  $S_i$  est représenté comme un vecteur à  $d$  composantes  $(N_{i1}, \dots, N_{id})$ .

Soit  $dist$  une distance euclidienne au carré entre objets

$$dist(\vec{S}_1, \vec{S}_2) = \|\vec{S}_1 - \vec{S}_2\|^2 = \sum_{j=1}^d (N_{1j} - N_{2j})^2 \quad (6.1)$$

$\mu$  la fonction centre de gravité (ou centroïde, ou moyenne) d'un ensemble de vecteurs.

$$\vec{\mu}(C_k) = \frac{\sum_{\vec{S} \in C_k} \vec{S}}{|C_k|} = \vec{\mu}(C_k) = \frac{\sum_{S_i \in C_k} N_{ij}}{|C_k|} \quad (6.2)$$

Soient  $\vec{m}_1^{(0)}, \vec{m}_2^{(0)}, \dots, \vec{m}_k^{(0)}$   $K$  centroïdes initiaux, définis via une partition initiale

arbitraire de segments en  $K$  clusters

**Pour**  $k$  de 1 à  $K$  faire : initialiser un cluster  $c_k$  vide

FIN = FAUX

**TANT QUE** NON FIN **FAIRE**

**POUR** chaque  $C_k$  **FAIRE**

Calculer les similarités entre  $S_i$  et les  $m_k$

Ajouter  $S_i$  au cluster  $c_k$  tel que la distance entre  $m_k$  et  $S_i$  est minimale

**FIN POUR**

**SI** aucun changement des valeurs des centroïdes **FAIRE**

FIN = VRAI

**SINON**

**POUR** pour chaque  $C_k$  **FAIRE**

recalculer  $m_k$  comme le centroïde de nouveau cluster

$$\vec{m}_k := \mu(c_k)$$

**FIN POUR**

**FIN SI**

**FIN TANT QUE**

La figure 6.3 montre une prise d'écran des résultats obtenus suite à une classification de texte de 6 segments en 3 clusters avec la méthode K-means.

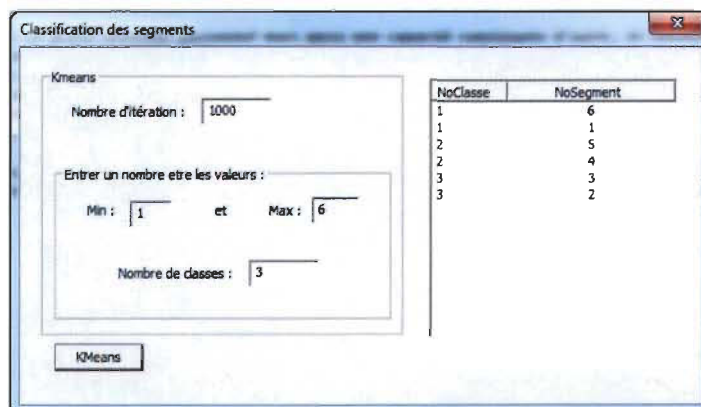


Figure 6.3 Partitionnement des segments retenus du corpus en 3 classes

### 6.3.3. Extraction des règles d'association maximales [50]

Les classes générées par le module précédent seront utilisées comme des transactions pour extraire les règles d'association maximales, visant à extraire des connaissances cachées, nouvelles et potentiellement utiles. L'intervention de l'utilisateur est nécessaire pour déterminer en premier lieu le mot pour lequel il veut trouver les associations les plus vraisemblables.

L'exemple suivant permettra de bien comprendre le processus d'extraction des règles d'association maximales  $X \xrightarrow{\text{max}} Y$  à partir des classes générées par le système:

Soit  $E1 = \{x, a, b, c, d, e, f\}$  l'ensemble du lexique de notre corpus. La classification donne en sortie les classes avec leur lexique respectif :  $C1 : \{x, a, b, c\}$ ,  $C2 : \{a, c, d\}$ ,  $C3 : \{x, e, f, d\}$ .

Le processus d'extraction et de calcul des règles d'association maximales à partir de ces classes se déroule en trois étapes :

**(1) Choix de l'ensemble X :** c'est l'utilisateur qui choisit parmi la liste des éléments de  $E1$  le lexique qui va représenter notre  $X$ . Supposons pour les besoins de l'explication  $X = \{x\}$ .

**(2) Identification de l'ensemble Y et de l'ensemble E2 :** l'identification de l'ensemble catégorie  $E2$  dans lequel  $Y$  serait un sous-ensemble dépend fortement de l'ensemble  $X$  choisi et des classes dont  $X$  est un sous-ensemble.

Dans le cas de notre illustration,  $X$  est inclus dans  $C1$  et dans  $C3$ .  $Y$  peut alors être un sous-ensemble soit de  $\{a, b, c\}$  soit de  $\{e, f, d\}$ . Autrement dit que,  $Y$  peut représenter un des sous-ensembles suivants :  $\{a\}$   $\{b\}$   $\{c\}$   $\{a, b\}$   $\{a, c\}$   $\{b, c\}$   $\{a, b, c\}$   $\{e\}$   $\{f\}$   $\{d\}$   $\{e, f\}$



$\{e, d\}$ .  $\{F, d\}$   $\{e, f, d\}$ . Les mesures du M-Support et de la M-confiance seront calculées par rapport à ces différentes possibilités de valeurs de Y. Un processus itératif permettra de tester l'ensemble de ces possibilités. Nous pouvons, toutefois, limiter le nombre d'itérations pour éviter un coût computationnel trop prohibitif. Par exemple en fixant (au moyen d'un paramètre) la cardinalité du sous-ensemble Y. Supposons que  $Y = \{a, c\}$ , pour construire E2 nous devons dans un premier temps établir les catégories respectives des éléments a et c. celles-ci sont obtenues à travers l'union des classes qui contiennent a (respectivement c). Suite à quoi  $E2 = \text{catégorie}(Y) = \text{catégorie}(\{a, c\})$  sera obtenu par l'intersection de la catégorie (a) avec la catégorie (c). Ainsi :

$$\text{catégorie}(a) = \{a, b, c\} \cup \{a, c, d\} = \{a, b, c, d\} \quad \text{et}$$

$$\text{catégorie}(c) = \{a, b, c\} \cup \{a, c, d\} = \{a, b, c, d\} \quad \text{donc :}$$

$$E2 = \text{catégorie}(Y) = \text{catégorie}(a, c) = \text{catégorie}(a) \cap \text{catégorie}(c) = \{a, b, c, d\}$$

(3) dès lors que les ensembles E1, E2, X et Y ainsi que les transactions ont été clairement identifiés, le calcul des mesures peut se faire.

Considérons l'association  $x \xrightarrow{\text{max}} a, c$ . En utilisant les classes C1 :  $\{x, a, b, c\}$ , C2 :  $\{a, c, d\}$ , C3 :  $\{x, e, f, d\}$  comme transactions, et  $E2 = \{a, b, c, d\}$ , il en découle un M-support égal à 1, puisque seulement la classe 1 contient  $X = \{x\}$  et  $Y = \{a, c\}$ , et une M-confiance de 0.5 puisque deux classes contiennent X alors qu'une seule contient X et Y.

### 6.3.4. Visualisation graphique du résultat

#### 6.3.4.1. Description de fonctionnement

Le lancement du module affichera à l'utilisateur une fenêtre principale. La figure 6.4 montre une prise d'écran de cette dernière.

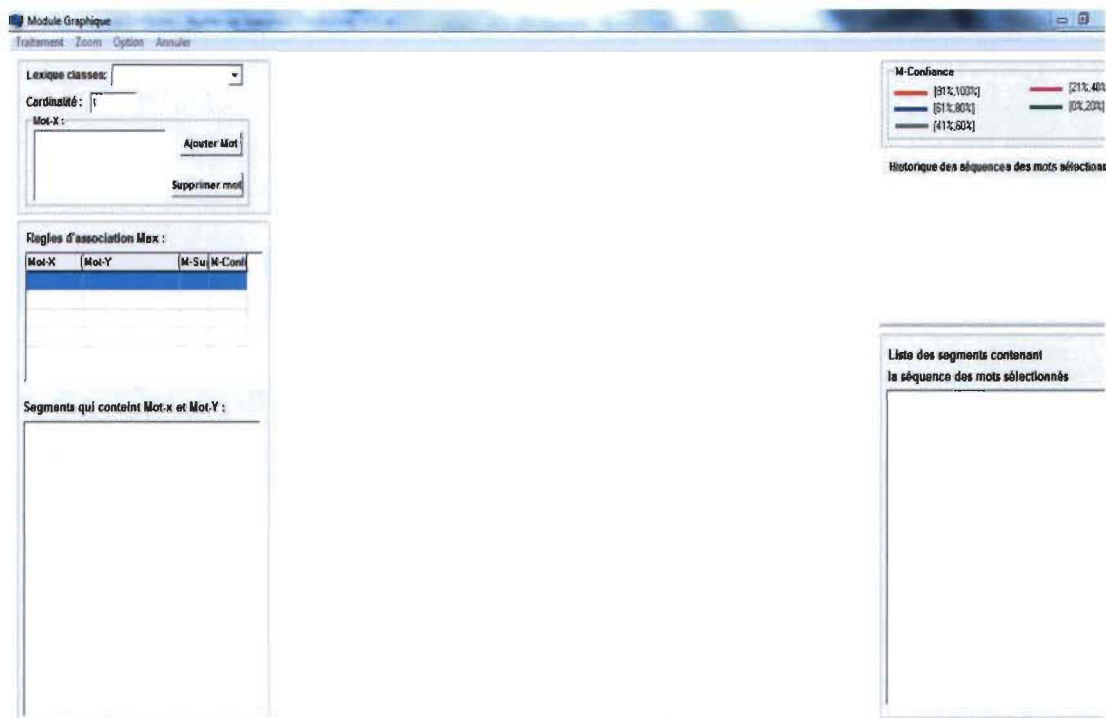


Figure 6.4 Fenêtre principale du module graphique.

#### 6.3.4.2. Acquisition des données

Le module prend en entrée le lexique des classes trouvées par le classifieur K-means ainsi que les différents segments qui leur appartiennent.

Comme on peut le voir au niveau de la figure 6.5, l'utilisateur a la possibilité de choisir un itemset en sélectionnant un mot à partir de la liste de lexique des classes, puis en appuyant sur le bouton « Ajouter Mot », le mot sera ajouté dans la liste des mots qui vont

construire l'itemset  $X$  (prémisse des associations), notant que le système prend en compte le cas où  $X$  est composé de plusieurs mots, il peut supprimer un mot en le sélectionnant d'abord et en cliquant ensuite sur le bouton « Supprimer mot ».

La cardinalité est par défaut initialisée à 1, cela signifie que le système prend en considération uniquement les sous-ensembles d'itemsets  $Y$  (conclusions des associations) qui sont constitués par un seul mot, mais l'utilisateur a la possibilité d'augmenter la valeur de la cardinalité, si par exemple la cardinalité est égale à deux, le système génère les différentes combinaisons un à un et un à deux pour l'itemset  $Y$  de l'association  $X \xrightarrow{\text{max}} Y$ . La performance de l'algorithme diminue proportionnellement à l'augmentation de la cardinalité.

The screenshot shows a user interface with the following elements:

- A dropdown menu labeled "Lexique classes:" with "COMPORTEMENT" selected.
- An input field labeled "Cardinalité:" containing the number "1".
- A section labeled "Mot-X:" containing a list with two items: "CAPACITÉ" and "COMPORTEMENT".
- A button labeled "Ajouter Mot" next to the list.
- A button labeled "Supprimer mot" below the list.

Figure 6.5 Interface permettant de saisir le Mot-X et le nombre de cardinalité.

Le système vérifiera s'ils existent des classes qui contiennent la combinaison des mots qui constituent l'itemset  $X$ . Dans le cas négatif, la combinaison ne sera pas considérée et l'utilisateur en sera averti.

Afin d'éviter une représentation graphique surchargée, le système offre à l'utilisateur deux façons pour limiter la quantité des règles d'association à afficher dans le graphe, la

figure 6.6 montre une prise d'écran de la boîte de dialogue qui permet de fixer ce nombre de mots Y à afficher.

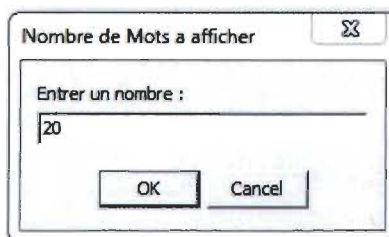


Figure 6.6 Limitation de règles à afficher par fixation du nombre d'itemsets Y.

La figure 6.7 représente une deuxième façon d'afficher uniquement les associations dont les *M-Confiances* sont inclus dans un intervalle choisi par l'utilisateur.

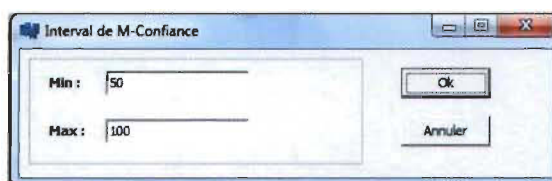


Figure 6.7 Limitation des règles à afficher par choix d'intervalle de M-Confiance.

#### 6.3.4.3. Génération de règles d'associations maximale

Une fois, la prémisse, la valeur de cardinalité et le nombre d'associations à afficher sont choisis, l'extraction de règles d'association maximales peut se faire en cliquant sur le menu [Traitement >> Calcul M-Support et M-Confiance]. Le résultat de calcul s'affiche sous deux formats différents.

### 6.3.4.3.1. Représentation textuelle :

La figure 6.8 montre une prise d'écran de ce type d'affichage.

Chaque ligne de tableau représente une association  $X \xrightarrow{\max} Y$ , dont l'élément de la 1ère colonne représente la *prémisse*, la 2ème colonne représente la *conclusion*. La 3ème et la 4ème colonne représentent respectivement le *M-Support* et le *M-confiance* de l'association.

X	Y	m support	m confiance
ROI	TUER	1	12
ROI	URBAIN	3	37
ROI	ÉMEUTE	2	25
ROI	VICTIME	1	12
ROI	TRAUMATISM	1	12
ROI	SATELLITE	1	12
ROI	TOURISME	1	12

Figure 6.8 Représentation sous forme textuelle.

### 6.3.4.3.2. Représentation graphique

Comme la figure 6.9 la montre, chaque règle  $X \xrightarrow{\max} Y$  est représentée par un arc reliant deux nœuds, dont le nœud situé au centre du cercle contient la prémisse de la règle (*mot X*) et le 2ème nœud contient la conclusion de la règle (*mot Y*). Lorsqu'on positionne la souris sur un *mot Y*, une description textuelle s'affiche pour indiquer le M-support et M-confiance de la règle.

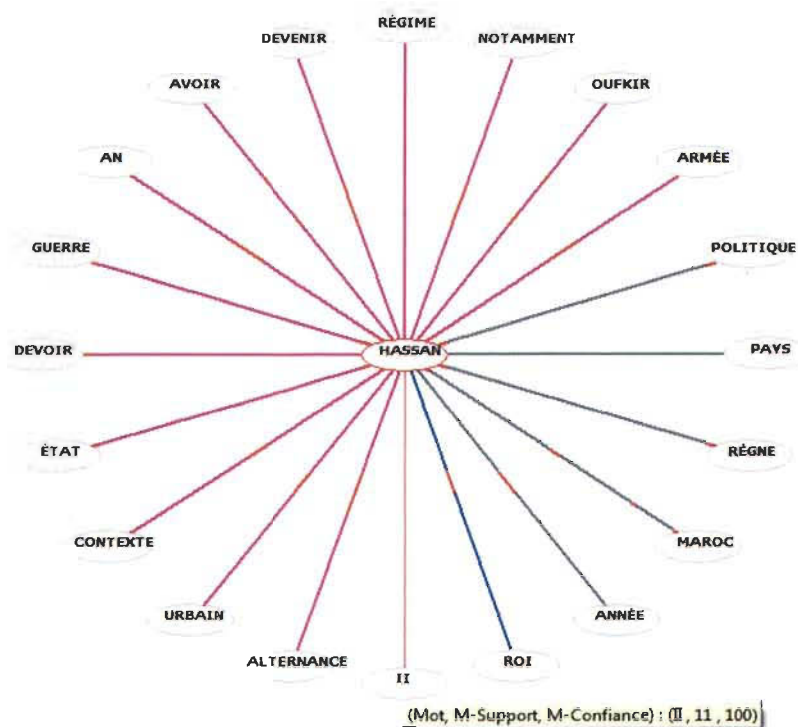


Figure 6.9 Représentation graphique des associations.

L'arc peut prendre différentes couleurs, chaque couleur correspond à un intervalle de *M-confiance*, cela permet à l'utilisateur de cibler directement les associations dont leur *M-confiance* appartient à un intervalle précis. La figure 6.10 présente bien la relation entre la couleur de l'arc et le *M-confiance*.

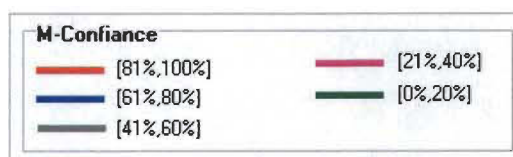


Figure 6.10 Couleur de l'arc selon le M-confiance.

#### 6.3.4.4. Approche dynamique et interactive

L'utilisateur peut poursuivre le processus de recherche des règles d'association maximales de façon itérative, en positionnant la souris sur un mot et en cliquant sur bouton droit, un menu contextuel apparaîtra permettant de déclencher l'algorithme de recherche de règles d'associations avec ce mot qui en retour provoque une réactualisation de la visualisation graphique.

D'abord, le système mémoriser les informations (deux parties de la règle, M-support et M-confiance) de l'association de ce mot avec l'ancien mot  $X$  dans une liste. Puis, le système nous redirigera vers un nouveau graphe avec le même principe de visualisation graphique qu'on a citée précédemment, et la représentation est mise à jour. Un rectangle va s'ajouter pour contenir l'information enregistrée dans la liste (voir figure 6.11).

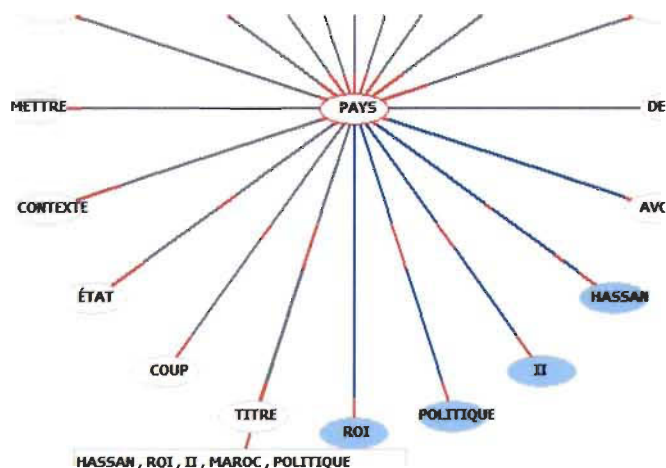


Figure 6.11 Chemin parcouru dans la fouille.

Après une séquence de requêtes de calcul de règles d'association maximales, le rectangle va contenir une séquence des mots (*séquence des mots sélectionnés*) qui ont été soumis à un processus d'extraction de règles d'association maximales.

Comme illustrée sur la figure 6.11, la séquence des mots sélectionnés (*Hassan, Roi, II, Maroc, Politique*), s'interprète comme suit :

$HAASAN \xrightarrow{\max} ROI$  : association avec une confiance de 72 % ;

$ROI \xrightarrow{\max} II$  : association avec une confiance de 100 % ;

$II \xrightarrow{\max} MAROC$  : association avec une confiance de 54 % ;

$MAROC \xrightarrow{\max} POLITIQUE$  : association avec une confiance de 40 % ;

$POLITIQUE \xrightarrow{\max} PAYS$  : association avec une confiance de 62 % .

La couleur bleue du nœud (voir figure 6.11) indique à l'utilisateur que le mot  $Y$  qu'il contient a été déjà soumis au processus de calcul des règles d'association (fais partie de la séquence des mots sélectionnés). Dans ce cas, le système juge qu'il est inutile de revenir à une étape antérieure. Un message d'erreur apparaîtra lorsque l'utilisateur tente de sélectionner ce mot  $Y$ .

