

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE
SCIENTIFIQUE
UNIVERSITE AKLI MOAND OULHADJE-BOUIRA



Faculté des Sciences et des Sciences Appliquées
Département d'Informatique

Mémoire de fin d'étude

Présenté par :

BAY Zahra

REZKI Tassadit

En vue de l'obtention du diplôme de **Master 02** en :

Filière : **INFORMATIQUE**

Option : ingénierie des systèmes d'informations et logiciels

Thème :

**Extraction des relations sémantique entre les mots en arabes :
Reconnaissance des entités temporelles**

Devant le jury composé de :

Mr .ABBAS AKLI	Dr	UAMOB	Président
Mr.ZERROUKI TAHA	Dr	UAMOB	Encadreur
Mm.BRAHIMI FARIDA	Dr	UAMOB	Examineur
Mr. HAYI MOHAMED YASSINE	Ma	UAMOB	Examineur

Année Universitaire 2018/2019

Remerciements

Tout d'abord, nous remercions le bon Dieu qui nous a aidés et qui nous a donné le courage et la patience pour réaliser ce travail.

Notre profonde gratitude et sincère remerciement vont à notre promoteur Mr Taha Zerrouki de nous avoir encadré, pour son suivi et son orientation. Nous le remercions pour ses conseils avisés, sa patience et pour le temps qu'il a bien voulu nous consacrer. Un grand merci particulier pour le Chef du Département de l'informatique Mr Djamel Benouar qui déploie de grands efforts pour assurer à leurs étudiants une formation actualisée.

Nous tenons à exprimer nos sincères remerciements à tous les enseignants de la qualité de cours dispensés, et une profonde satisfaction intellectuelle.

Nous tenons à présenter notre reconnaissance aux membres du jury pour nous avoir fait l'honneur d'accepter d'évaluer ce travail.

Nous remercions également nos parents pour leur contribution, leur soutien et leur patience.

En fin, nous adressons nos sincères remerciements à tous nos proches et amis, et à toute personne ayant contribué de près ou de loin à la concrétisation de ce travail.

Merci à toutes et à tous

Dédicaces

Je dédie ce modeste travail à :

*A mes très chers parents qui m'ont couvert
d'amour et de soutien*

A mes frères j'espère que la vie réserve le meilleur pour eux

*A mes oncles pour toute l'affection qu'ils m'ont
donné et pour leurs précieux encouragements.*

A tous mes chers amies surtout : Djamila, Hadda, Tassadit,

Yosra, Zoulikha et Fatma

A ma famille et tous mes proches.

Zahra

Dédicaces

Je dédie l'effort de ce modeste travail à :

*Mes chers parents, que nulle dédicace ne puisse exprimer
mes sincères sentiments , pour leur patience illimité, leur encouragement contenu,
leur aide, on témoignage de mon profond amour et respect pour leurs grands sacrifices.*

*Ma sœur Saliha, mon frère Djamel, pour leur grand amour et leur soutien qu'ils
trouvent, ici l'expression de ma haute gratitude.*

Mes très chères amies Kahina, Djahida, Sarah, Imane, Ahlem, Razika et Zahra.

Ma famille et à tous ceux que j'aime.

Tassadit.

Résumé

Face à la croissance de l'information, il existe une demande croissante d'applications robustes à utiliser pour la recherche des informations dans des textes et pour créer des modèles permettant d'extraire des connaissances sémantiques de textes. À partir de ça explorer les expressions temporelles et historiques particulier dans un texte donné, un texte historique ou une nouvelle, afin de mieux comprendre les textes et d'extraire les relations les plus importantes entre les mots et les phrases.

Les relations entre les mots dans un texte peuvent être syntaxiques comme sujet-verbe, verbe-objet, ou bien sémantiques comme les expressions idiomatiques, et les entités nommées (noms propres, places, villes,).

Dans ce projet, nous étudierons les expressions temporelles dans le cadre d'une étude détaillée afin de développer un système d'information capable d'extraire des dates du texte arabe et de les identifier sous différentes formes. Le but de ce travail est d'aider les chercheurs à extraire une partie des composantes linguistiques dans les actualités presse, les reportages d'actualité et les textes historiques, contribuant aussi à une meilleure compréhension du texte arabe.