#### 6.3.4.5 Historique d'exploration

L'exploration des règles peut également être facilitée en procurant à l'utilisateur une liste de la navigation qu'il a effectuée. Ce dispositif de mémorisation des règles permettrait, d'une certaine façon, de donner à l'utilisateur une vue globale de l'ensemble de règles qu'il



visite. Qui peut par la suite retrouver, visualiser et comparer les règles qu'il a jugées intéressantes lors de ses explorations passées (voir la figure 6.12)

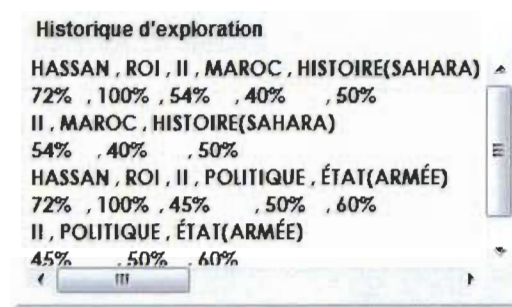


Figure 6.12 Historique d'exploration.

#### 6.3.4.6. Annulation et restauration

Le système donne la possibilité de revenir à une séquence des mots sélectionnés, en d'autres termes, le système permet de reculer ou d'avancer dans le chemin parcouru. Soit de faire un retour en arrière en cliquant sur le menu [annuler], soit de choisir un point de restauration, en sélectionnant le mot de la séquence des mots sélectionnés à laquelle l'utilisateur veut revenir et en cliquant sur le bouton droit de la souris. Avant que la restauration s'effectue, le système sauvegarde le chemin parcouru complet et le chemin parcouru à partir du point de récupération et les affiche dans une liste, qu'on a nommée « Historique d'exploration ». Pour récupérer un chemin, il suffit de le sélectionner (voir la figure 6.13).

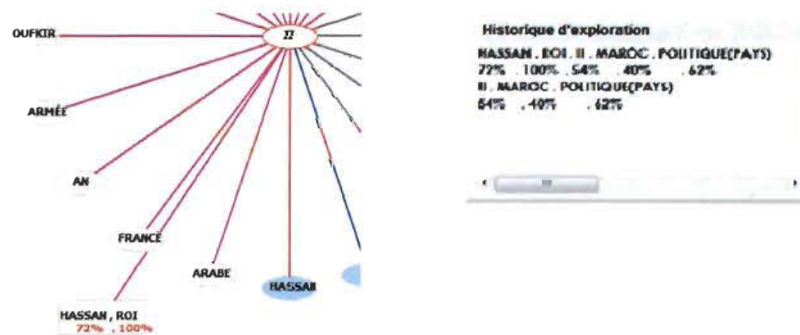


Figure 6.13 Sauvegarde et restauration de chemin parcouru.

#### 6.3.4.7. Recherche de segments

L'utilisateur peut afficher les segments qui contiennent le mot  $X$  et le mot  $Y$ , en sélectionnant ce dernier, puis en cliquant sur l'option « Afficher segments » du bouton droit de la souris. La figure 6.14 montre le résultat de la recherche des segments qui contiennent les deux mots : PAYS et ETAT.

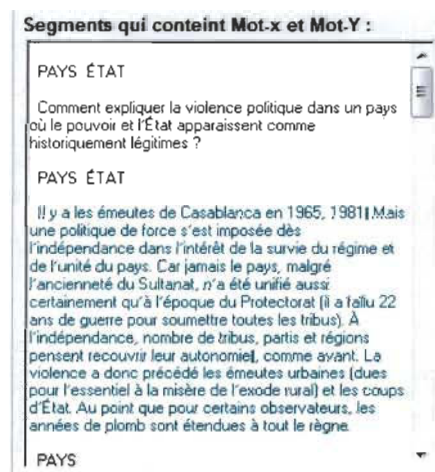


Figure 6.14 Segments contenant  $Mot-X$  et  $Mot-Y$ .

Le système permet aussi d'afficher les segments qui contiennent la séquence des mots sélectionnés, comme illustrés sur la figure 6.15, le système les affiche par ordre décroissant selon le nombre de mots de la séquence qui sont inclus dans le segment.

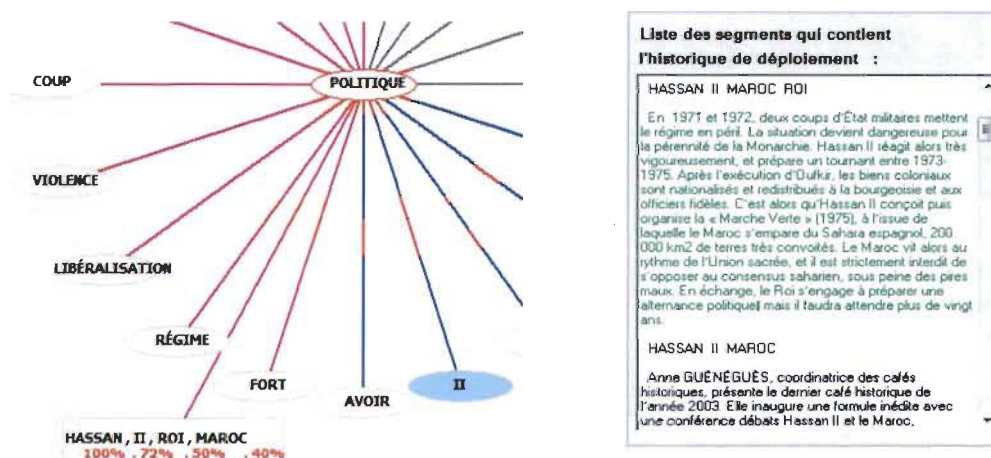


Figure 6.15 Segments contenant des mots de la séquence des mots sélectionnés.

## 6.4 Conclusion

Dans ce chapitre, nous avons présenté notre outil de fouille de règles d'association maximales. Le système comporte cinq modules fortement interconnectés : analyse du texte, classification numérique, calcul des règles d'association maximales et visualisation graphique des règles. Nous avons présenté d'une manière détaillée le fonctionnement du module de visualisation graphique qui permet d'assister l'utilisateur dans son processus de fouille au sein d'une grande quantité de règles et de diriger sa navigation en visualisant à chaque itération une quantité limitée de règles afin d'identifier les informations pertinentes.

## Chapitre 7 - Expérimentations et résultats

Ce chapitre présente la méthode d'évaluation qui a été utilisée ainsi que les résultats qui ont été produits par les tests.

En effet, notre logiciel, offre des outils de manipulation des données qui aident à la lecture, l'analyse et la compréhension rapide des textes, il permet de réaliser un doucement explicatif (interprétation subjective) du contenu du texte original construit au travers de ce que l'utilisateur a compris en explorant les segments les plus pertinents correspondants aux règles d'association maximales de bonne qualité.

L'utilisation d'une méthode objective pour déterminer la qualité des résultats produits par notre système n'étant pas possible. En effet, ce qui est jugé pertinent pour un utilisateur ne l'est pas nécessairement pour un autre.

Pour justifier la pertinence de nos résultats, nous nous sommes appuyés sur l'analyse des deux critères suivants :

- temps de lecture manuelle du texte original par rapport au temps investi dans la navigation et l'extraction de l'information pertinente en utilisant le logiciel.
- Cohérence et pertinence de l'information extraite par rapport au texte original.

Nous avons travaillé sur des fichiers qui ne dépassent pas 20 pages, afin de pouvoir réaliser des tests rapides. Ces fichiers sont codés en Unicode (codage qui prend en considération plusieurs langues tel que l'arabe).

L'environnement de tests a les caractéristiques suivantes :

- PC avec microprocesseur DUAL-CORE 2.30 GHz
- Mémoire vive à 2.00 GB
- Système d'exploitation Windows 7

### **7.1. Exemple 1 : « 2011 : le printemps arabe ? »**

Les données étudiées dans cet exemple proviennent du texte « 2011 : le printemps arabe ? » de Samir Amin, économiste marxiste franco-égyptien, publié le 17 mai 2011 sur le site *Europe solidaire sans frontières* [57] qui éclaire selon son point de vue les transformations en cours en Égypte.

Nous avons opté pour les paramètres de classifications suivants :

#### **Analyse du texte :**

- Une segmentation par "phrase "
- 3 caractères pour la taille des N-gram.
- Les lettres minuscules sont identiques aux lettres majuscules.
- La suppression des N-gram avec espace.

#### **Classification :**

- Classification du texte avec le classifieur K-means.

- Division des segments en 30 classes.

Dans notre exemple, nous avons débuté notre fouille par choisir l'itemset "*PRINTEMPS ARABE*" qui figure dans le titre de texte et qui semble le plus pertinent. Bien évidemment, le choix de l'itemset dépend de la subjectivité d'individu. Nous avons opté pour visualiser 25 règles d'associations qui possèdent les valeurs de M-confiance les plus élevées.

Afin de découvrir les itemsets qui sont en associations maximales avec l'itemset "*PRINTEMPS ARABE*", nous avons déclenché l'algorithme de découverte des règles d'association maximales qui a donné en retour le graphe suivant (figure 7.1) :

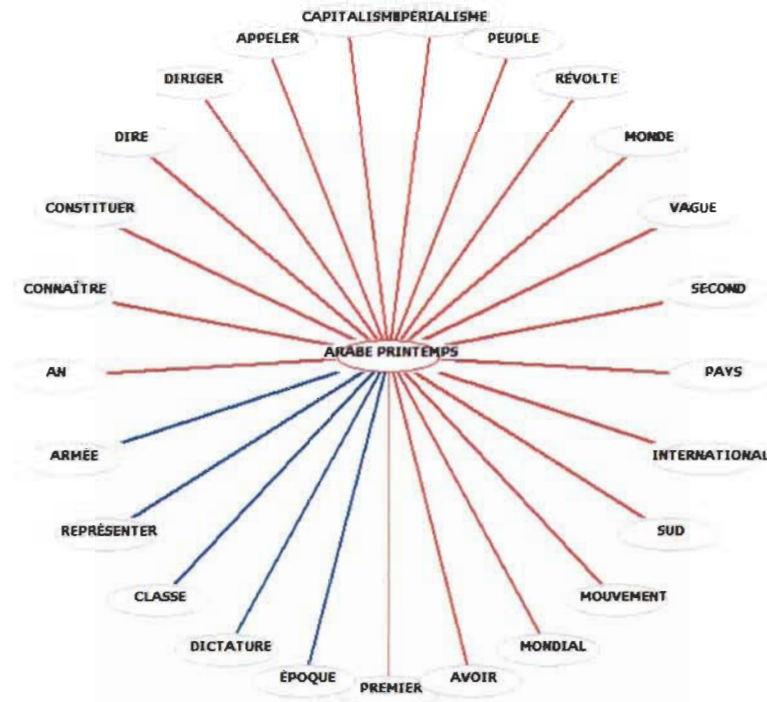


Figure 7.1 Extraction des règles pertinentes avec "*PRINTEMPS ARABE*" en prémisse.

Les arcs de couleur rouge désignent des règles qui possèdent en effet de très bonnes valeurs de M-confiance (de 80% à 100%).

Nous montrons dans le tableau 7.1 la cohérence de ces règles avec le contenu du texte original

Règle, M-confiance	Analyse de la règle
$\begin{aligned} & \textit{PRINTEMPS ARABE} \\ & \xRightarrow{\textit{max}} \textit{MONDE}, 100\% \\ & \textit{ARABE} \xRightarrow{\textit{max}} \textit{MONDIAL}, 100\% \\ & \textit{PRINTEMPS ARABE} \\ & \xRightarrow{\textit{max}} \textit{INTERNATIONAL}, 100\% \end{aligned}$	<p>L'itemset "<i>PRINTEMPS ARABE</i>" est fortement lié avec les l'itemsets : "<i>MONDE</i>", "<i>MONDIAL</i>" et "<i>INTERNATIONAL</i>".</p> <p>Ces résultats sont justifiés puisque le printemps arabe est un événement <b>mondial</b> et <b>international</b>.</p>
$\begin{aligned} & \textit{PRINTEMPS ARABE} \\ & \xRightarrow{\textit{max}} \textit{PAYS}, 100\% \\ & \textit{PRINTEMPS ARABE} \\ & \xRightarrow{\textit{max}} \textit{SUD}, 100\% \end{aligned}$	<p>La règle montre l'étroite relation entre l'itemset "<i>PRINTEMPS ARABE</i>" et l'item "<i>PAYS</i>". En effet le terme printemps arabe a été proposé pour désigner des événements connus par plusieurs <b>pays</b> du <b>sud</b> de la planète</p>
$\begin{aligned} & \textit{PRINTEMPS ARABE} \\ & \xRightarrow{\textit{max}} \textit{MOUVEMENT}, 100\% \\ & \textit{PRINTEMPS ARABE} \\ & \xRightarrow{\textit{max}} \textit{RÉVOLTE}, 100\% \\ & \textit{PRINTEMPS ARABEE} \\ & \xRightarrow{\textit{max}} \textit{PEUPLE}, 100\% \\ & \textit{PRINTEMPS ARABE} \\ & \xRightarrow{\textit{max}} \textit{IMPERIALISME}, 100\% \\ & \textit{PRINTEMPS ARABE} \\ & \xRightarrow{\textit{max}} \textit{CAPITALISME}, 100\% \end{aligned}$	<p>L'itemset "<i>PRINTEMPS ARABE</i>" a une forte relation avec les itemsets : "<i>MOUVEMENT</i>", "<i>RÉVOLTE</i>", "<i>PEUPLE</i>", "<i>IMPERIALISME</i>" et "<i>CAPITALISME</i>". En effet le printemps arabe est un <b>mouvement</b> de <b>révolte</b> sociale et pacifique du <b>peuple</b> arabe anti-<b>capitalisme</b> et anti-<b>impérialisme</b></p>

Tableau 7.1 Cohérence de résultat obtenu pour la prémisse "*PRINTEMPS ARABE*" avec le texte original.

Pour chacune de ces règles, nous évaluons sa qualité et sa pertinence, en consultant les segments qui contiennent les parties de la règle.

Prenons par exemple les deux règles :  $PRINTEMPS\ ARABE \xRightarrow{max} PEUPLE$  et  $PRINTEMPS\ ARABE \xRightarrow{max} IMPÉRIALISME$ , le tableau 7.2 illustre les segments fouillés correspondants à ces deux règles

$PRINTEMPS\ ARABE \xRightarrow{max} PEUPLE$

Les « printemps » des peuples arabes, comme ceux que les peuples d'Amérique latine connaissent depuis deux décennies, que j'appelle la seconde vague de l'éveil des peuples du Sud – la première s'était déployée au 20 ième siècle jusqu'à la controffensive du capitalisme/impérialisme néo libéral – revêt des formes diverses allant des explosions dirigées contre les autocraties qui ont précisément accompagné le déploiement néo-libéral à la remise en cause de l'ordre international par les « pays émergents »

2011 : LE **PRINTEMPS ARABE**

Le « **printemps arabe** » s'inscrit dans cette réalité

L'Égypte, c'est-à-dire son **peuple**, ses élites, la nation qu'elle représente, n'a jamais accepté ce statut

Ce mouvement gigantesque du **peuple** égyptien associe trois composantes actives : les jeunes « re-politisés » par leur propre volonté et dans des formes « modernes » qu'ils ont inventées, les forces de la gauche radicale, celles rassemblées par les classes moyennes démocrates

Les Frères musulmans dont les dirigeants avaient appelé à boycotter les manifestations pendant les quatre premiers jours (persuadés que celles-ci seraient mises en déroute par la répression) n'ont accepté le mouvement que tardivement, lorsque l'appel, entendu par l'ensemble du **peuple** égyptien, a produit des mobilisations gigantesques de 15 millions de manifestants

L'appel à la manifestation formulé par les trois composantes actives du mouvement a été rapidement entendu par l'ensemble du **peuple** égyptien

La généralisation du mouvement à l'ensemble du **peuple** égyptien constitue par elle-même un défi positif

Car ce **peuple** est, comme tous les autres, loin de constituer un « bloc homogène »



Cette bourgeoisie est compradore (dans la langue politique courante en Égypte le **peuple** les qualifie de « parasites corrompus »)

La direction est constituée exclusivement d'hommes immensément riches (grâce, entre autre, au soutien financier de l'Arabie Saoudite, c'est-à-dire de Washington), l'encadrement par des hommes issus des fractions obscurantistes des classes moyennes, la base par des gens du **peuple** recrutés par les services sociaux de charité offerts par la confrérie (et financés toujours par l'Arabie Saoudite), tandis que la force de frappe est constituée par les milices (les baltaguis) recrutés dans le lumpen

Ils feront tout pour que l'Égypte ne soit pas démocratique et, à cette fin, soutiendront, par tous les moyens, mais avec hypocrisie, la fausse alternative Frères musulmans qui ont montré n'être qu'en minorité dans le mouvement du **peuple** égyptien pour un changement réel

Mao n'avait pas tort lorsqu'il affirmait que le capitalisme (réellement existant, c'est-à-dire impérialiste par nature) n'avait rien à offrir aux **peuples** des trois continents (la périphérie constituée par l'Asie, l'Afrique et l'Amérique latine – cette « minorité » qui rassemble 85% de la population de la planète

Le discours dominant aujourd'hui appelle à la mise en œuvre d'un « droit international » qui autorise en principe l'intervention lorsque les droits fondamentaux d'un **peuple** sont bafoués

Les mouvements partent, comme ceux du siècle précédent, de la reconquête de l'indépendance des **peuples** et des Etas des périphéries du système, reprenant l'initiative dans la transformation du monde

La lutte pour la démocratisation laïque prend alors une dimension décisive dans le moment actuel qui oppose la perspective d'une émancipation des **peuples** à celle de la barbarie généralisée

*PRINTEMPS ARABE  $\xrightarrow{\text{max}}$  IMPÉRIALISME*

Les « **printemps** » des **peuples arabes**, comme ceux que les **peuples** d'Amérique latine connaissent depuis deux décennies, que j'appelle la seconde vague de l'éveil des **peuples** du Sud – la première s'était déployée au 20<sup>ième</sup> siècle jusqu'à la controffensive du capitalisme/**impérialisme** néo libéral – revêt des formes diverses allant des explosions dirigées contre les autocraties qui ont précisément accompagné le déploiement néo libéral à la remise en cause de l'ordre international par les « pays émergents »

Dans la première hypothèse, les avancées du monde **arabe** s'inscriront nécessairement ans le mouvement de dépassement du capitalisme / **impérialisme** à l'échelle mondiale

## 2011 : LE PRINTEMPS ARABE

Le « printemps arabe » s'inscrit dans cette réalité

Il reste que sans l'agression de l'impérialisme ces contradictions auraient probablement pu être surmontées, comme l'exemple japonais le suggère

Cette interconnexion des objectifs n'est d'ailleurs rien d'autre que l'expression des effets de l'intégration de l'Égypte moderne dans le système du capitalisme / impérialisme mondialisé de l'époque

L'essoufflement offert à l'impérialisme, dirigé désormais par les États-Unis, l'occasion de briser le mouvement, en mobilisant à cet effet leur instrument militaire régional : Israël

Les Frères musulmans, depuis leur création en 1927 jusqu'à ce jour, ont toujours été un allié utile pour l'impérialisme et le bloc réactionnaire local

Une première vague de « révolutions », si on les appelle ainsi, avait balayé certaines dictatures en Asie (les Philippines, l'Indonésie) et en Afrique (le Mali), qui avaient été mises en place par l'impérialisme et les blocs réactionnaires locaux

L'impérialisme ne veut ni le progrès social, ni la démocratie pour ces pays

On ne peut que déplorer que la « gauche » européenne, même radicale, ait cessé de comprendre ce qu'est l'impérialisme

Tableau 7.2 Liste des segments contenant les règles :  $PRINTEMPS\ ARABE \xRightarrow{max} PEUPLE$   
et  $PRINTEMPS\ ARABE \xRightarrow{max} IMPÉRIALISME$ .

L'exploration de la liste des segments nous permet de dégager pour chaque règle les traits caractérisant le contenu essentiel de ces segments (voir tableau 7.3).

règle	Sens général des segments correspondants
$\begin{array}{l} \textit{PRINTEMPS ARABE} \\ \xRightarrow{\textit{max}} \textit{PEUPLE} \end{array}$	Le printemps arabe est déclenché par le soulèvement des peuples arabes pour un changement réel de leurs situations
$\begin{array}{l} \textit{PRINTEMPS ARABE} \\ \xRightarrow{\textit{max}} \textit{IMPÉRIALISME} \\ \textit{ARABE} \xRightarrow{\textit{max}} \textit{CAPITALISME} \end{array}$	Les pays du sud de la planète ont été exploités par les forces impérialistes et capitalistes afin de les contrôler et protéger leurs intérêts dans la région.
$\begin{array}{l} \textit{PRINTEMPS ARABE} \\ \xRightarrow{\textit{max}} \textit{MOUVEMENT} \end{array}$	Aperçu historique sur les mouvements égyptiens pour le changement jusqu'au printemps arabe.
$\begin{array}{l} \textit{PRINTEMPS ARABE} \\ \xRightarrow{\textit{max}} \textit{RÉVOLTE} \end{array}$	Le printemps arabe s'agit de révoltes sociales qui ont touché plusieurs pays arabes : Égypte, Lybie, Tunisie...
$\begin{array}{l} \textit{PRINTEMPS ARABE} \\ \xRightarrow{\textit{max}} \textit{INTERNATIONAL} \end{array}$	Stratégie internationale suivie par l'Égypte avant le printemps arabe.

Tableau 7.3 Extraction du sens général des segments.

Nous avons choisi l'item " *PEUPLE* " pour poursuivre notre processus de fouille, en cliquant sur le mot " *PEUPLE* " le graphe est mis à jour pour contenir les règles de plus grande qualité correspondantes à l'itemset " *PEUPLE* " (voir Figure 7.2).

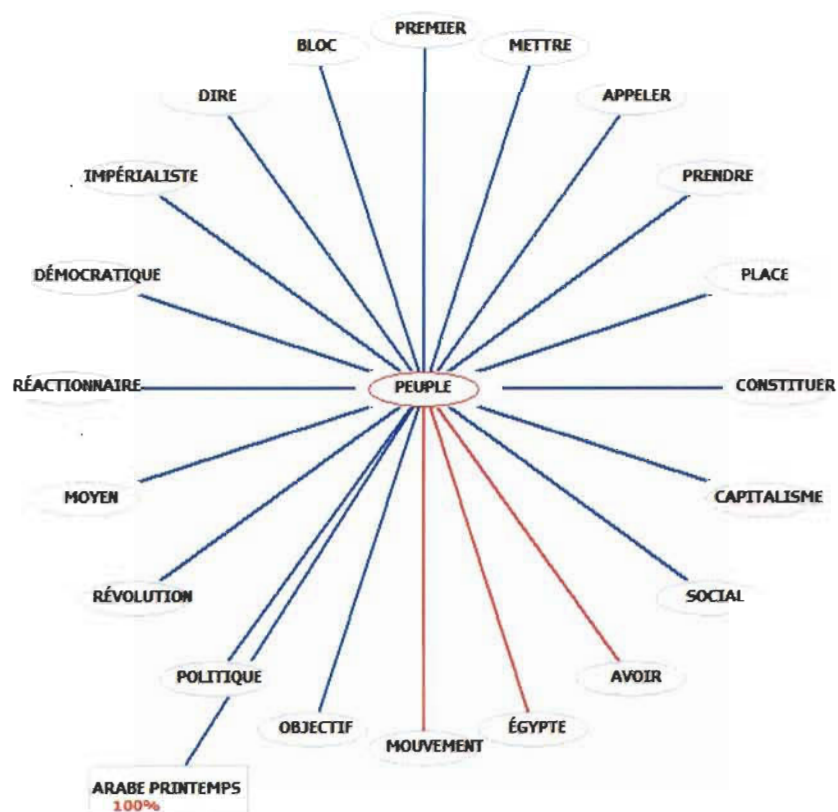


Figure 7.2 Extraction des règles pertinentes avec " *PEUPLE* " en prémisse.

Toujours, avec le même principe, en cherchant les meilleures règles et en consultant les segments qui les contiennent pour comprendre le sens général de ces segments. À la fin, nous avons achevé notre processus de fouille avec une séquence des mots qui semble suffisante pour trouver la majorité des informations pertinentes. Notons que nous avons changé le chemin de navigation plusieurs fois. La figure 7.3 illustre le chemin parcouru durant le processus de fouille.

La liste des segments qui contiennent les éléments de la séquence des mots sélectionnés est présentée à l'annexe A.

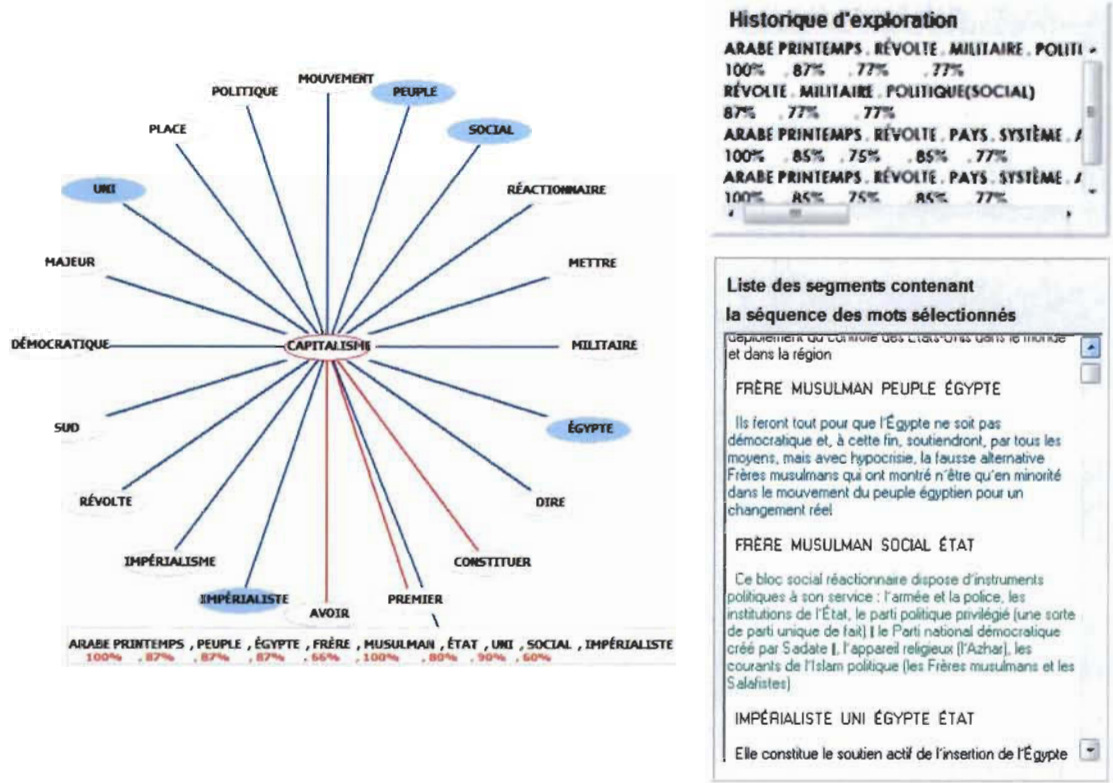


Figure 7.3 Chemins parcourus durant le processus de fouille pour le texte «2011 : Printemps arabe ».

La figure 7.4 contient une explication sur ce qu'on a compris du texte à l'aide du système sans avoir lu le texte original à l'avance.

Dans ce texte, l'auteur a décrit son point de vue sur les avancées du monde arabe, qu'il a analysé à travers le cas de l'Égypte, vu son rôle pivot dans la région.

Dans une approche historique devisée en trois grandes périodes, l'auteur a donné une analyse des forces sociales et politiques existées en Égypte, de leurs identités sociales et politiques, de leurs visions politiques et religion dans lesquelles ils s'inscrivent.

Selon l'auteur, dans la première période (1800-1920) l'Égypte a tenté plusieurs fois d'émerger et de faire sortir de la périphérie du capitalisme mondialisé. Cette émergence d'une Égypte moderne est mise en échec par l'occupation militaire de l'Angleterre (en 1882)

La deuxième période (1919-1967) est caractérisée par la montée des luttes pour l'émancipation et le progrès, et c'est dans cette période où Nasser qui représente un projet anti-impérialiste a pris le contrôle de l'Égypte.

La troisième période (1967-2011) est caractérisée par la domination néolibérale qui est constituée par une alliance entre les hauts militaires, les hommes d'affaires corrompus, et les Frères musulmans. Cette alliance avait été construite par Sadate et Moubarak.

Selon l'auteur, l'intérêt commun entre les Frères musulmans, qui ont une vision économique basée sur le marché libéral et l'États-Unis, capitalisme des monopoles dominants à l'échelle mondiale, va avorter le mouvement démocratique en Égypte. Et les États-Unis va soutenir un « régime islamique » dirigé par les Frères musulmans.

Figure 7.4 Interprétation subjective du contenu de texte original (2011 : printemps arabe).

Sachant que nous avons consacré 66 minutes pour une lecture manuelle du texte original, la lecture du texte à l'aide de notre outil, nous a pris 39 minutes pour extraire l'information illustrée sur la figure 7.4. Le résultat semble satisfaisant au niveau informationnel et comporte une certaine cohérence avec le texte original.

## 7.2. Exemple2: 8 MAI 1945 - MAI 1985, 40ème anniversaire

Le texte que nous avons utilisé dans cet exemple pour évaluer notre logiciel provient de la revue *SOCIALISTE*, N° 21, MAI 1985. Publié le 12 juin 2010 sur le site « *socialgerie.net* » [58]. Ce texte traite un évènement tragique qui a eu lieu dans la région de Sétif (Algérie) en mai 1945 causant la mort de 45000 Algériens contre 104 Français.

Les paramètres du prétraitement de texte ont été fixés comme suit :

- Une segmentation par "paragraphe "
- 2 caractères pour la taille des N-gram.
- Les lettres minuscules sont identiques aux lettres majuscules.
- La suppression des N-gram avec espace.
- Classification du texte avec le classifieur K-means.
- Division des segments en 20 classes.

Nous avons commencé notre recherche de l'information pertinente par la visualisation de 20 meilleures règles d'association maximales concernant la prémisse "*MAI*".

L'algorithme d'extraction de règles d'association maximales a produit le graphe suivant (Figure 7.5) :

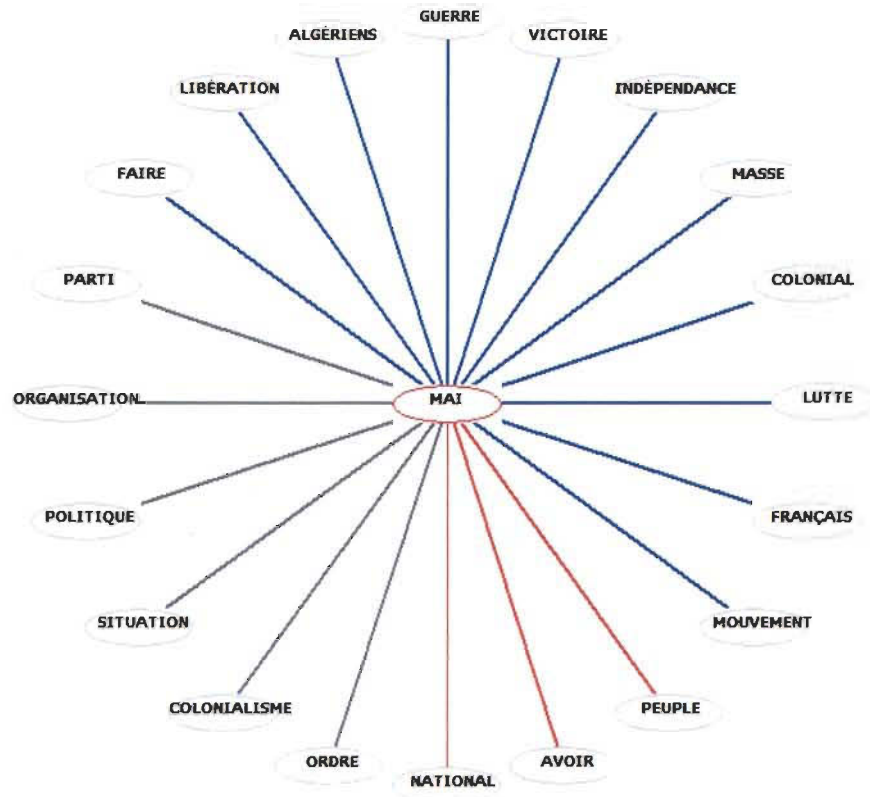


Figure 7.5 Extraction des règles pertinentes avec " *MAI* " en prémisses.

Notons que les arcs de couleur rouge désignent des règles de très bonne qualité, ce sont des règles qui ont un M-confiance entre 80% et 100%. Les arcs de couleur bleu représentent des règles moins bonnes dont le M-confiance vari entre 60% et 80%.

Les arcs de couleur rouge désignent des règles qui possèdent en effet de très bonnes valeurs de M-confiance (de 80% à 100%). L'analyse de la cohérence des résultats obtenus pour l'itemset "*MAI*" par rapport au texte original se résume dans le tableau 7.4 :



Règle, M-confiance	Analyse de la règle
$MAI \xRightarrow{max} NATIONAL$ , 92%	L'itemset "MAI" a une forte relation avec les itemsets : "NATIONAL", "PEUPLE", "FRANCAIS", "MOUVEMENT" et "LUTTE", "COLONIAL", "INDÉPENDANCE", "ALGÉRIENS", "LIBÉRATION". En effet en lisant le texte, on voit très bien que le 8 Mai est une journée <b>nationale</b> du souvenir de victimes <b>algériennes</b> dans le massacre commis en 1945 par la <b>France coloniale</b> contre le <b>peuple algérien</b> .  Ansi que les manifestations de 8 mai 1945 sont des <b>mouvements</b> de revendications <b>nationaux</b> pour <b>lutter</b> pour la reconquête de la <b>liberté</b> et <b>l'indépendance</b> du <b>peuple algérien</b> .
$MAI \xRightarrow{max} PEUPLE$ , 84%	
$MAI \xRightarrow{max} FRANCAIS$ , 76%	
$MAI \xRightarrow{max} MOUVEMENT$ , 76%	
$MAI \xRightarrow{max} LUTTE$ , 76%	
$MAI \xRightarrow{max} COLONIAL$ , 69%	
$MAI \xRightarrow{max} INDÉPENDANCE$ , 61%	
$MAI \xRightarrow{max} ALGÉRIENS$ , 61%	
$MAI \xRightarrow{max} LBÉRATION$ , 61%	

Tableau 7.4 Cohérence de résultat obtenu pour la prémisse " MAI " par apport au texte original

Nous avons exploré les segments correspondants aux meilleures règles, ce qui nous a permis d'extraire le sens général de ces segments.

Le tableau 7.5 illustre la liste des segments extraits par notre logiciel concernant les deux règles ( $\xRightarrow{max} NATIONAL$ , 92%) et ( $MAI \xRightarrow{max} PEUPLE$ , 84%).

$MAI \xrightarrow{max} NATIONAL$

Mais l'hommage que nous rendons à nos milliers de martyrs de **Mai** 45, victimes de la répression barbare et précurseurs de la guerre de libération, est lié à la célébration par l'humanité progressiste du 40ème anniversaire de la victoire sur le fascisme qui a eu une portée considérable dans la vie de tous les peuples du monde. Nous partageons, avec de très larges secteurs de l'opinion publique mondiale, la conviction que la victoire antifasciste de **Mai** 45 a été pour toute l'humanité un moment décisif. Elle a non seulement mis fin aux terribles souffrances dues à la guerre et aux méthodes barbares de l'agression fasciste, mais elle a été aussi un puissant facteur d'émancipation **nationale** et sociale. Elle a été un facteur de démocratisation dans la vie de chaque peuple et de l'instauration de nouveaux rapports entre les nations.

Le 8 **Mai** 1945 une manifestation de la conscience **nationale**

Les évènements de **Mai** 1945 dans notre pays (1er **Mai** à Alger et 8 **Mai** et les jours suivants dans le Constantinois ont été avant tout une manifestation de la conscience **nationale** de notre peuple et de sa volonté d'indépendance et de libération sociale et culturelle.

Les enseignements du 8 **Mai** 1945 ont appris à notre peuple et à l'ensemble de ses organisations **nationales** à s'unir, à s'organiser et à se mobiliser pour forger les conditions de la victoire contre le colonialisme français.

Pour ce qui est du mouvement communiste et ouvrier algérien on peut affirmer que les événements du 8 **Mai** 1945 ont été déterminants pour forger progressivement dans les faits (et pas seulement en principe) sa ligne **nationale** et son caractère révolutionnaire. Ils ont beaucoup contribué à créer les conditions pour que le PCA soit en état de définir et de suivre une ligne juste pendant la guerre de libération **nationale**. La lutte menée par le PCA a pu ainsi être partie intégrante de celle de notre peuple. Le PCA a su intégrer son action propre dans celle du FLN en tant que large rassemblement patriotique. Ses militants armés ont été intégrés dans l'ALN, à la suite d'un accord réalisé entre les directions du FLN et du PCA en juin 1956.

En célébrant ce double anniversaire du 8 **Mai** 45, celui de la victoire mondiale antifasciste et celui de la monstrueuse répression colonialiste contre les patriotes algériens qui célébraient cette victoire en réclamant l'indépendance, nous tirons deux enseignements, étroitement liés au plan **national** et **international** :

Dès l'Été 45, il apparaît que les événements de **mai** ont provoqué des divisions profondes au sein du mouvement **national**, rejetant les forces "modérées" vers le retour d'une politique réformiste vis-à-vis du colonialisme, et les forces radicales vers une clandestinité très difficile pour les conditions de l'époque.

C'est dans ces conditions que les communistes, qui ont mesuré rapidement le caractère erroné et nocif de leurs positions sur les événements du 8 **Mai**, prennent l'initiative de

lancer une campagne pour la libération des détenus. Rapidement les comités d'amnistie se mettent en place, au niveau **national** (avec comme SG Larbi Bouhali) et aux niveaux locaux. Des meetings, assemblées se multiplient, la presse communiste révèle chaque jour les méfaits de la répression de Mai etc.

Ce mouvement pour l'Amnistie aboutit au vote d'une loi, dont le projet est déposé à l'Assemblée **Nationale** Française par le groupe communiste. Cette loi remet en liberté des centaines de détenus de Mai. Ce mouvement apparaît avec le recul comme un véritable maillon intermédiaire pour la reconstitution et la relance de la lutte du mouvement **national**

$MAI \xrightarrow{max} PEUPLE$

Le 8 **Mai** 1945 a une charge émotive très grande pour nous. Il évoque avant tout la volonté de lutte de notre **peuple** et les massacres commis par le colonialisme français : des dizaines de militants d'Algériens froidement assassinés en particulier à Sétif et dans sa région, et à Guelma pour le seul "crime" de revendiquer l'indépendance de leur pays.

Mais l'hommage que nous rendons à nos milliers de martyrs de **Mai** 45, victimes de la répression barbare et précurseurs de la guerre de libération, est lié à la célébration par l'humanité progressiste du 40ème anniversaire de la victoire sur le fascisme qui a eu une portée considérable dans la vie de tous les **peuples** du monde. Nous partageons, avec de très larges secteurs de l'opinion publique mondiale, la conviction que la victoire antifasciste de **Mai** 45 a été pour toute l'humanité un moment décisif. Elle a non seulement mis fin aux terribles souffrances dues à la guerre et aux méthodes barbares de l'agression fasciste, mais elle a été aussi un puissant facteur d'émancipation nationale et sociale. Elle a été un facteur de démocratisation dans la vie de chaque **peuple** et de l'instauration de nouveaux rapports entre les nations.

Les événements de **Mai** 1945 dans notre pays (1er **Mai** à Alger et 8 **Mai** et les jours suivants dans le Constantinois ont été avant tout une manifestation de la conscience nationale de notre **peuple** et de sa volonté d'indépendance et de libération sociale et culturelle.

Le fait essentiel demeure que la signification profonde des manifestations de **Mai** 1945 et leur prolongement dans les mois suivants ont créé de nouvelles conditions et une nouvelle qualité dans la lutte de notre **peuple** et de toutes ses organisations.