Mots clés : Relation sémantique, Expression temporelle, Langue arabe, ...

Abstract

In the face of growth information, the demand for powerful applications is growing to be used to search for information in texts and to build models to extract semantic knowledge from texts. In addition, the historical and temporal phrases within a specific text were explored as historical text, in order to better understand the texts and draw the most important semantic relations between words and phrases. The words in the text are related to grammatical relationships such as : verb, doer of the verb , verb, object, or noun such as : named entities (names of places, names of persons, expressions of time , ect).

In this project, we will study the time expressions in detail, in order to develop an information system capable of extracting dates from the Arabic text and specifying them in

different versions. The aim of this work is to assist researchers in extracting part of the linguistic components in press releases, news reports and historical texts, thus helping to better understand the Arabic text.

Key words : Semantic relation, Temporal expression, Arabic language, ...

ملخص

في مواجهة النمو المعلوماتي، يزداد الطلب على تطبيقات قوية لاستغلالها في البحث عن المعلومات في النصوص وبناء نماذج لاستخلاص المعرفة الدلالية من النصوص. ومن ذلك، استكشاف العبارات الزمنية والتاريخية ضمن نص معين خبر كان أو نص تاريخي، من أجل فهم أكثر للنصوص واستخلاص أهم العلاقات الدلالية بين الكلمات والعبارات.

فالكلمات في النص ترتبط بعلاقات نحوية مثل فعل وفاعل، وفعل ومفعول، أو دلالية مثل

الكيانات المسماة (أسماء الأماكن، أسماء الأشخاص، تعابير الزمن...)

في مشروعنا هذا سندرس التعبيرات الزمنية دراسة تفصيلية، من أجل تطوير نظام معلوماتي قادر على استخلاص التواريخ من النص العربي وتحديدتها بمختلف صيغها. ويهدف هذا العمل إلى مساعدة الباحثين في استخراج جزء من المكونات اللغوية في الأخبار الصحفية والتقارير الإخبارية والنصوص التاريخية مما يساعد على فهم أكبر للنص العربي.

الكلمات المفتاحية: اللغة العربية، العبارات الزمنية، العلاقات الدلالية...

Table des matières

Table des matières	i
Table des figures	v
Liste des tableaux	vi
Liste des abréviations	vii
Introduction générale	1
1 Les relations sémantiques entre les mots	3
1.1 Introduction	3
1.2 Travaux antérieurs	3
1.3 Définition de la sémantique	4
1.4 Les types de la sémantique	4
1.4.1 Sémantique du lexique	4
1.4.2 Sémantique formelle et traits sémantiques	5
1.4.3 Sémantique fonctionnelle	5
1.4.4 Sémantique structuraliste	5
1.5 Définition des relations sémantiques	6
1.6 Les types des relations sémantiques	7
1.6.1 Les relations d'équivalence et d'opposition	7
1.6.2 Les relations hiérarchiques et d'inclusion	8
1.7 Le champ sémantique	10
1.8 La relation entre la sémantique et les expressions temporelles	11

1.9	Conclusion	11
2	Les aspects des expressions temporelle arabes	12
2.1	Introduction	12
2.2	Le rôle des entités temporelles dans les applications TAL	12
2.2.1	Recherche d'information	13
2.2.2	Système de question-réponse	13
2.2.3	Affectation automatique des périodes	14
2.3	Information temporelle	14
2.3.1	Information	14
2.3.2	La temporalité linguistique [22]	14
2.3.3	Le temps et l'information	17
2.4	Les classes d'information temporelle [26]	20
2.4.1	Les expressions temporelles	20
2.4.2	Les expressions absolues	22
2.4.3	Les expressions relatives	22
2.4.4	Les itératifs	22
2.4.5	Les événements	23
2.4.6	La circonstance du temps	23
2.5	Les différentes formules de l'expression temporelle arabe	24
2.5.1	Le jour	25
2.5.2	Le mois	25
2.5.3	L'année	27
2.6	Les propriétés spécifiques des expressions temporelles en langue arabe	27
2.7	Les difficultés pour la RET arabe	28
2.7.1	Agglutination	28
2.7.2	Les étiquettes des mois	29
2.7.3	Analyse morphologique	29
2.8	Conclusion	29
3	Reconnaissance et extraction des expressions temporelles	31
3.1	Introduction	31
3.2	Les approches pour la Reconnaissance des Entités Temporelles RET	31