Les enseignements du 8 **Mai** 1945 ont appris à notre **peuple** et à l'ensemble de ses organisations nationales à s'unir, à s'organiser et à se mobiliser pour forger les conditions de la victoire contre le colonialisme français.

La haine de notre **peuple** contre le colonialisme français se confirma et se renforça après les massacres du 8 **Mai** 45.

Pour ce qui est du mouvement communiste et ouvrier algérien on peut affirmer que les événements du 8 **Mai** 1945 ont été déterminants pour forger progressivement dans les faits (et pas seulement en principe) sa ligne nationale et son caractère révolutionnaire. Ils ont

beaucoup contribué à créer les conditions pour que le PCA soit en état de définir et de suivre une ligne juste pendant la guerre de libération nationale. La lutte menée par le PCA a pu ainsi être partie intégrante de celle de notre **peuple**. Le PCA a su intégrer son action propre dans celle du FLN en tant que large rassemblement patriotique. Ses militants armés ont été intégrés dans l'ALN, à la suite d'un accord réalisé entre les directions du FLN et du PCA en juin 1956.

Nous rendons hommage à la mémoire de nos 45. 000 compatriotes tombés en **Mai** 1945 sous les balles et le feu des colonialistes, comme nous rendons hommage à celle des millions d'autres morts dans le combat antifasciste, ainsi qu'à tous les **peuples** du monde, en premier lieu les **peuples** de l'URSS. La victoire contre le fascisme, celles remportées par d'autres **peuples** comme le VietNam face au colonialisme français, ont encouragé notre **peuple** pour se dresser et mettre fin au cauchemar colonialiste, au prix de plus d'un million de martyrs. La reconquête de notre indépendance a, à son tour favorisé en Afrique une longue série de victoires anticolonialistes renforçant ainsi le Front Mondial contre l'Impérialisme. Ainsi s'alimentent réciproquement les luttes de chaque **peuple** et celles du large front des trois grandes forces révolutionnaires mondiales.

Le 8 **Mai** 45 se situe dans la trajectoire du long combat qui a opposé notre **peuple** à l'occupant colonial.

Tableau 7.5 Liste des segments contenant les règles :  $MAI \xRightarrow{max} NATIONAL$  et  $MAI \xRightarrow{max} PEUPLE$ .

Le sens général déduit en fouillant ces deux segments est donné dans le tableau suivant :

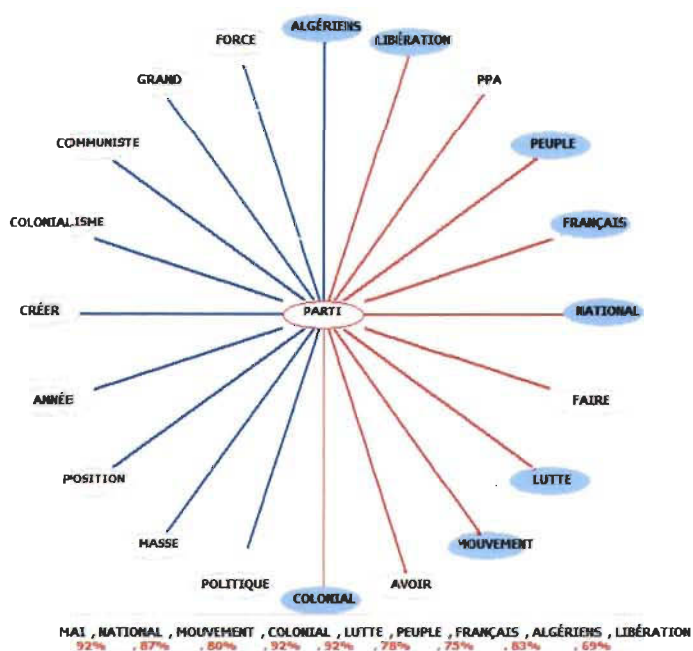
règle	Sens général des segments correspondants
$MAI \xRightarrow{max} NATIONAL$	Après une longue durée de faiblesse et de divisions profondes au sein du mouvement national, les événements de 8 Mai ont donné une leçon à ces derniers, de se réunir et de s'organiser pour forger les conditions de l'indépendance et de la liberté national.

$MAI \xRightarrow{max} PEUPLE$	<p>Les événements de 8 Mai 45 se situent, dans la trajectoire du long combat qui a opposé le peuple algérien au colonialisme français ainsi la conscience nationale pour l'indépendance du peuple se confirma, et se renforça après ces massacres.</p>
--------------------------------	--

Tableau 7.6 Extraction du sens général des segments.

Le scénario de requêtes poursuivi dans notre processus de fouille à la recherche de l'information pertinente est illustré sur la figure 7.6.

La liste des segments qui contiennent les éléments de la séquence des mots sélectionnés (*MAI, NATIONAL, MOUVEMENT, COLONAL, LUTTE, PEUPLE, FRANÇAIS, ALGÉRIENS, LIBÉRATION*) est présentée à l'annexe B.



## Historique d'exploration

MAI , NATIONAL , MOUVEMENT , COLONIAL(LUTTE)  
92% , 87% , 80% , 92%  
MOUVEMENT , COLONIAL(LUTTE)  
80% , 92%  
MAI , NATIONAL , MOUVEMENT , PPA , COMMUNISTE  
92% , 87% , 66% , 72% , 88%  
NATIONAL , MOUVEMENT , PPA , COMMUNISTE  
87% , 66% , 72% , 88%

Liste des segments contenant  
la séquence des mots sélectionnés

LIBÉRATION LUTTE MAI MOUVEMENT  
NATIONAL PEUPLE

Pour ce qui est du mouvement communiste et ouvrier algérien on peut affirmer que les événements du 8 Mai 1945 ont été déterminants pour forger progressivement dans les faits (et pas seulement en principe) sa ligne nationale et son caractère révolutionnaire. Ils ont beaucoup contribué à créer les conditions pour que le PCA soit en état de définir et de suivre une ligne juste pendant la guerre de libération nationale. La lutte menée par le PCA a pu ainsi être partie intégrante de celle de notre peuple. Le PCA a su intégrer son action propre dans celle du FLN en tant que large rassemblement patriotique. Ses militants armés ont été intégrés dans l'ALN, à la suite d'un accord réalisé entre les directions du FLN et du PCA en juin 1956.

ALGÉRIENS COLONIAL FRANÇAIS MOUVEMENT  
NATIONAL

Après le discours de De Gaulle à Constantine, à la même période, un plan de réformes est envisagé pour

Figure 7.6 chemins parcourus durant le processus de fouille pour le texte «8 MAI 1945 - MAI 1985, 40ème anniversaire ».

A la fin de ce processus, nous avons pu faire ressortir une idée sur le contenu général du texte original. Cette dernière a été mise dans un document appelé document explicatif du texte (voir figure 7.7).

À l'occasion du 40ème anniversaire des événements de Sétif en mai 1945, l'auteur a présenté dans une approche historique les causes et les conséquences de ces événements.

La lutte des algériens pour la liberté a commencé dès le premier jour où les Français ont mis leurs pieds dans l'Algérie

Dans la période du 1920 à 1940, une nouvelle phase est apparue : la création des syndicats, partis politiques, associations culturelles et religieuses, pour réaliser la jonction entre lutte politique et lutte armée.

Dans la période du 1939 à 1945 (2ème guerre mondiale), les algériens ont connu une détérioration catastrophique des conditions de vie, vu la dépendance étroite de l'économie algérienne par rapport à la puissance coloniale française.

La défaite rapide des armées françaises face à l'Allemagne a donné espoir aux algériens pour se libérer, par contre cette victoire n'a fait qu'accroître l'oppression coloniale.

La période 1942, et jusqu'à 1945, a connue un puissant essor du mouvement de revendication nationale.

Des manifestations pacifiques ont eu lieu le 8 mai à travers la plupart des grandes villes algériennes pour fêter la victoire mondiale sur le nazisme et revendiquer l'indépendance. Mais à Sétif un policier tire sur un jeune tenant un drapeau de l'Algérie et le tue, ce qui déclenche une émeute meurtrière des manifestants, avant que l'armée n'intervienne sauvagement.

Le bilan a été de 45 000 morts algériens et une centaine de morts européens.

Pour l'auteur, à travers ces massacres le peuple a pris conscience que la lutte armée reste la seule solution.

Figure 7.7 Interprétation subjective du texte (8 MAI : 8 MAI 1985, 40ème anniversaire).

Sachant que nous avons investi environ 73 minutes dans la lecture manuelle du texte original, à l'aide du logiciel, nous avons consacré 40 minutes afin de comprendre le

contenu général du texte et de dégager le document explicatif illustré sur la figure 7.5. Le résultat obtenu donne une grande satisfaction au niveau informationnel et cohérence.

### **7.3. Conclusion**

Nous avons consacré ce chapitre à la présentation des résultats obtenus par le traitement de deux documents textuels.

Notre évaluation nous a permis de valider la pertinence de l'information extraite, En effet, le document explicatif du contenu de texte original construit par notre système montre une cohérence et un niveau informationnel satisfaisantes avec le texte original.



## Chapitre 8 - Conclusion

Dans le cadre de ce mémoire, nous nous sommes intéressés à l'exploration de grandes quantités des règles qui peuvent être générées par les algorithmes d'extraction de connaissances. Ce qui ne facilite pas la tâche de fouille de connaissances de l'utilisateur. La technique de visualisation graphique s'avère très bénéfique pour l'assister.

Nous avons développé un module opérationnel pour l'extraction des règles d'association maximales centré sur l'utilisateur via une représentation graphique et interactive des règles générées par l'outil **GRAMEXCO**, permettant d'explorer de grands volumes de règles et de faire sortir des informations pertinentes.

Ce module est conçu sur une méthodologie de telle sorte que l'utilisateur explore la représentation graphique des règles par un algorithme d'extraction d'associations maximales qui permet de n'extraire qu'un sous-ensemble de règles que l'utilisateur souhaite visualiser et faire interactivement évoluer ce graphe de règles selon son point de vue, afin de cibler les règles intéressantes qui permettent d'extraire l'information pertinente.

Nous avons apporté une certaine optimisation à l'algorithme d'extraction des règles M-association de la version précédente de l'outil GRAMEXCO en termes de temps de calcul et de ressources en mémoire utilisées afin d'améliorer la performance du processus de fouille de données.

L'outil a été testé sur un ensemble de corpus réels de taille moyenne (environ 20 pages), issus des sites web, les résultats obtenus donnent une grande satisfaction, soit pour la pertinence de l'information extraite en utilisant notre approche de visualisation interactive des règles, soit pour le temps de calcul des règles M-associations qui est devenu plus raisonnable.

De nombreuses perspectives s'offrent à la suite de notre travail à savoir :

- l'application de ce modèle pour la recherche documentaire sur le Web. En effet, la séquence des mots sélectionnés à la fin du processus de fouille pour un document *D*, peut être utilisée dans la requête de recherche sur le Web pour trouver des documents similaires à *D*.
- Utilisation d'autres mesures de qualité des règles largement répandues telles que : coefficient de corrélation, Lift, complétude,... [59].
- Ajouter un autre type de représentation telle que la visualisation tridimensionnelle, pour cela nous nous référons au modèle réalisé par BLANCHARD [30].

## Bibliographie (ou Références)

- [1] Haytham Elghazel (2007). *Classification et Prévission des Données Hétérogènes : Application aux Trajectoires et Séjours Hospitaliers*, préparée au sein du laboratoire LIESP, Université Claude Bernard Lyon 1.
- [2] J.Han et M. Kamber (2001), *Data mining: concepts and techniques*. Morgan Kaufmann Publisher.
- [3] Jean-Pierre Nakache et Josiane Confais (2004). *Approche pragmatique de la classification : Arbres hiérarchiques - Partitionnements*. Edition TECHNIP.
- [4] Anh Tuan, IFI Hanoi (2004). *Réduction de base de données par la classification automatique*, Institut de la Francophonie pour l'Informatique Hanoi-Vietnam.
- [5] R. Agrawal, T. Imielinski, et A. Swami (1993). *Mining Association Rules between sets of Items in Large Databases*. In Proceedings of the ACM SIGMOD International Conference on Management of Data, pages 207–216, Washington D.C., May 1993.
- [6] Rakesh Agrawal et Ramakrishnan Srikant (1994). *Fast algorithms for mining association rules in large databases*. In *International Conference on Very Large Data Bases (VLDB '94)*, pages 487–499, San Francisco, Ca., USA, Septembre 1994. Morgan Kaufmann Publishers, Inc.
- [7] A. Jain, M. Murty et P. Flynn (2009). Data clustering : a review. *ACM Computing Surveys*.
- [8] Claudia Victoria, Isaza Narvaez (2007). *Diagnostic par techniques d'apprentissage floues : Conception d'une méthode de validation et d'optimisation des partitions*. CNRS Toulouse-France.
- [9] Ali Meghaoui (2008). *Exploitation des arbres fréquents de dépendance pour la représentation et la classification automatique de textes*. In NATO Advanced study Institute on Digital Image Processing and Analysis, Bonas-France.
- [10] E. Diday, G. Govaert, Y. Lechevallier, et J. Sidi (1980). *Clustering in pattern recognition*. *IEEE Trans. Electron Devices*, vol. ED-11, pp. 34–39.
- [11] Nicolas. Pasquier (2000). *Data Mining : algorithmes d'extraction et de réduction des règles d'association dans les bases de données*. Université de Clermont-Ferrand II, France, 2000.

- [12] U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, et R. Uthurusamy (1996). *Advances in Knowledge Discovery and Data Mining*. AAAI Press.
- [13] Hassane Hilali (2009). *Application de la classification textuelle pour l'extraction des règles d'association maximales*. Université de Québec à Trois-Rivières, avril 2009.
- [14] M.-S. Chen, J. Han, et P. S. Yu(1996). *Data mining: An overview from a database perspective*. IEEE Tansansaction on Knowledge and Data Engineering.
- [15] Berrani, Sid-Ahmed (2004). *Recherche approximative de plus proches voisins avec contrôle probabiliste de la précision ; application à la recherche d'images par le contenu*. Université de Rennes I.
- [16] Joachims, T. (1998). *Text categorization with support vector machines: learning with many relevant features*. In C. Nédellec et C. Rouveirol (Eds.), Proceedings of ECML-98, 10<sup>th</sup> European Conference on Machine Learning, Number 1398, Chemnitz, DE, pp. 137–142.
- [17] Sami Laroum, Nicolas Béchet, Hatem Hamza et Mathieu Roche (2009). *Classification automatique de documents bruités à faible contenu textuel*. INRA Nantes – France.
- [18] Mariam TANANA(2009). *Évaluation formative du savoir-faire des apprenants à l'aide d'algorithmes de classification*. MINSa de Rouen-France.
- [19] Réhel, Simon(2005). *Catégorisation automatique de textes et cooccurrence de mots provenant de documents non étiquetés*. Université Laval-Québec.
- [20] Jérôme Azé (2003). *Extraction de connaissances à partir de données numériques et textuelles*. Université Paris Sud.
- [21] Cheikh Talibouya Diop, Moussa Lo et Fatou Kamara Sangaré(2007). *Intégration de règles d'association pour améliorer la recherche d'informations XML*. 4eme conférence francophone en Recherche d'Informations et Applications. École Nationale supérieure des Mines de Saint-Étienne.
- [22] Clément Fauré (2007). *Découvertes de motifs pertinents par l'implémentation d'un réseau bayésien : application à l'industrie aéronautique*. L'Institut National des Sciences appliquées de Lyon.
- [23] Benoît Vaillant (2006). *Mesurer la qualité des règles d'association : études formelles et expérimentales*. Nationale Supérieure des Télécommunications de Bretagne.
- [24] Ronen Feldman, Yonatan Aumann, Amihod Amir, Amir Zilberstein et Willi Kloesgen(1997). *Maximal Association Rules: a New Tool for Mining for Keyword co-occurrences in Document Collections*. Department of Mathematics and Computer Science Department, Bar-Ilan University, Ramat-Gan, ISRAEL”, “German National Research Center for Information Technology, Germany”.

- [25] A. Amir, Y. Aumann, R. Feldman, M. Fresko (2004). *Maximal Association Rules: A Tool for Mining Associations in Text*. Journal of Intelligent Information Systems. Department of Computer Science, Bar Ilan University, Israel
- [26] Tutut Herawan et Mustafa Mat Deris (2011). *A soft set approach for association rules mining*. Knowledge-Based Systems 24, CSSE, University Malaysia Pahang ITM, University Tun Hussein Onn Malaysia.
- [27] P. C. Wong, P. Whitney & J. Thomas. Visualizing association rules for text. in Proceedings of the 1999 IEEE symposium on information visualization, IEEE Computer Society, 1999, p. 120–123.
- [28] L. Szathmary, S. Maumus, P. Petronin, Y. Toussaint, et A. Napoli (2006). *Vers l'extraction de motifs rares*. In G. Ritschard and C. Djeraba, editors, Extraction et gestion des connaissances, Lille, pages 499-510. Editions Toulouse.
- [29] Julien BLANCHARD, Fabrice Guillet, Henri Briand(2003). *Une visualisation orientée qualité pour la fouille anthropocentrée de règles d'association*. IRIN École polytechnique de l'université de Nantes.
- [30] Julien BLANCHARD (2005). *Un système de visualisation pour l'extraction, l'évaluation, et l'exploration interactives des règles d'association*. École Polytechnique de l'Université de Nantes.
- [31] [www.dbminer.com](http://www.dbminer.com)
- [32] [www.purpleinsight.com](http://www.purpleinsight.com)
- [33] J. Han (1998). *Towards on-line analytical mining in large databases*. SIGMOD Special Interest Group on Management Of Data. Record 27, no 1, p. 97–107.
- [34] C. Brunk, J. Quelly et R. Kohavi (1997). *Mineset : An integrated system for data mining*. in Proceedings of the third ACM SIGKDD international conference on knowledge discovery and data mining (D. Heckerman, H. Mannila, D. Pregibon & R. Uthurusamy, eds.), AAAI Press, 1997, p. 135–138.
- [35] [www.ibm.com/software/data/iminer/visualization/index.html](http://www.ibm.com/software/data/iminer/visualization/index.html)
- [36] M. Klemettinen, H. Mannila, P. Ronkainen, H. Toivonen et A. I. Verkamo (1994). *Finding interesting rules from large sets of discovered association rules*. in Proceedings of the third international conference on information and knowledge management. Gaithersburg, Maryland, 1994.
- [37] C. P. Rainsford et J. F. Roddick (2000). *Visualisation of temporal interval association rules*, in Proceedings of the second international conference on intelligent data engineering and automated learning, Springer-Verlag, p. 91–96.
- [38] H. Hofmann et A. Wilhelm(2001). *Visual comparison of association rules*, Computational Statistics 16, no. 3, p. 399–415.

- [39] Sharma Chakravarthy and Hongen Zhang (2003). *Visualization of association rules over relational dbmss*. In SAC, pages 922–926.
- [40] Dario Bruzzese and Paolo Buono(2004). *Combining visual techniques for association rules exploration*. In AVI, pages 381–384.
- [41] Marie NDIAYE(2010). *Exploration de grands ensembles de motifs*. École doctorale SST Laboratoire d'Informatique, Équipe BdTIn, Université François Rabelais tours-France.
- [42] Dario Bruzzese and Paolo Buono (2004). *Combining visual techniques for association rules exploration*. In AVI, pages 381–384.
- [43] Jean-François Boulicaut, Patrick Marcel, and Christophe Rigotti (1999). *Query driven knowledge discovery in multidimensional data*. In DOLAP '99 : Proceedings of the 2nd ACM international workshop on Data warehousing and OLAP, pages 87–93, New York, NY, USA. ACM Press.
- [44] Julien Blanchard, Fabrice Guillet, and Henri Briand. Exploratory visualization for association rule rummaging. In KDD-03 Workshop on Multimedia Data Mining (MDM-03, pages 107–114, 2003.
- [45] Ismaïl Biskri, Hassane Hillali, Louis Rompré(2010). *Extraction de relations d'association maximales dans les textes*. Université de Québec à Trois-Rivières-canada.
- [46] P. Fule & J. F. Roddick (2004). *Experiences in building a tool for navigating association rule result sets*. Australien Computer Society, Inc. p. 103–108.
- [47] I. Kopanakis & B. Theodoulidis (2001). *Visual data mining and modeling techniques in Proceedings of the KDD*. workshop on visual data mining.
- [48] Nicolas Pasquier. *Extraction de Bases pour les Règles d'Association à partir des Itemsets Fermés Fréquents*. Laboratoire d'Informatique (LIMOS) - Université Clermont-Ferrand II.
- [49] Ismail Biskri & Sylvain Delisle (2001). *Les n-grams de caractères pour l'aide à l'extraction de connaissances dans des bases de données textuelles multilingues*. Université de Québec à Trois-Rivières
- [50] Ismaïl Biskri, Hassane Hillali, Steve Descoteaux et Louis Rompré. *Aide à l'interprétation des classes de similarité* (chapitre 12).
- [51] Laettia Jourdan (2003). *Méta heuristiques pour l'extraction de connaissances : application a la génomique*. Université des sciences et technologies de Lille U.F.R. D'I.E.E.A.
- [52] Belacel Nabil (2000). *Méthodes de Classification Multicritère : Méthodologie et Applications à l'aide au diagnostic médical*. INSTITUT DE STATISTIQUE ET DE RECHERCHE OPERATIONNELLE, Université Libre de Bruxelles.

- [53] Sabrina Tollari (2000). *Indexation et recherche d'images par fusion d'informations textuelles et visuelles*. Université du Sud Toulon-Var-France.
- [54] Gilles Lebrun (2006). *Sélection de modèles pour la classification supervisée avec des SVM(Séparateurs à VasteMarge). Application en traitement et analyse d'images*. UNIVERSITE de CAEN/BASSE-NORMANDIE.
- [55] Mounzer BOUBOU (2006). *Contribution aux méthodes de classification non supervisée via des approches prétopologiques et d'agrégation d'opinions*. Université Claude Bernard - Lyon I
- [56] G. CELEUX, Y. Lechevalier (1980). *Méthodes de discrimination non paramétrique asymptotiquement efficaces au sens de bayes*. Rapport de recherche n°52. INRIA
- [57] <http://www.europe-solidaire.org/spip.php?article21675>
- [58] <http://www.socialgerie.net/>
- [59] Benoît Vaillant (2006). *Mesurer la qualité des règles d'association : études formelles et expérimentales*. École Nationale Supérieure des Télécommunications de Bretagne

## Annexe A

Liste des segments du texte « 2011 : LE PRINTEMPS ARABE » contenant la séquence des mots sélectionnés : *PRINTEMPS ARABE, PEUPLE, ÉGYPTE, RÈRE, MUSULMAN, ÉTAT, UNI, SOCIAL, IMPÉRIALISTE*

### FRÈRE MUSULMAN SOCIAL UNI ÉTAT

Les Frères musulmans ne sont donc « modérés » que dans le double sens où ils ont toujours refusé de formuler un programme économique et social quelconque et que, de fait, il ne remettent pas en cause les politiques néo-libérales réactionnaires, et qu'ils acceptent de facto la soumission aux exigences du déploiement du contrôle des États-Unis dans le monde et dans la région

### FRÈRE MUSULMAN PEUPLE ÉGYPTE

Ils feront tout pour que l'Égypte ne soit pas démocratique et, à cette fin, soutiendront, par tous les moyens, mais avec hypocrisie, la fausse alternative Frères musulmans qui ont montré n'être qu'en minorité dans le mouvement du peuple égyptien pour un changement réel

### FRÈRE MUSULMAN SOCIAL ÉTAT

Ce bloc social réactionnaire dispose d'instruments politiques à son service : l'armée et la police, les institutions de l'État, le parti politique privilégié (une sorte de parti unique de fait) – le Parti national démocratique créé par Sadate –, l'appareil religieux (l'Azhar), les courants de l'Islam politique (les Frères musulmans et les Salafistes)

### IMPÉRIALISTE UNI ÉGYPTE ÉTAT

Elle constitue le soutien actif de l'insertion de l'Égypte dans la mondialisation impérialiste contemporaine, l'allié inconditionnel des États-Unis

### MUSULMAN SOCIAL ÉGYPTE ÉTAT

Ils sont également le soutien majeur de l'Islam conservateur dans les campagnes et, par leurs rapports étroits (souvent de parenté) avec les représentants des appareils d'État et de la religion (l'Azhar est en Égypte l'équivalent d'une Église musulmane organisée) dominant la vie sociale rurale

### FRÈRE MUSULMAN ÉTAT

Cette bourgeoisie compte dans ses rangs de nombreux généraux de l'armée et de la police, de « civils » associés à l'État et au parti dominant (« National démocratique ») créé par Sadate et Moubarak, de religieux (la totalité des dirigeants des Frères musulmans et des cheikhs majeurs de l'Azhar, sont tous des « milliardaires »)



## FRÈRE MUSULMAN ÉTAT

Des opérations difficiles à imaginer sans le soutien tacite (et parfois davantage de complicité) de l'appareil d'État, en particulier de la Justice, largement confiée aux Frères musulmans

## ARABE PEUPLE PRINTEMPS

Les « printemps » des peuples arabes, comme ceux que les peuples d'Amérique latine connaissent depuis deux décennies, que j'appelle la seconde vague de l'éveil des peuples du Sud – la première s'était déployée au 20<sup>ème</sup> siècle jusqu'à la controffensive du capitalisme/impérialisme néo libéral – revêt des formes diverses allant des explosions dirigées contre les autocraties qui ont précisément accompagné le déploiement néo libéral à la remise en cause de l'ordre international par les « pays émergents »

## FRÈRE MUSULMAN PEUPLE

Les Frères musulmans dont les dirigeants avaient appelé à boycotter les manifestations pendant les quatre premiers jours (persuadés que celles-ci seraient mises en déroute par la répression) n'ont accepté le mouvement que tardivement, lorsque l'appel, entendu par l'ensemble du peuple égyptien, a produit des mobilisations gigantesques de 15 millions de manifestants

## IMPÉRIALISTE SOCIAL ÉGYPTE

Tout comme dans la période de flux des luttes du passé, le mouvement démocratique anti-impérialiste et social se heurte en Égypte à un bloc réactionnaire puissant

## IMPÉRIALISTE UNI ÉTAT

Il y a là une duplicité permanente dans le langage des dirigeants de la triade impérialiste (États-Unis, Europe occidentale, Japon)

## FRÈRE UNI ÉTAT

Ils sont donc des alliés utiles pour Washington (y-a-t-il un meilleur allié des États-Unis que l'Arabie Saoudite, patron des Frères

## MUSULMAN UNI ÉTAT

L'Islam politique est tout également l'allié stratégique des États-Unis et de leurs partenaires subalternes de l'OTAN à travers le monde musulman

## ARABE UNI ÉTAT

Les États-Unis et l'Europe veulent répéter dans le monde arabe ce qui s'est passé au Mali, aux Philippines et en Indonésie : tout changer pour que rien ne change

## FRÈRE MUSULMAN SOCIAL

Ce n'est pas un hasard si, pour en réduire le danger, l'ambassade britannique et le Palais royal ont alors soutenu activement la création des Frères musulmans (1927) qui s'inspiraient de la pensée « islamiste » dans sa version « salafiste » (passéiste) wahabite formulée par Rachid Reda, c'est à dire la version la plus réactionnaire (antidémocratique et anti progrès social) du nouvel « Islam politique »

## FRÈRE MUSULMAN ÉGYPTE

En Égypte, les Frères musulmans sont désormais épaulés par le courant salafiste (« traditionaliste »), tout également largement financé par les pays du Golfe

## FRÈRE MUSULMAN ÉGYPTE

L'introduction de l'islam wahabite en Égypte a été amorcée par Rachid Reda dans les années 1920 et repris par les Frères musulmans dès 1927

#### SOCIAL UNI ÉTAT

Les jeunes et la gauche radicale poursuivent trois objectifs communs : la restauration de la démocratie (la fin du régime militaire et policier), la mise en œuvre d'une nouvelle politique économique et sociale favorable aux classes populaires (la rupture avec la soumission aux exigences du libéralisme mondialisé), et celle d'une politique internationale indépendante (la rupture avec la soumission aux exigences de l'hégémonie des États-Unis et du déploiement de son contrôle militaire sur la planète)

#### UNI ÉGYPTÉ ÉTAT

Les États-Unis, (et derrière eux l'Europe) craignent plus que tout une Égypte réellement démocratique qui, certainement, remettrait en cause son alignement sur le libéralisme économique et la stratégie agressive des États-Unis et de l'OTAN

#### UNI ÉGYPTÉ

L'Égypte, soumise aux exigences du libéralisme mondialisé et aux stratégies des États-Unis, a cessé d'exister comme acteur actif régional et international

#### ARABE ÉGYPTÉ

Bien avant le Japon et la Chine, dès le début du XIXe siècle Mohammed Ali avait conçu et mis en œuvre un projet de rénovation de l'Égypte et de ses voisins immédiats du Mashreq arabe

#### UNI ÉTAT

L'essoufflement offert à l'impérialisme, dirigé désormais par les États-Unis, l'occasion de briser le mouvement, en mobilisant à cet effet leur instrument militaire régional : Israël

#### UNI ÉTAT

La duplicité cynique du discours de l'establishment des États-Unis (et sur ce plan Obama n'est pas différent de Bush) sert parfaitement ses objectifs

#### UNI ÉTAT

L'aide militaire octroyée par les États-Unis à l'armée égyptienne (1,5 milliard de dollars annuels) n'a jamais été destinée à renforcer la capacité de défense du pays mais au contraire à en annihiler le danger par la corruption systématique, non pas connue et tolérée, mais soutenue positivement, avec cynisme

#### UNI ÉTAT

Mais les États-Unis ne peuvent avouer que leur stratégie vise à mettre en place des régimes « islamiques » dans la région

#### UNI ÉTAT

Par ce moyen, ils légitiment leur « guerre permanente au terrorisme », qui poursuit en réalité d'autres objectifs : le contrôle militaire de la planète destiné à réserver aux États-Unis-Europe-Japon l'accès exclusif aux ressources

#### UNI ÉTAT

Il est plus que jamais nécessaire de faire apparaître clairement cette véritable duplicité de la stratégie des États-Unis, dont les opinions publiques – manipulées avec efficacité – sont dupes

## UNI ÉTAT

Mais il n'a pris toute sa vigueur qu'après la seconde guerre mondiale lorsque la rente pétrolière des pays du Golfe, soutenus par les États-Unis en conflit avec la vague de libération nationale populaire des années 1960, a permis d'en démultiplier les moyens financiers

## UNI ÉTAT

Aujourd'hui, les États-Unis ont besoin de transférer l'Africom (le commandement militaire des États-Unis pour l'Afrique, une pièce importante du dispositif du contrôle militaire de la planète, toujours localisé à Stuttgart

## UNI ÉTAT

Mais ici les États-Unis et l'Europe étaient parvenus à faire avorter la dynamique de ces mouvements populaires, parfois gigantesques par les mobilisations qu'ils ont suscitées

## UNI ÉTAT

Car toute intervention militaire des États-Unis et de l'OTAN dans les affaires des pays du Sud, sous quelque prétexte que ce soit fût-il d'apparence sympathique - comme l'intervention « humanitaire » - doit être proscrite

## UNI ÉTAT

Elle se résume à l'ambassadeur des États-Unis, suivi automatiquement par ceux de l'Europe

## IMPÉRIALISTE SOCIAL

Le nassérisme a alors réalisé ce qu'il pouvait donner : une posture internationale résolument anti-impérialiste (associée aux mouvements panarabe et panafricain), des réformes sociales progressistes (mais non « socialistes »)

## IMPÉRIALISTE SOCIAL

La révolution démocratique à laquelle ils appellent est une révolution démocratique anti-impérialiste et sociale

## PEUPLE SOCIAL

La direction est constituée exclusivement d'hommes immensément riches (grâce, entre autre, au soutien financier de l'Arabie Saoudite, c'est-à-dire de Washington), l'encadrement par des hommes issus des fractions obscurantistes des classes moyennes, la base par des gens du peuple recrutés par les services sociaux de charité offerts par la confrérie (et financés toujours par l'Arabie Saoudite), tandis que la force de frappe est constituée par les milices (les baltaguis) recrutés dans le lumpen

## IMPÉRIALISTE UNI

Le projet de contrôle militaire de la planète par les forces armées des États Unis et de leurs alliés subalternes de l'Otan, toujours en cours, le déclin de la démocratie dans les pays du centre impérialiste, le refus passéiste de la démocratie dans les pays du Sud en révolte (qui prend la forme d'illusions para religieuses « fondamentalistes » que les Islam, Hindouisme et Bouddhisme politiques proposent) opèrent ensemble dans cette perspective abominable

## IMPÉRIALISTE ÉGYPTE

L'Égypte émergente battue a été alors soumise pour près de quarante ans (1880-1920) au statut de périphérie dominée, dont les structures ont été refaçonnées pour servir le modèle de l'accumulation capitaliste / impérialiste de l'époque

#### PEUPLE ÉGYPTÉ

L'Égypte, c'est à dire son peuple, ses élites, la nation qu'elle représente, n'a jamais accepté ce statut

#### PEUPLE ÉGYPTÉ

Cette bourgeoisie est compradore (dans la langue politique courante en Égypte le peuple les qualifie de « parasites corrompus »)

#### ARABE PRINTEMPS

#### 2011 : LE PRINTEMPS ARABE

#### ARABE PRINTEMPS

Le « printemps arabe » s'inscrit dans cette réalité

#### IMPÉRIALISTE PEUPLE

Mao n'avait pas tort lorsqu'il affirmait que le capitalisme (réellement existant, c'est-à-dire impérialiste par nature) n'avait rien à offrir aux peuples des trois continents (la périphérie constituée par l'Asie, l'Afrique et l'Amérique latine – cette « minorité » qui rassemble 85% de la population de la planète

#### FRÈRE MUSULMAN

Là encore, les forces de la réaction égyptienne soutenues par Londres ont réagi avec violence et mobilisé à cet effet les Frères musulmans qui ont soutenu une seconde dictature de Sedki Pacha, sans parvenir à faire taire le mouvement

#### FRÈRE MUSULMAN

Le Wafd revenu au gouvernement, sa dénonciation du Traité de 1936, l'amorce de la guérilla dans la zone du Canal encore occupée, n'ont été mis en déroute que par l'incendie du Caire (1951), une opération dans laquelle les Frères musulmans ont trempé

#### FRÈRE MUSULMAN

Sadate qui succède, accentue la portée de la dérive à droite et intègre les Frères musulmans dans son nouveau système autocratique

#### FRÈRE MUSULMAN

), partis dans la défense de la « démocratie » - à l'américaine- développent souvent, en Égypte, des arguments destinés à légitimer les Frères Musulmans

#### FRÈRE MUSULMAN

La résistance opiniâtre des petits paysans aux expropriations rendues possibles par l'annulation de la réforme agraire (les Frères musulmans ont voté au parlement pour ces lois scélérates, sous prétexte que la propriété privée serait « sacrée » dans l'Islam et que la réforme agraire était inspirée par le diable communiste

#### FRÈRE MUSULMAN

Ces « pauvres » peuvent donner l'impression (par les barbes, les voiles, les accoutrements vestimentaires) que le pays profond est « islamique », voire mobilisé par les Frères musulmans

## FRÈRE MUSULMAN

Les Frères Musulmans et les partis de droite ont refusé de participer à ce Conseil, réaffirmant ainsi ce qu'on sait : leur opposition à la poursuite du mouvement

## FRÈRE MUSULMAN

Parmi ces mesures une loi scélérate antigrevé (sous prétexte de remettre en route l'économie du pays), une loi imposant des restrictions sévères à la constitution des partis politiques qui vise à ne permettre la possibilité d'entrer dans le jeu électoral qu'aux courants de l'Islam politique (les Frères musulmans en particulier) déjà bien organisés grâce au soutien systématique du régime ancien

## FRÈRE MUSULMAN

Les Frères musulmans constituent la seule force politique dont le régime avait non seulement toléré l'existence, mais dont il avait soutenu activement l'épanouissement

## FRÈRE MUSULMAN

Les Frères musulmans n'ont jamais été et ne peuvent pas être « modérés », encore moins « démocratiques »

## FRÈRE MUSULMAN

Les Frères musulmans sont acquis à un système économique basé sur le marché et totalement dépendant de l'extérieur

## FRÈRE MUSULMAN

Les Frères musulmans, depuis leur création en 1927 jusqu'à ce jour, ont toujours été un allié utile pour l'impérialisme et le bloc réactionnaire local

## FRÈRE MUSULMAN

Cette étrange division du travail permet aux Frères musulmans de paraître modérés ; ce que Washington feint de croire

## FRÈRE MUSULMAN

La mobilisation des Frères Musulmans, qui avaient été à l'origine il y a quelques années des insurrections de Hama et de Homs, n'est peut-être pas étrangère au complot de Washington, qui s'emploie à mettre un terme à l'alliance Syrie/Iran, essentielle au soutien de Hezbollah au Liban et de Hamas à Gaza

## SOCIAL ÉGYPTÉ

L'Égypte de Nasser avait mis en place un système économique et social critiquable mais cohérent

## SOCIAL ÉTAT

La vague des mouvements d'émancipation qui a balayé l'Amérique du Sud a par contre permis des avancées réelles dans les trois directions que représentent la démocratisation de l'État et de la société, l'adoption de postures anti-impérialistes conséquentes, l'engagement sur la voie de réformes sociales progressistes

## PEUPLE

Ce mouvement gigantesque du peuple égyptien associe trois composantes actives : les jeunes « re-politisés » par leur propre volonté et dans des formes « modernes » qu'ils ont inventées, les forces de la gauche radicale, celles rassemblées par les classes moyennes démocrates

PEUPLE

L'appel à la manifestation formulé par les trois composantes actives du mouvement a été rapidement entendu par l'ensemble du peuple égyptien

PEUPLE

La généralisation du mouvement à l'ensemble du peuple égyptien constitue par elle-même un défi positif

PEUPLE

Car ce peuple est, comme tous les autres, loin de constituer un « bloc homogène »

PEUPLE

Le discours dominant aujourd'hui appelle à la mise en œuvre d'un « droit international » qui autorise en principe l'intervention lorsque les droits fondamentaux d'un peuple sont bafoués

PEUPLE

Les mouvements partent, comme ceux du siècle précédent, de la reconquête de l'indépendance des peuples et des Etats des périphéries du système, reprenant l'initiative dans la transformation du monde

PEUPLE

La lutte pour la démocratisation laïque prend alors une dimension décisive dans le moment actuel qui oppose la perspective d'une émancipation des peuples à celle de la barbarie généralisée

ARABE

Dans la première hypothèse, les avancées du monde arabe s'inscriront nécessairement ans le mouvement de dépassement du capitalisme / impérialisme à l'échelle mondiale

ARABE

L'échec maintiendrait le monde arabe dans son statut actuel de périphérie dominée, lui interdisant de s'ériger au rang d'acteur actif dans le façonnement du monde

ARABE

Il est toujours dangereux de généraliser en parlant du « monde arabe », en ignorant par là même la diversité des conditions objectives qui caractérisent chacun des pays de ce monde

ARABE

L'épuisement de ce système (la substitution d'immigrés asiatiques à ceux en provenance des pays arabes) a entraîné la renaissance des résistances

ARABE

Il est intéressant de noter à ce propos que nos blogueurs, qui s'expriment en anglais plutôt qu'en arabe (

ARABE

Leur chef - le mourchid (traduction arabe de « guide » - Führer) est autoproclamé et l'organisation repose sur le principe de la discipline et de l'exécution des ordres des chefs, sans discussions d'aucune sorte

ARABE

Je n'ai pas l'intention, dans cet article, d'examiner avec autant de précision chacun des mouvements en cours dans le monde arabe (Tunisie, Libye, Syrie, Yémen et autres)

#### ARABE

La « révolte arabe » ne constitue pas l'exemple unique, même si elle en est l'expression la plus récente, de la manifestation de l'instabilité inhérente à la « zone des tempêtes »

#### MUSULMAN

Des élections « normales » permettaient alors non seulement à des Coptes d'être élus par des majorités musulmanes, mais encore davantage à ces mêmes Coptes d'exercer de très hautes fonctions dans l'Etat, sans que cela ne pose le moindre problème

#### MUSULMAN

Il est intéressant de constater que dans les pays musulmans (Mali, Indonésie), l'Islam politique a été mobilisé à cet effet

#### FRÈRE

La course est donc engagée : qui des Frères et de leurs associés islamistes (les Salafistes) ou de l'alliance démocratique parviendra à formuler des alliances efficaces avec les masses désorientées, voire à les « encadrer » (terme que je récuse)

#### UNI

Dans la région, les alliés majeurs des Etats-Unis – l'Arabie saoudite et Israël – occupent le devant de la scène

#### UNI

Des avancées non négligeables dans la construction du front uni des forces démocratiques et des travailleurs sont en cours en Egypte

#### SOCIAL

Sans doute, les contradictions sociales propres à ce projet comme les conceptions politiques, idéologiques et culturelles sur la base desquelles il se déployait ont-elles leur part de responsabilité dans cet échec

#### SOCIAL

La régression imposée a frappé, au-delà du système productif du pays, ses structures politiques et sociales, comme elle s'est employée à renforcer systématiquement des conceptions idéologiques et culturelles passées et réactionnaires utiles pour le maintien du pays dans son statut subordonné

#### SOCIAL

L'objectif était triple : démocratie, indépendance nationale, progrès social

#### SOCIAL

La « révolution égyptienne » en cours illustre la possibilité de la fin annoncée du système « néolibéral », remis en cause dans toutes ses dimensions politiques, économiques et sociales

#### SOCIAL

Bien que le mouvement des jeunes reste diversifié dans sa composition sociale et ses expressions politiques et idéologiques, il se situe dans l'ensemble « à gauche »

## SOCIAL

La CIA tente néanmoins de renverser le sens du mouvement, d'éloigner les militants de leurs objectifs de transformation sociale progressiste et de les dévoyer sur d'autres terrains

## SOCIAL

Parallèlement un Conseil National (Maglis Watany) à été constitué par toutes les forces politiques et sociales acteurs du mouvement (les partis à orientation socialiste, les partis démocratiques divers, les syndicats indépendants, les organisations paysannes, les réseaux de jeunes, de nombreuses associations sociales)

## SOCIAL

Ce bloc peut être identifié dans les termes de ses composantes sociales (de classes, évidemment) mais il doit l'être tout également dans ceux qui définissent ses moyens d'intervention politique et des discours idéologiques au service de celle-ci

## SOCIAL

En termes sociaux, le bloc réactionnaire est dirigé par la bourgeoisie égyptienne considérée dans son ensemble

## SOCIAL

Il s'agit de révoltes sociales potentiellement porteuses de la cristallisation d'alternatives, qui peuvent à long terme s'inscrire dans la perspective socialiste

## SOCIAL

Or, il en a eu, et, très banalement, aggravé les difficultés sociales pour la majorité

## SOCIAL

L'impérialisme ne veut ni le progrès social, ni la démocratie pour ces pays

## IMPÉRIALISTE

Ils n'avaient pas tort, car le nassérisme ne s'est cristallisé comme projet anti-impérialiste qu'après Bandoung (avril 1955)

## IMPÉRIALISTE

Car les composantes du mouvement sont différents d'un pays à l'autre, tout comme le sont les formes de leur intégration dans la mondialisation impérialiste et les structures des régimes en place

## IMPÉRIALISTE

Mais ces révoltes, pour devenir des avancées révolutionnaires, devront surmonter de nombreux obstacles : d'une part, surmonter les faiblesses du mouvement, construire des convergences positives entre ses composantes, concevoir et mettre en œuvre des stratégies efficaces, mais aussi d'autre part mettre en déroute les interventions (y compris militaires) de la triade impérialiste

## ÉGYPTE

Je centrerai donc les réflexions qui suivent sur l'Égypte, dont on reconnaîtra sans difficulté le rôle majeur qu'elle a toujours rempli dans l'évolution générale de la région

## ÉGYPTE

L'Égypte a été le premier pays de la périphérie du capitalisme mondialisé qui a tenté « d'émerger »



## ÉGYPTE

Par deux fois, en 1840, puis dans les années 1870 par la prise du contrôle des finances de l'Égypte khédivale, enfin par l'occupation militaire (en 1882), l'Angleterre a poursuivi avec acharnement son objectif : la mise en échec de l'émergence d'une Égypte moderne

## ÉGYPTE

Cette interconnexion des objectifs n'est d'ailleurs rien d'autre que l'expression des effets de l'intégration de l'Égypte moderne dans le système du capitalisme / impérialisme mondialisé de l'époque

## ÉGYPTE

Le premier moment de ce demi-siècle de montée des luttes d'émancipation en Égypte avait mis l'accent – avec la constitution du Wafd en 1919 – sur la modernisation politique par l'adoption d'une forme bourgeoise de démocratie constitutionnelle et sur la reconquête de l'indépendance

## ÉGYPTE

Tout l'effort de la puissance britannique, avec le soutien actif du bloc réactionnaire constitué par la monarchie, les grands propriétaires et les paysans riches, s'est employé à faire reculer les avancées démocratiques de l'Égypte wafdiste

## ÉGYPTE

Israël peut alors s'engager dans la voie de l'expansion de sa colonisation de la Palestine occupée, avec la complicité tacite de l'Égypte et des pays du Golfe

## ÉGYPTE

Les puissances impérialistes prétendaient que ce régime « protégeait » l'Égypte de l'alternative islamiste

## ÉGYPTE

Les grèves ouvrières de 2007 – les plus fortes du continent africain depuis 50 ans – la résistance obstinée des petits paysans menacés d'expropriation par le capitalisme agraire, la formation de cercles de protestation démocratique dans les classes moyennes (les mouvements "Kefaya" et "du 6 avril") annonçaient l'inévitable explosion - attendue en Égypte, même si elle a surpris les « observateurs étrangers »

## ÉGYPTE

Les classes moyennes se rassemblent dans l'ensemble autour du seul objectif démocratique, sans nécessairement remettre intégralement en cause le « marché » (tel qu'il est) et l'alignement international de l'Égypte

## ÉGYPTE

Les paysans riches ont toujours constitué dans l'Égypte moderne une classe réactionnaire et ils le sont plus que jamais

## ÉGYPTE

Cette « aide » a permis aux plus hauts officiers de s'approprier des segments importants de l'économie compradore égyptienne, au point qu'on parle en Égypte de « la société anonyme / armée » (Sharika al geish)

## ÉGYPTE

La collusion entre les puissances impérialistes et l'Islam politique n'est d'ailleurs ni nouvelle, ni particulière à l'Égypte

ÉGYPTE

Ils ont toujours été un ennemi féroce des mouvements démocratiques en Égypte

ÉGYPTE

Il y a néanmoins des luttes violentes en perspective au sein des courants religieux islamistes en Égypte

ÉGYPTE

La Libye n'est ni la Tunisie, ni l'Égypte

ÉGYPTE

Le bloc au pouvoir (Khadafi) et les forces qui se battent contre lui n'ont rien d'analogues avec ce qu'ils sont en Tunisie et en Égypte

PRINTEMPS

Ces printemps coïncident donc avec « l'automne du capitalisme », le déclin du capitalisme des monopoles généralisés, mondialisés et financiarisés

ÉTAT

Le premier coup d'État des Officiers libres (1952), mais surtout le second inaugurant la prise de contrôle de Nasser (1954) sont alors venus pour « couronner » cette période de flux continu des luttes selon les uns, ou pour y mettre un terme, selon les autres

ÉTAT

Si, évidemment, l'appareil policier demeure intact (aucune poursuite n'est envisagée contre ses responsables) comme l'ensemble de l'appareil d'État (les nouveaux gouverneurs sont tous des anciens du régime), le Parti national démocratique a par contre disparu dans la tourmente et sa dissolution prononcée par la justice

ÉTAT

Islam ouvert, tolérant, insistant sur la conviction individuelle plutôt que sur la pratique des rites (« il y a autant de voies vers Dieu que d'individus » disent-ils), le soufisme égyptien a toujours été tenu en suspicion par les pouvoirs d'État, lesquels, néanmoins, maniant la carotte et le bâton, se gardaient d'entrer en guerre ouverte contre lui

ÉTAT

En réaction, les soufistes d'aujourd'hui sont laïcisants, sinon laïques ; ils appellent à la séparation entre la religion et la politique (le pouvoir d'État et celui des autorités religieuses reconnues par lui, l'Azhar)

ÉTAT

Khadafi avait dès 1969 exigé l'évacuation des bases britanniques et états-uniennes mises en place au lendemain de la seconde guerre mondiale

ÉTAT

Or l'Union Africaine refuse de l'accepter et jusqu'à ce jour aucun État africain n'a osé le faire

## Annexe B

Liste des segments du texte « 8 MAI 1945 – MAI 1985, 40ème anniversaire » contenant la séquence des mots sélectionnés : *MAI, NATIONAL, MOUVEMENT, COLONIAL, LUTTE, PEUPLE, FRANÇAIS, ALGÉRIENS, LIBÉRATION.*

### LIBÉRATION LUTTE MAI MOUVEMENT NATIONAL PEUPLE

Pour ce qui est du mouvement communiste et ouvrier algérien on peut affirmer que les événements du 8 Mai 1945 ont été déterminants pour forger progressivement dans les faits (et pas seulement en principe) sa ligne nationale et son caractère révolutionnaire. Ils ont beaucoup contribué à créer les conditions pour que le PCA soit en état de définir et de suivre une ligne juste pendant la guerre de libération nationale. La lutte menée par le PCA a pu ainsi être partie intégrante de celle de notre peuple. Le PCA a su intégrer son action propre dans celle du FLN en tant que large rassemblement patriotique. Ses militants armés ont été intégrés dans l'ALN, à la suite d'un accord réalisé entre les directions du FLN et du PCA en juin 1956.

### ALGÉRIENS COLONIAL FRANÇAIS MOUVEMENT NATIONAL

Après le discours de De Gaulle à Constantine, à la même période, un plan de réformes est envisagé pour l'Algérie. Mais les gros colons et la haute administration réussissent à en empêcher la mise en œuvre. Cela ne fait qu'aiguiser encore la contradiction entre l'ordre colonial et le mouvement national ascendant. Au point que l'ordonnance de mars 1944, abrogeant le code de l'indigénat et octroyant la citoyenneté française à quelques dizaines de milliers d'Algériens, il est désormais trop tard. Même les éléments modérés et assimilationnistes du mouvement national ont changé de camp : « L'assimilation est à éliminer. Depuis 115 ans, nous en sommes toujours au même point. Reste l'autonomie que les AML réclament » déclare F. Abbas dans un meeting à Sétif.

### ALGÉRIENS FRANÇAIS LUTTE MAI PEUPLE

Le 8 Mai 1945 a une charge émotive très grande pour nous. Il évoque avant tout la volonté de lutte de notre peuple et les massacres commis par le colonialisme français : des dizaines de militants d'Algériens froidement assassinés en particulier à Sétif et dans sa région, et à Guelma pour le seul "crime" de revendiquer l'indépendance de leur pays.

### COLONIAL LUTTE MOUVEMENT NATIONAL PEUPLE

D'ailleurs les patriotes qui avaient pu se faire des illusions sur les proclamations d'amitié avec les arabes prodiguées par la propagande nazie à notre peuple pour le détourner de la

lutte antifasciste, n'ont pas tardé à comprendre ce que signifiaient de telles proclamations. Dès l'instauration du régime pro-hitlérien de Vichy en France et de ses représentants en Algérie, le régime colonial a été extrêmement aggravé. La grosse colonisation a vu ses positions renforcées, plus que jamais, alors que le mouvement national, dans toutes ses composantes, a subi les rigueurs d'une répression impitoyable : incarcération de très nombreux démocrates et responsables, militants des différents partis et organisations nationaux, dans les prisons militaires et camps de concentration où beaucoup sont morts de maladies et de mauvais traitements, misère plus atroce pour les larges masses de notre peuple.

#### FRANÇAIS LIBÉRATION LUTTE NATIONAL PEUPLE

Les communistes (en particulier français, dont le poids était alors très grand en Algérie) ont lié d'une façon trop exclusive les appels à la lutte antifasciste à l'objectif partiel de la libération de la France occupée. Ils subordonnaient ainsi mécaniquement la lutte pour la satisfaction (ultérieure) des aspirations nationales de notre peuple à la libération de la France.

#### FRANÇAIS LUTTE MAI MOUVEMENT NATIONAL

Ce mouvement pour l'Amnistie aboutit eu vote d'une loi, dont le projet est déposé à l'Assemblée Nationale Française par le groupe communiste. Cette loi remet en liberté des centaines de détenus de Mai. Ce mouvement apparaît avec le recul comme un véritable maillon intermédiaire pour la reconstitution et la relance de la lutte du mouvement national.

#### LIBÉRATION LUTTE MOUVEMENT NATIONAL PEUPLE

Chaque fois qu'il a su assurer ce lien, le mouvement national algérien a tiré bénéfice de la lutte générale contre l'impérialisme et pour la paix mondiale. C'est ce qu'a souligné l'appel du 1er Novembre 1954 du Front de Libération Nationale, qui appréciait la détente internationale comme un facteur favorable au passage de notre peuple à la lutte armée.

#### COLONIAL FRANÇAIS MOUVEMENT NATIONAL

La poussée profonde et impétueuse du mouvement de revendication nationale s'alimente et aggrave en même temps l'affaiblissement de l'ordre colonial. Mais la résistance da la grosse colonisation et les tendances pro-colonialistes françaises ne sont pas vaincues.

#### ALGÉRIENS FRANÇAIS LIBÉRATION PEUPLE

Avec la libération des détenus et la reprise de certaines activités politiques et syndicales, conséquences des premiers grands revers du fascisme, une action militante intense se développe. La "Charte de l'Atlantique", dont l'un des articles se prononce en faveur « du droit de chaque peuple de choisir la forme de gouvernement sous laquelle il doit vivre » constitue la référence principale de la campagne de propagande. Les Américains avaient pris l'initiative, en relation avec leurs intérêts impérialistes à long terme, de faire connaître cette "charte" aux Algériens. Très vite, elle sera traduite en français et en arabe et diffusée à travers tout le pays.

#### FRANÇAIS LUTTE MAI PEUPLE

Nous rendons hommage à la mémoire de nos 45. 000 compatriotes tombés en Mai 1945 sous les balles et le feu des colonialistes, comme nous rendons hommage à celle des millions d'autres morts dans le combat antifasciste, ainsi qu'à tous les peuples du monde, en premier lieu les peuples de l'URSS. La victoire contre le fascisme, celles remportées par d'autres peuples comme le VietNam face au colonialisme français, ont encouragé notre

peuple pour se dresser et mettre fin au cauchemar colonialiste, au prix de plus d'un million de martyrs. La reconquête de notre indépendance a, à son tour favorisé en Afrique une longue série de victoires anticolonialistes renforçant ainsi le Front Mondial contre l'Impérialisme. Ainsi s'alimentent réciproquement les luttes de chaque peuple et celles du large front des trois grandes forces révolutionnaires mondiales.

#### FRANÇAIS LUTTE MOUVEMENT PEUPLE

Dans le même temps, cette victoire a accru le prestige de l'URSS et des organisations progressistes résistantes, en premier lieu les partis communistes qui tels le parti communiste français (PCF) et le parti communiste algérien (PCA), avaient déployé de grands efforts pour propager les idées libératrices du socialisme. C'est en grande partie, sur la pression de la lutte gigantesque menée par l'URSS et la poussée démocratique irrésistible déclenchée par la guerre des peuples contre l'agression et l'occupation fascistes, qu'a été inscrit dans la Charte de l'Atlantique le droit des peuples à disposer d'eux-mêmes, défendu jusque là essentiellement par le mouvement communiste et ouvrier international.

#### FRANÇAIS MAI NATIONAL PEUPLE

Les enseignements du 8 Mai 1945 ont appris à notre peuple et à l'ensemble de ses organisations nationales à s'unir, à s'organiser et à se mobiliser pour forger les conditions de la victoire contre le colonialisme français.

#### FRANÇAIS MOUVEMENT NATIONAL PEUPLE

Au moment où débute la 2ème guerre mondiale, en 1939, l'espoir est retombé au sein du mouvement national de voir une amélioration du sort du peuple algérien telle qu'elle avait pu se dessiner en 1936 avec l'instauration du Front Populaire en France. Après les conquêtes sociales arrachées par la classe ouvrière française, le Front Populaire, sous la direction des socialistes, évolue rapidement vers un abandon des aspirations populaires. En Algérie, les travailleurs qui s'étaient mobilisés ont eux aussi réussi à imposer certains acquis sociaux et syndicaux importants dont ils se servirent pour donner un nouvel essor à leurs organisations politiques. Mais le mouvement marqué par le Congrès Musulman Algérien regroupant les Oulémas, les Communistes et la Fédération des Elus, reflua à la suite de l'échec du projet Blum-Violette et des obstacles opposés par le gouvernement socialiste qui cède aux pressions de la grosse colonisation. Un fort mécontentement se crée.

#### ALGÉRIENS FRANÇAIS LUTTE PEUPLE

En ce sens, la lutte antifasciste ne pouvait pas et ne devait pas être indifférente au peuple algérien. Elle ne devait pas être laissée aux seuls antifascistes français ou alliés de la France installés sur notre territoire, elle devait être prise en charge et renforcée par les patriotes algériens dans une voie juste et conforme aux aspirations et aux intérêts immédiats et à long terme de notre peuple.

#### LIBÉRATION LUTTE MOUVEMENT NATIONAL

- d'une part la composition ethnique (à dominante européenne) du mouvement ouvrier en Algérie, reflet d'une forte colonie de peuplement. Ce mouvement ouvrier eut le mérite d'animer des luttes sociales importantes et d'introduire les idées socialistes en Algérie. Plus encore, l'organisation des communistes d'Algérie a été la seule sur le territoire national à revendiquer l'indépendance jusqu'en 1936. De nombreux communistes d'origine européenne sont tombés les armes à la main au cours de la guerre de libération,

ont connu la prison et la torture et participé activement dans des formes diverses au combat libérateur ;

#### LIBÉRATION LUTTE MOUVEMENT NATIONAL

- L'importance qu'il y a d'assurer correctement l'orientation du large mouvement de masse par une organisation et une direction révolutionnaires sensibles et ouvertes aux aspirations populaires, idéologiquement formées, politiquement expérimentées. C'est en ce sens que les communistes estiment qu'il faut maîtriser le contenu de classe à donner aux luttes de libération à chacune de leurs étapes. C'est ce qui explique l'importance que notre parti attache, encore aujourd'hui à la réalisation des tâches nationales démocratiques et à l'approfondissement de leur contenu social.

#### LIBÉRATION MAI NATIONAL PEUPLE

Mais l'hommage que nous rendons à nos milliers de martyrs de Mai 45, victimes de la répression barbare et précurseurs de la guerre de libération, est lié à la célébration par l'humanité progressiste du 40ème anniversaire de la victoire sur le fascisme qui a eu une portée considérable dans la vie de tous les peuples du monde. Nous partageons, avec de très larges secteurs de l'opinion publique mondiale, la conviction que la victoire antifasciste de Mai 45 a été pour toute l'humanité un moment décisif. Elle a non seulement mis fin aux terribles souffrances dues à la guerre et aux méthodes barbares de l'agression fasciste, mais elle a été aussi un puissant facteur d'émancipation nationale et sociale. Elle a été un facteur de démocratisation dans la vie de chaque peuple et de l'instauration de nouveaux rapports entre les nations.

#### LIBÉRATION MAI NATIONAL PEUPLE

Les événements de Mai 1945 dans notre pays (1er Mai à Alger et 8 Mai et les jours suivants dans le Constantinois ont été avant tout une manifestation de la conscience nationale de notre peuple et de sa volonté d'indépendance et de libération sociale et culturelle.

#### LUTTE MOUVEMENT NATIONAL PEUPLE

On peut considérer que les années 1915-1920 marquent la fin d'une phase dans le combat de notre peuple. Le rôle des villes est plus important, les nouvelles couches formées au sein de notre peuple par son contact avec la colonisation (lettrés, bourgeoisie moderne, prolétaires urbains et ruraux etc..) sont au-devant de la scène politique. Pendant deux décennies, de 1920 à 1940, vont se former les bases du mouvement national moderne avec la constitution des syndicats et partis politiques, la création des associations culturelles et religieuses, avec l'expérience des luttes électorales, la confrontation aux idées modernes de démocratie, de socialisme, etc.

#### COLONIAL LUTTE MOUVEMENT

La grande portée historique de la victoire antifasciste réside justement dans le fait que malgré toutes les tentatives de faire tourner en arrière la roue de l'histoire, ces 40 dernières années ont vu la naissance et le renforcement considérable du système socialiste mondial et le démantèlement presque total du système colonial, un déploiement sans précédent des luttes démocratiques et sociales et un puissant mouvement pour la paix, pour la vie et pour le développement économique jusqu'aux coins les plus reculés du globe et au cœur même des bastions de l'impérialisme.

#### COLONIAL LUTTE PEUPLE

En effet cette victoire antifasciste a abattu l'impérialisme le plus agressif et le danger le plus mortel et le plus menaçant pour les peuples du monde : le fascisme hitlérien, le nazisme. Elle nous a épargné le risque d'une occupation coloniale encore plus tragique. Elle a surtout créé un nouveau rapport de forces international plus favorable à notre lutte, d'autant plus que la France, puissance colonisatrice, a vu son poids affaibli dans le monde (objectivement et aux yeux de nos compatriotes).

#### COLONIAL MAI PEUPLE

Le 8 Mai 45 se situe dans la trajectoire du long combat qui a opposé notre peuple à l'occupant colonial.

#### COLONIAL MOUVEMENT NATIONAL

Il en fut de même pour les partis politiques. Malgré les restrictions persistantes - lois coloniales et lois de Vichy qui ne furent pas toutes abrogées - le PCA, le PPA et les autres courants du mouvement national multiplièrent leurs activités, soit de façon publique, soit semi-clandestine.

#### COLONIAL MOUVEMENT NATIONAL

Effondrement de la puissance mondiale de la France, affaiblissement de l'ordre colonial, possibilités nouvelles créées par le rapport des forces internationales d'un côté, situation économique et sociale catastrophique, puissant essor du mouvement de revendication nationale de l'autre côté : les éléments principaux d'une situation de crise sont en place. Cette dernière va se développer dans la première moitié de 1945, au moment même où se réalise en Europe l'écrasement des armées nazies et fascistes.

#### COLONIAL MOUVEMENT NATIONAL

Les premiers mois de 1945 précipitent l'évolution en cours depuis les deux dernières années. La perspective de la fin de la 2ème guerre mondiale avec la victoire des Alliés, la tenue de la conférence de San Francisco des Nations Unies le 24 avril, celle d'Alexandrie le 22 avril, qui doit poser les fondements de la Ligue Arabe, agissent comme des stimulants pour chacun des protagonistes -ordre colonial d'un côté, mouvement national de l'autre- en vue d'une solution à la crise.

#### ALGÉRIENS COLONIAL PEUPLE

Le régime des réquisitions s'installa : les fellahs devaient, fournir moutons et céréales en grandes quantités, une bonne part, de la production de la vigne était envoyée en Allemagne. Les colons et administrateurs, par esprit de revanche pour la peur qu'ils avaient eu en 1936, se faisaient arrogants et provocants vis à vis des Algériens. "Radio-Berlin" s'était tue. Le peuple algérien comprit la signification de la victoire du fascisme : loin de le libérer ou d'apporter une amélioration à son sort, comme on le lui promettait, cette victoire n'a fait qu'accroître l'oppression coloniale. Les organisations et courants politiques qui fondaient quelque espoir sur l'aide des hitlériens, ou plus prudemment sur la générosité du gouvernement de Vichy, ont vite perdu leurs illusions.

#### ALGÉRIENS COLONIAL NATIONAL

Après la constitution du régime collaborationniste du Maréchal Pétain en France (gouvernement de Vichy), les changements qui s'en suivirent dans l'administration coloniale en Algérie et l'installation des commissions d'Armistice allemande et italienne à Alger, la réalité ouvrit les yeux des Algériens. Les promesses de "Radio-Berlin" étaient loin d'être tenues et au contraire, l'Allemagne nazie s'allia et soutint le gouvernement

colonialiste de Vichy et la grosse colonisation qui releva ainsi la tête. L'oppression et la répression s'intensifièrent : les représentants des partis nationaux furent envoyés dans les camps : Kaddour BELKAIM, secrétaire du PCA et M. DOUAR, dirigeant du PPA, y perdront la vie. Comme il le faisait en France occupée, le fascisme voulait mettre à contribution l'Algérie pour faire vivre sa machine de guerre désormais tournée principalement contre l'URSS.

#### COLONIAL FRANÇAIS PEUPLE

En relation avec cette situation en Algérie et qui se retrouve, dans des conditions et formes différentes, en Indochine, Madagascar etc. se tient, en janvier 1944, à Brazzaville, une rencontre des hauts fonctionnaires français sous la présidence du Général De Gaulle. Le but était de mettre au point une politique permettant à la France, aux prix de quelques réformes, de conserver son empire colonial. Des oppositions fermes à toute évolution du statut colonial s'y expriment également, telle cette affirmation de Pleven, commissaire aux colonies au sein du gouvernement provisoire français : « Nous lisons de temps à autre que cette guerre doit se terminer par ce qu'on appelle un affranchissement des peuples coloniaux. Dans la grande France coloniale, il n'y a ni peuples à affranchir, ni discriminations raciales à abolir. Il y a des populations... qui n'entendent connaître d'autre indépendance que celle de la France".

#### COLONIAL LIBÉRATION LUTTE

L'expérience a montré enfin que si le renforcement de la lutte antifasciste dans le monde contribuait puissamment à favoriser les luttes de libération, cela ne pouvait constituer en soi la garantie que la victoire antifasciste déboucherait automatiquement sur l'octroi de l'indépendance par la puissance coloniale ou par ses alliés.

#### ALGÉRIENS MAI NATIONAL

En célébrant ce double anniversaire du 8 Mai 45, celui de la victoire mondiale antifasciste et celui de la monstrueuse répression colonialiste contre les patriotes algériens qui célébraient cette victoire en réclamant l'indépendance, nous tirons deux enseignements, étroitement liés au plan national et international :

#### ALGÉRIENS NATIONAL PEUPLE

Les nationalistes du Parti du Peuple Algérien (PPA), encore dans une semi-clandestinité, élargissent leur influence à travers le territoire national touchant les villages et campagnes en particulier dans le Constantinois. Le PPA édite clandestinement le journal "Action Algérienne", des tracts et organise des campagnes d'inscriptions murales et de boycott de lieux fréquentés par les européens (bars, cinémas, etc..). Cette action du PPA était relayée par un ensemble d'organisations contrôlées par ses militants, telles que les scouts musulmans Algériens (SMA), la Fédération des Petits Commerçants et Artisans, de nombreuses Associations culturelles et sportives.

#### COLONIAL FRANÇAIS LIBÉRATION

Les masses populaires et de nombreux dirigeants patriotiques constatèrent que la libération du joug colonial ne serait pas le résultat d'une explosion populaire demi spontanée (ainsi qu'il était pensé fréquemment dans des milieux nationalistes) ou par une France progressiste après sa libération (thèse défendue avec force par les communistes français et non combattue par le PCA).

#### FRANÇAIS LUTTE PEUPLE



La conviction que la France affaiblie pouvait être battue se développe : il suffisait simplement, pensait-on, de mobiliser et organiser les masses populaires et de les faire manifester leur volonté d'indépendance. Etant donné l'état moral et physique dans lequel se trouvait notre peuple., les difficultés d'organisation et de luttes politiques, l'expérience amère du Front Populaire, les gens n'envisageaient de possibilité de se libérer du colonialisme français que par l'intervention d'une autre puissance ; ce fut d'abord l'Allemagne, ensuite les Etats-Unis. On ne percevait pas bien encore la nature profonde du colonialisme et la solidarité fondamentale qui le liait aux autres pays impérialistes quelle que soit la forme de leur système politique.

#### FRANÇAIS MAI PEUPLE

La haine de notre peuple contre le colonialisme français se confirma et se renforça après les massacres du 8 Mai 45.

#### FRANÇAIS MOUVEMENT NATIONAL

Avec le recul, on peut penser qu'ils auraient sans doute contribué à une meilleure orientation du mouvement des masses de façon à désamorcer partiellement la provocation ainsi qu'à faire fructifier plus rapidement et plus efficacement les manifestations patriotiques dans l'opinion française et internationale ou à atténuer certaines conséquences immédiates de la provocation (reflux temporaire du mouvement dans l'opinion nationale).

#### FRANÇAIS MOUVEMENT NATIONAL

- d'autre part, l'influence déterminante qu'exerçait alors le PCF par divers canaux sur ce mouvement ouvrier et sur le PCA de naissance toute récente. En effet les dirigeants du parti communiste français appliquèrent à l'Algérie de façon mécanique dans les mêmes formes et à l'image de ce qui convenait à leur propre pays les justes orientations antifascistes du mouvement communiste international. Cela n'a pas facilité la tâche du PCA. Cela ne lui a pas permis d'imprégner ces orientations antifascistes et son action des particularités nationales résolument anticolonialistes comme cela fut réalisé par exemple par les communistes au VietNam. Ces derniers ont su appliquer et adapter les principes du marxisme-léninisme à leurs réalités nationales.

#### LIBÉRATION LUTTE NATIONAL

Ainsi cette orientation et forme de lutte se confirmait dans l'action. La répression reculait effectivement avec la libération de milliers d'emprisonnés et d'internés sous la pression de l'opinion nationale et internationale. Il se posa alors la question des perspectives de lutte de plus en plus clairement à toutes les organisations patriotiques, et provoquant à leur niveau des réajustements et des décantations.

#### LIBÉRATION LUTTE NATIONAL

Une des formes les plus nocives de ce sectarisme, ce sont les calomnies et manœuvres de diversion anticomunistes qui tendent à nier, minimiser au déformer l'action menée par les communistes, que ce soit à l'échelle internationale dans leurs luttes passées contre le fascisme, et leurs luttes actuelles pour la préservation de la paix mondiale, que ce soit au plan national dans leur apport patriotique et unitaire à la cause de la libération nationale et sociale.

#### LIBÉRATION MAI NATIONAL

C'est dans ces conditions que les communistes, qui ont mesuré rapidement le caractère erroné et nocif de leurs positions sur les événements du 8 Mai, prennent l'initiative de

lancer une campagne pour la libération des détenus. Rapidement les comités d'amnistie se mettent en place, au niveau national (avec comme SG Larbi Bouhali) et aux niveaux locaux. Des meetings, assemblées se multiplient, la presse communiste révèle chaque jour les méfaits de la répression de Mai etc.

#### LIBÉRATION MOUVEMENT NATIONAL

Dans cet esprit, l'enseignement majeur qui se dégage pour nous, de ce double anniversaire, c'est que la victoire antifasciste, comme l'a montré la suite, a été un puissant facteur d'impulsion et de maturation pour notre mouvement de libération nationale et sociale dans toutes ses composantes.

#### LIBÉRATION MOUVEMENT NATIONAL

Désormais la perspective est ouverte vers le développement de l'action unie des différentes tendances du mouvement national. Elle évoluera à travers des phases de rapprochement, et d'autres de conflit, sous des formes diverses et originales, et débouchera sur le combat commun dans la guerre de libération nationale.

#### LUTTE MAI PEUPLE

Le fait essentiel demeure que la signification profonde des manifestations de Mai 1945 et leur prolongement dans les mois suivants ont créé de nouvelles conditions et une nouvelle qualité dans la lutte de notre peuple et de toutes ses organisations.

#### LUTTE MOUVEMENT NATIONAL

C'est dans ces conditions "flottantes" que fut lancé le mot d'ordre d'action des communistes en faveur des "comités d'amnistie" et une campagne de dénonciation politique de la répression et du complot colonialiste, en s'appuyant sur toutes les possibilités d'un mouvement de masse légal. Ce mouvement prit rapidement une grande ampleur, entraînant en particulier les milieux patriotiques divers, encouragés par cette forme de lutte. Cette action apparut à beaucoup comme un maillon intermédiaire opportun entre la situation difficile du moment et une nouvelle période où le mouvement national pourrait reprendre l'offensive. C'est ce même type d'action unitaire qu'engagèrent les communistes et le journal "Alger Républicain" quelques années plus tard après les arrestations de centaines de membres de l'organisation spéciale, branche armée du PPA.

#### MAI MOUVEMENT NATIONAL

Dès l'Été 45, il apparaît que les événements de mai ont provoqué des divisions profondes au sein du mouvement national, rejetant les forces "modérées" vers le retour d'une politique réformiste vis-à-vis du colonialisme, et les forces radicales vers une clandestinité très difficile pour les conditions de l'époque.

#### MOUVEMENT NATIONAL PEUPLE

Tandis qu'au sein de notre peuple, au fur et à mesure que les informations lui parviennent, l'émotion et l'indignation dominant, les prises de position des divers courants du mouvement national reflètent la gravité de la situation et la volonté de préserver le minimum de possibilités légales d'action.

#### ALGÉRIENS NATIONAL

Dès le début de la guerre franco-allemande, les partis nationaux sont interdits : PPA et PCA sont dissous, leurs dirigeants arrêtés ou contraints à la clandestinité. La mobilisation militaire commence, des dizaines de milliers de jeunes algériens iront sur le front.

### ALGÉRIENS LIBÉRATION

Avec le débarquement anglo-américain en Afrique du Nord, en novembre 1942, l'espoir pour la libération est investi dans l'intervention des Alliés et en particulier des Américains. L'étalage de la puissance et de la richesse des Américains face à la faiblesse de la France, fait penser aux Algériens qu'une possibilité réelle de se débarrasser du colonialisme se présente.

### ALGÉRIENS FRANÇAIS

Cet amalgame grossier a pourtant été fait dans les premiers jours des événements par les communistes français et par les communistes algériens. Cet amalgame a constitué une grave erreur de jugement politique. Cette erreur prolongeait une erreur plus ancienne qui consistait à mettre sur le même pied certains mots d'ordre nationalistes avec les mots d'ordre des fascistes français, sous prétexte qu'ils se prononçaient les uns et les autres contre certaines positions de la gauche française.

### ALGÉRIENS FRANÇAIS

La grosse colonisation pousse dans le sens de la répression : Abbo, un de ses représentants, déclare : « Il y aura des événements et le Gouvernement sera obligé de revenir sur l'ordonnance du 7 mars », ordonnance qui rend citoyen français pouvant voter au 1er collège quelques milliers d'Algériens. De son côté le sous-préfet de Constantine, Lestrade-Carbone l'avoue, en avril : « il y aura des troubles et un grand parti sera dissous » .

### ALGÉRIENS LUTTE

Chaque fois battus les Algériens ont repris la lutte, profitant de la moindre occasion, du moindre changement dans le rapport des forces, global ou local.

### ALGÉRIENS LUTTE

On sait le rôle joué dans cette phase par les travailleurs algériens émigrés en France, notamment avec la création de l'Etoile Nord Africaine (l'ENA), née en liaison avec le dynamisme révolutionnaire du prolétariat mondial. Bien sûr dans la réalité les choses sont plus complexes que le schéma qu'on peut en dégager ; la lutte politique n'excluait pas par exemple l'utilisation de la violence ; mais dans l'ensemble, la forme violente de lutte était devenue adjacente, secondaire au cours de cette période.

### FRANÇAIS MAI

À la suite des affrontements qui ont lieu à Alger le 1er Mai et au cours desquels des soldats français tirent sur le défilé des AML -une dizaine de morts- la direction des AML décide de laisser aux organisations locales le choix des modalités pratiques des manifestations prévues pour le 8 Mai à travers la plupart des grandes villes. Elle insiste seulement sur la nécessité de veiller à ne pas tomber dans le piège de la provocation policière et pour cela recommande de ne pas être armés.

### LIBÉRATION MOUVEMENT

On ne peut qu'être surpris par la disproportion entre les événements eux-mêmes et la répression qui les a suivi. Cette répression était planifiée et avait un objectif clair : briser par la violence et la terreur la montée du mouvement populaire de libération.

### COLONIAL MOUVEMENT

Et il est vrai que l'administration coloniale ne restait pas les bras croisés face à l'évolution de la situation. Grâce à ses services de renseignements (on le sait maintenant) elle était au

courant des faits essentiels concernant le mouvement, ses projets, ses moyens. Après une période d'hésitation sur l'attitude à avoir (dissoudre les AML ou non ?), l'option avait été prise de laisser les choses se faire et casser le mouvement au moment propice.

#### COLONIAL MOUVEMENT

Divers éléments montrent que l'administration coloniale mettait en place, de façon calculée, les moyens d'une répression du mouvement.....

#### COLONIAL MOUVEMENT

Dans certaines localités, les forces de répression opèrent des arrestations et assassinats à partir de listes de militants nationalistes. L'ordre colonial veut effrayer les masses et aussi décapiter le mouvement.

#### LIBÉRATION NATIONAL

Victoire antifasciste mondiale : stimulant pour la libération nationale et sociale

#### ALGÉRIENS COLONIAL

Immédiatement après ces événements, la répression coloniale se met en place, de façon méthodique et décidée. Y prendront part le gendarmerie, les parachutistes, les troupes sénégalaises ainsi que les milices composées d'Européens. On utilise les automitrailleuses, les blindés, l'aviation et même la canonnade à partir d'un bateau dans la rade de Béjaia. Toute la région entre Béjaia et Sétif est ratissée, bombardée. Des massacres sauvages sont commis. On tue les Algériens par groupes entiers. On enterre dans des fosses communes ou on brûle à l'essence, voire dans des fours à chaux (à Héliopolis près de Guelma).

#### MOUVEMENT PEUPLE

Autour de ce mouvement se regroupent les "Élus", les Oulémas et le PPA. Les communistes inquiets des manœuvres néocolonialistes américaines, insuffisamment liés aux aspirations profondes de notre peuple et pour d'autres raisons, ne se sont pas associés à cette initiative. Ils rompaient de la sorte avec la tradition unitaire qu'ils avaient mis en oeuvre depuis la fondation de l'ENA et de l'expérience du "Bloc Ouvrier et Paysan" avec l'Émir Khaled (1926) jusqu'au Congrès Musulman de 1936. Fait plus grave, le PCF et le PCA ont considéré que ce mouvement faisait diversion aux exigences du combat prioritaire contre le fascisme et donc rendait service à ce dernier.

#### MOUVEMENT PEUPLE

Au début 1945, le mouvement se radicalise encore et à la conférence des sections AML en mars, les thèses du PPA l'emportent. La résolution adoptée ne fait plus référence à une « Algérie fédérée à la France » comme le stipulait le texte fondateur des AML. La conférence exprime aussi sa conviction que « les victoires (des Alliés) assureront bientôt la liberté à tous les peuples sur tous les continents ».

#### LUTTE MAI

Si on examine les événements du 8 Mai 45 dans cette perspective, en fidélité à cette logique populaire de résistance anticolonialiste à l'œuvre depuis plus d'un siècle, ils apparaissent comme l'annonce d'une nouvelle phase : celle où se réalise le début de jonction entre lutte politique et lutte armée, entre villes et campagnes, entre le nationalisme populaire des couches. petites bourgeoises et semi prolétariennes des villes avec l'irréductibilité de la paysannerie sans terre et de la petite paysannerie pauvre.

#### NATIONAL PEUPLE

- de la revendication nationale (développement des "Amis du Manifeste et de la Liberté" (AML) regroupant tendances nationalistes révolutionnaires et réformistes dont le PPA (Parti du Peuple Algérien de Messali), les Oulemas (association dirigée par des chefs religieux musulmans), le courant bourgeois libéral essentiellement avec Ferhat Abbas.

#### NATIONAL PEUPLE

Nous voyons donc que le lien est étroit et permanent (même s'il ne reste pas très visible, car l'impérialisme fait tout pour le cacher), entre l'intérêt national, l'indépendance dans la démocratie celui de l'édification économique et le progrès social, avec tous ces facteurs internationaux qui de notre peuple et des autres sont déterminants pour l'avenir des peuples de notre région.

#### FRANÇAIS MOUVEMENT

Une année après, en mars 44, à la suite d'un accord entre F. Abbas, Cheikh Ibrahimi et Messali, était fondé le mouvement des A.M.L. Dans l'article 4 du document fondateur on lit : « Créer un courant d'opinion en faveur du Manifeste, rendre familière l'idée d'une nation algérienne et désirable la constitution en Algérie d'une république autonome fédérée à une république française anticolonialiste et anti-impérialiste ».

#### LUTTE MOUVEMENT

C'est ainsi qu'au congrès du PPA de février 1947, furent critiqués les flottements et certaines initiatives aventuristes de la direction qui avaient caractérisé les orientations du mouvement pendant et après les événements de 1945. A ce congrès marqué par des débats houleux fut adopté également le principe de la préparation de la lutte armée. L'application de ce principe subit elle aussi plus tard les contrecoups des luttes entre tendances révolutionnaires et réformistes au sein du principal mouvement nationaliste.

#### FRANÇAIS PEUPLE

Cette réalité profonde, un journaliste français de la fin du XIX<sup>ème</sup> siècle l'a bien saisi : "Nous étions, dit-il, l'ennemi en 1894, nous le serons jusqu'à la fin des siècles. Nous serons éternellement pour l'Arabe, les gens qui lui ont volé ses terres, ses biens et sa liberté et cela s'apprend dans les écoles afin que les enfants s'en souviennent. Nous avons des tendances à l'oublier nous, mais là-bas, on le répète journellement et la plaie de la défaite saigne toujours. Faut-il que l'on soit insensé pour s'imaginer qu'un peuple de fanatiques, ajoutait le journaliste, qui a lutté trente ans pour son indépendance, qui a livré autour de ses étendards des combats de géants, qui a déjoué dans dix campagnes la science de nos généraux et l'héroïsme de nos meilleures troupes, se change tout à coup en un troupeau de petits apôtres dans les mains de nos pacifiques administrateurs ! .. On croit les musulmans réconciliés parce que décapités de leurs chefs et énervés dans leur force ; ils penchent prudemment la tête sous notre puissance et parce que dans les cortèges officiels, on voit parfois caracoler une douzaine d'Imams bariolés que l'on applaudit, sans se douter qu'ils sont en exécution à leurs coreligionnaires et traités par eux de traîtres et de chiens vendus aux chiens. Qui donc ici pourrait deviner quelle profondeur de ruse cache cette apparente soumission, quelles espérances de revanche brûlent sous cette résignation forcée et avec quelle fervente adoration dessous la tente, on chante, en langage mystérieux, les exploits du grand chef Abdelkader, en cherchant dans le ciel le signe infailible de son successeur pour le triomphe de l'Islam et l'extermination des infidèles."

#### FRANÇAIS PEUPLE

La défaite rapide des armées françaises face à l'Allemagne a été accueillie avec un grand intérêt par les masses algériennes : c'était d'abord la défaite de l'ennemi et la preuve que la France était vulnérable. Un fort courant de sympathie existait envers l'Allemagne, basé sur le raisonnement simpliste que l'ennemi de mon ennemi est mon ami ! Une active propagande hitlérienne dirigée vers l'Afrique du Nord, cultivait et exploitait ce sentiment d'espoir des peuples colonisés par la France. "Radio-Berlin" présentait Hitler et l'Allemagne comme la force qui allait libérer l'Algérie. "Houna-Berlin" était religieusement écouté.

#### LUTTE NATIONAL

A la source des erreurs et faiblesses des nationalistes, il y a la sous-estimation de l'importance pour la cause nationale de la lutte antifasciste. Certains (très minoritaires) ont appelé à des prises de position en faveur de la coalition hitlérienne. On peut cependant affirmer que la position la plus répandue était une position attentiste vis-à-vis de la lutte antifasciste, elle avait une signification essentiellement antifranaise et non de ralliement aux thèses du fascisme.

#### LUTTE NATIONAL

- l'importance qu'il y a de comprendre et d'assurer correctement le lien entre la lutte des forces patriotiques et progressistes sur le sol national et la lutte qui se déroule à l'échelle internationale contre l'impérialisme.

#### LUTTE PEUPLE

Pour notre peuple, ce 40ème anniversaire a donc une double signification : victoire mondiale sur le fascisme et lutte pour l'indépendance. Ce lien indispensable entre ces deux enseignements doit être perçu d'une façon scientifique conforme à la vérité historique objective et avec un esprit d'ouverture, en faisant reculer les interprétations unilatérales et étroitement partisans.

#### MAI MOUVEMENT

Au sein des formations nationalistes se poursuivirent également les déchantements. Le mouvement des AML, se scinda avec une aile réformiste (effrayée en particulier par la poussée populaire de Mai 1945) qui constitua l'Union Démocratique du Manifeste Algérien (UDMA) et l'aile majoritaire plus radicale et populaire, représentée par le PPA et sa branche légale le Mouvement pour le Triomphe des Libertés Démocratiques (MTLD). Au sein de ce dernier mouvement, se poursuivirent d'autres déchantements.

#### MAI MOUVEMENT

Malgré la volonté de certains dirigeants du PPA de généraliser l'insurrection à tout le territoire, afin notamment d'alléger la pression exercée par les forces répressives sur le Constantinois, il est désormais trop tard. Dès la fin Mai, c'est le reflux du mouvement.

#### MAI NATIONAL

Le 8 Mai 1945 une manifestation de la conscience nationale

#### MOUVEMENT NATIONAL

- La direction du mouvement nationaliste a connu pendant cette période des flottements, une impréparation et un manque de cohésion (dans les mots d'ordre politiques comme dans les consignes pratiques et d'organisation). On ne peut que mesurer le manque à gagner causé par l'erreur faite par les communistes en ne réalisant pas la jonction avec les

autres composantes du mouvement national sur la base de ce qu'il y avait de révolutionnaire dans leur Programme : la revendication de l'indépendance.

#### MOUVEMENT NATIONAL

Il est bien vrai que certaines positions étroites ou sectaires du PPA dans certains domaines contrastaient avec les activités plus ouvertes de l'Etoile Nord Africaine à ses origines. Mais l'erreur d'appréciation des communistes avait surtout ses racines dans la sous-estimation du mouvement national, de sa force et de son rôle anti-impérialiste, pour des raisons diverses.

#### MOUVEMENT NATIONAL

Ce réajustement radical de la ligne du PCA après 1945 se poursuit et trouva son couronnement au début des années 50 (6ème congrès de février 1952), en liaison avec la montée du mouvement ouvrier algérien et des courants progressistes au sein de notre mouvement national.

#### MOUVEMENT NATIONAL

la reprise et l'essor des activités du Mouvement National

#### MOUVEMENT NATIONAL

Dès novembre 1942, et jusqu'à 1945, on assiste à un développement impétueux, en surface mais surtout en profondeur, du mouvement national.

#### MOUVEMENT NATIONAL

Les "Élus", les Oulémas et le PCA dénoncent la provocation colonialiste, condamnent les "extrémistes du PPA" qui ont poussé le mouvement à l'aventure. Parmi ces prises de position, celle du PCA est violente quand elle établit un parallèle entre les ultra-colonialistes et les "pseudo-nationalistes". Il ne s'agit pas là simplement des effets d'une connaissance incomplète de ce qui s'était passé dans l'Est du pays. C'était en fait plus profondément le reflet du déséquilibre qui s'était opéré dans la ligne politique des communistes au détriment de leurs responsabilités nationales.

#### LUTTE

Grâce à la lutte antifasciste il s'est établi entre les années 1942 et 1945 dans notre pays (après le débarquement des troupes anglo-américaines) une situation où les soutiens les plus directs du fascisme (c'est-à-dire les colons et autres ultra-colonialistes) ont été neutralisés relativement. Cette situation a ouvert la voie à un essor et à une expression jamais vus jusque là :

#### LUTTE

- des revendications sociales ouvrières et paysannes contre les méfaits du colonialisme (luttés soutenues activement par le PCA et les syndicats CGT).

#### LUTTE

Il devenait plus clair que l'indépendance serait le fruit de luttes longues et difficiles, dans lesquelles le colonialisme n'hésiterait pas à renouveler les méthodes sanglantes qu'il perpétuait depuis plus d'un siècle.

#### LUTTE

La lutte pour la reconquête de la liberté a commencé le jour même où le colonialisme s'est implanté dans le pays. Les batailles ont été nombreuses, depuis celles dirigées par l'Emir Abdelkader jusqu'à l'insurrection des Aurès en 1916, en passant par celles de Fatma N'Soumer en Grande Kabylie (1857), des Ouled Sidi Cheikh (1864), El Mokrani et Cheikh

El Haddad (1871), Bouâmama (1881), Beni Chougrane (1911), etc... Les révoltes et insurrections ont marqué toute la fin du XIXème et les débuts du XXème siècles.

#### ALGÉRIENS

Durant cette période la production agricole connaît une chute constante due notamment à des années répétées de sécheresse (en 43, 44 et 45). Des produits essentiels tels sucre, café, huile, savon, ..tissus, etc. manquent. On institue des mesures de restrictions et de rationnement de la consommation : c'est le régime du "bon de ravitaillement". Mais ce rationnement est détourné par des fonctionnaires corrompus, des caïds et des administrateurs rapaces. Le "marché noir" et la corruption font leur apparition au détriment des plus pauvres et surtout dans les campagnes au seuil de la famine. De véritables fortunes se constituent grâce au "marché noir", aussi bien chez les Européens que parmi certains Algériens.

#### ALGÉRIENS

l'évolution de la guerre et l'attitude des Algériens

#### ALGÉRIENS

- participation immédiate et effective des musulmans algériens au gouvernement de leur pays ;

#### ALGÉRIENS

Le courant modéré voulait une démonstration pacifique de la volonté des Algériens de voir appliquer à leur profit les dispositions de la charte de l'Atlantique, tandis que les courants radicaux, représentés notamment par Lamine Debaghine, estimaient qu'il fallait faire une démonstration de force, voire même passer à l'action directe.

#### MOUVEMENT

Ces faiblesses du mouvement nationaliste, liées à une inexpérience assez normale et à l'impréparation idéologique, ont pu fournir des prétextes supplémentaires aux provocations colonialistes. Mais elles ne pouvaient en aucune façon être assimilées à une collusion entre les ultracolonialistes fascistes et les dirigeants nationalistes initiateurs ou organisateurs des manifestations de masse.

#### MOUVEMENT

Cet éveil de l'activité politique, ce mouvement profond de mobilisation qui traverse le pays, dans ses différentes couches sociales, va se manifester et se cristalliser à travers la création et l'Association des « Amis du Manifeste et de la Liberté »(A.M.L).

#### MOUVEMENT

Rapidement, le mouvement des AML connaît un immense succès : dans tout le pays se constituent des sections et très vite, il apparaît que les militants PPA y jouent un rôle déterminant. En avril 45, plus de 250 sections sont recensées regroupant, selon certaines estimations, environ 500 000 adhérents. L'hebdomadaire des AML "Égalité" tire à 130 000 exemplaires !

#### MOUVEMENT

Le courant modéré des AML, représenté par F. Abbas, Boumendjel, Ibrahim, etc. est débordé. Il n'a plus de prise sur l'orientation du mouvement. D'ailleurs c'est un mouvement dont la maîtrise échappe à tous, même si le PPA y joue un rôle dominant. La propagande menée durant des années a mis en éveil la profonde irréductibilité paysanne qui



sommeillait au sein des masses. Des rumeurs incontrôlées se développent. On parle de "La nécessité de collecter des armes", de "constituer des maquis", de "se tenir prêt pour le passage à l'action", etc.

#### MOUVEMENT

Une atmosphère fiévreuse règne au sein du mouvement dont deux caractères se dégagent : d'une part une extension en surface très large et très rapide, englobant des militants prêts à passer à l'action armée ; d'autre part, une organisation, une orientation et une direction relativement confuses.

#### MOUVEMENT

L'allusion de la direction des PC et du PCA aux « suspects, provocateurs, fauteurs de troubles », visaient en particulier les éléments infiltrés au sein du mouvement et manipulés par la police.

#### MOUVEMENT

A l'occasion du transfert provocateur de Messali Hadj, des incidents éclatent à Ksar Chellala qui permettent à la police d'opérer une cinquantaine d'arrestations de militants du PPA. « Les événements de Ksar Chellala, note Gazagne, nous ont donné une occasion unique pour prendre légalement des mesures qui ont décapité le mouvement PPA clandestin et nous ont vraisemblablement évité l'insurrection dans le département d'Alger, insurrection qui aurait gagné l'Oranie ».

#### MOUVEMENT

Ces deux tendances exprimaient la réalité profonde du mouvement, fort hétérogène dans ses conceptions et méthodes. C'est pourquoi, la faiblesse de l'organisation et la difficulté des communications aidant, une confusion certaine a régné sur le terrain.

#### MAI

#### 40ème ANNIVERSAIRE DU 8 MAI 1945 UNE DOUBLE SIGNIFICATION

#### MAI

Cette erreur d'appréciation sur la signification patriotique des manifestations du 8 Mai 45 fut heureusement corrigée dans l'action. Dès les premiers jours, malgré l'amalgame signalé plus haut, la répression colonialiste a été très activement dévoilée. Puis la responsabilité unique du complot colonialiste fut dénoncée, dénonciation suivie par des prises de positions et des initiatives dont la justesse ira en s'approfondissant au cours des mois suivants.

#### MAI

Ce que le 8 Mai 1945 nous a appris

#### MAI

#### LE 8 MAI 1945 EN ALGÉRIE ORIGINES, FAITS, ET CONSÉQUENCES

#### MAI

#### LES ÉVÈNEMENTS DU 8 MAI 1945 ET LEURS CONSÉQUENCES IMMÉDIATES

#### MAI

Des arrestations ont eu lieu également à Alger (notamment après le 1er Mai) et dans l'Oranie. Parallèlement, le G.G fait venir, en avril, des troupes stationnées en Tunisie et les consigne à Constantine.

#### MAI

Rappelons les faits essentiels : « Cela commence à Sétif, le 8 Mai, un mardi jour de marché. Le défilé des AML est autorisé et a lieu le matin, tandis que celui de la CGT est prévu pour l'après-midi. Des milliers de personnes se rassemblent près de la mosquée, y déposent tous les objets pouvant servir d'armes. Le défilé s'ébranle avec les SM en tête. Des banderoles portent les mots d'ordre : "Libérez Messali", "Vive l'Indépendance", "A bas l'impérialisme", "Vive les Nations Unies". Au milieu des drapeaux des Alliés est brandi le drapeau algérien.

#### MAI

L'émeute partie de Sétif s'étend, dès la soirée du 8 Mai, le lendemain 9 et le 10 Mai, à toute la région et en particulier vers le Nord. Des actions armées sont menées en divers endroits : El Ouricia, Aoaucha, Kherrata, etc. Des fermes sont attaquées, des postes pris d'assaut, des voies routières et ferrées coupées etc.

#### FRANÇAIS

- Ainsi, il ne fait pas de doute qu'il y a eu provocations délibérées et minutieusement préparées par les ultra-colonialistes (qui se tenaient jusque là sur la défensive et attendaient leur moment). Ces derniers ont fait jonction avec le néo-colonialisme gaulliste qui s'était efforcé de cacher par ses propositions de réformes (déclaration de Brazzaville) sa volonté intransigeante de défendre par tous les moyens l'intégrité de "l'empire français" !

#### FRANÇAIS

Très rapidement, avec la défaite des armées françaises, en 1940, une situation nouvelle va s'ouvrir où les choses ne seront plus comme avant. Peu à peu, l'aggravation de la situation économique et sociale dans le pays, l'évolution de la guerre, les transformations politiques internes, vont converger pour créer une situation de crise.

#### LIBÉRATION

Son programme politique est résumé dans les mots d'ordre suivants : Libération de Messali - Application de la charte de l'Atlantique - Indépendance de l'Algérie.

#### LIBÉRATION

- libération de tous les condamnés et internés à quelque parti qu'ils appartiennent.

#### NATIONAL

Le lien vivant entre le national et l'international

#### NATIONAL

Le PCA visait à élargir les libertés démocratiques, à épurer l'administration des éléments profascistes, à condamner les sabotages et méfaits des gros colons, administrateurs et Caïds. Régulièrement « Liberté » rendait compte en détail et dénonçait la situation économique et sociale dont souffraient les masses. Bien qu'au plan des principes le PCA réaffirma son attachement à l'indépendance nationale -notamment à travers l'expression de son adhésion intégrale et sans réserve aux clauses de la "Charte de l'Atlantique" à l'occasion de la conférence de Reprise en Août 1943- il estimait le moment inopportun pour mettre cette question au premier plan.

#### NATIONAL

Pour les communistes particulièrement, cette campagne va aider à redresser la ligne politique du PCA dans le sens d'une prise en charge plus ferme des exigences du combat

national. C'est ce qu'affirmeront et manifesteront les Comité Central et Congrès de 46 et 47.

#### COLONIAL

Du fait de la dépendance étroite de l'économie algérienne par rapport à la puissance coloniale, la rupture des relations commerciales consécutives à la guerre va entraîner une pénurie de pièces de rechange, etc. Des usines et ateliers sont contraints d'arrêter leur production ou de la réduire substantiellement.

#### COLONIAL

De tous les courants qui avaient convergé dans l'action commune au Congrès Musulman en 1936, seules les personnalités de la « Fédération des Elus » (Abbas, Bendjelloul, etc..) et certains Oulémas, ont maintenu une activité légale depuis 39. Au moment du régime de Vichy, des illusions ont amené ces personnalités à des contacts et démarches en vue de réformer le statut colonial de l'Algérie. Après l'échec, et le changement de situation, de nouveau ils perçoivent la présence des Américains comme une planche de salut.

#### COLONIAL

Le "Manifeste" était adressé beaucoup plus aux Américains et aux nations Alliées qu'à la France elle-même. Il était fondé sur l'idée qu'étant donnée la position difficile de la France, des pressions internationales, conformes aux dispositions de la Charte de l'Atlantique, pourraient aboutir à libérer l'Algérie de la domination coloniale.

#### COLONIAL

L'administration coloniale était persuadée, comme le note un rapport de police, que « les partis nationalistes veulent profiter de cet événement (l'armistice) pour créer des incidents et se mettre sous l'autorité des Alliés »

#### COLONIAL

Les autorités coloniales feront tout pour minimiser l'ampleur et la sauvagerie de la répression, feront le black-out sur les informations grâce à la censure de la presse. Mais la vérité se fait jour peu à peu et le bilan s'établira à 45 000 morts. En même temps, plus de 5000 personnes sont arrêtées à travers tout le territoire, parmi lesquelles des dirigeants et des personnalités politiques.

#### COLONIAL

Alors que les colonialistes se réjouissent du rétablissement par la violence de l'ordre colonial ("nous avons gagné dix ans de tranquillité"), c'est le désarroi et l'amertume au sein des masses. On pleura les morts et on a peur pour les détenus.

#### PEUPLE

Nous devons d'autant plus prendre en considération ce facteur international, que l'impérialisme et la réaction internationale ne se sont à aucun moment résignés aux reculs qu'ils ont enregistrés après la victoire antifasciste de 1945. Ils sont prêts à remettre en cause localement et dès qu'ils le peuvent, les conséquences du nouveau rapport de forces international, en noyant dans le sang les aspirations des travailleurs et des peuples. Comme ils l'avaient fait dès 1945-1946 en Algérie, en Grèce, en Indochine, à Madagascar et ailleurs.

#### PEUPLE

Mais il nous appartient à tous, nationalistes, communistes et autres patriotes, de voir que rien n'est définitivement acquis sans la vigilance et la mobilisation active des peuples, sans des orientations anti-impérialistes claires et résolues au niveau des gouvernements et des organisations politiques.

PEUPLE

Des tâches pressantes et vitales pour tous les peuples

PEUPLE

Sur cet arrière-fond économique et social, les effets de la guerre mondiale, son évolution ont un impact très fort, avec des alternances d'espoir et de déceptions au sein de notre peuple.

PEUPLE

Pour le PCA, l'impératif de l'heure était le soutien à l'effort de guerre des Alliés en vue de la victoire sur le nazisme et le fascisme, ennemis de tous les peuples.

PEUPLE

Ainsi, dès février 1943, Ferhat ABBAS et ses amis politiques, exprimant cette fois le sentiment de larges courants nationalistes, rédigent le « Manifeste du peuple algérien ». Dans ce document, il était demandé :

PEUPLE

- application pour tous les pays, petits et grands, du droit des peuples à disposer d'eux-mêmes ;