3.2.1	Approche à base des règles	32
3.2.2	Approche d'apprentissage	33
3.3	Système à base d'apprentissage	33
3.3.1	ZamAn and Raqm : Extraction des expressions temporelles et numériques en arabe	33
3.4	Systèmes à base de règles	34
3.4.1	HeidelTime	34
3.5	Quelques systèmes d'annotation des expression temporelles	35
3.5.1	Extraction d'événements dans le cadre de TempEval 2	35
3.5.2	TimeML - un langage de balisage	36
3.6	Conclusion	38
4	Conception	39
4.1	Introduction	39
4.2	L'approche proposée	39
4.3	Description des données	40
4.3.1	Données en entrée	40
4.3.2	Données en sortie	40
4.4	L'algorithme de l'application	40
4.5	Protocole de système	41
4.6	Description détaillée du protocole de développement	42
4.6.1	Le pré-traitement	42
4.6.2	La segmentation	43
4.6.3	Le traitement linguistique	43
4.6.4	Application des règles	44
4.7	Construction des règles :	44
4.7.1	Les dates	45
4.7.2	Date relative	48
4.7.3	Les quantificateurs temporels :	50
4.7.4	Les heurs	50
4.8	Construction du corpus :	52
4.9	Conclusion	52

5	Implémentation	53
5.1	Introduction	53
5.2	Environnement de Travail	53
5.2.1	Environnement matériel	53
5.2.2	Environnement logiciel	53
5.3	Déroulement	54
5.4	Evaluation du système	55
5.5	Discussion	56
5.6	Conclusion	57
	Conclusion générale et perspectives	58
	Bibliographie	60
	A Les dictionnaires des mots qui expriment le temps	64
	B Titre de l'annexe	66

Table des figures

1.1	les relations sémantiques	6
2.1	La morphomogie d'un mot arabe	29
3.1	Les balises TimeML utilisées en français[39].	37
4.1	Processus général de la méthodologie du système.	42
4.2	Exemple d'agglutination.	43
4.3	Les classes et sous classes de l'expression temporelle.	45
4.4	AEF d'annotation des dates absolues en chiffre	46
4.5	AEF d'annotation des dates absolues en lettres.	46
4.6	AEF d'annotation des dates absolues en lettres et en chiffres	47
4.7	AEF d'annotation des dates absolues sous des formes irrégulières	48
4.8	AEF d'annotation des dates relatives	49
4.9	AEF d'annotation des quantificateurs temporels.	50
4.10	AFE d'annotations des heures relative	51
4.11	AEF d'annotations des heures absolue	51

Liste des tableaux

- 2.1 Les différentes formules du jours 25
- 2.2 Les différents formules du mois 26
- 2.3 Les différents formules d'année 27

- 4.1 L'algorithme générale de système. 41

- 5.1 Les résultats de déroulement de quelques textes 55
- 5.2 Les résultats de la mesure 56

- A.1 Les dictionnaires utilisées. 65

- B.1 Une partie de dataset. 67

Liste des abréviations

TAL	Traitement Automatique de Langue
ANLP	Arabic Natural Language Processing
IA	Intelligence Artificiel
ALESCO	Arab League Éducation Cultural and Sientific Organization
PRS	Patron de Relation Sémantique
TS	Type Sémantique
TimeMI	Time Machine Learning
ET	Expression Temporelle
RET	Reconnaissances des Entité Temporelles
AEF	Automate d'État Fini

Introduction générale

Introduction générale.

Le traitement automatique de la langue arabe a suscité l'écoulement de beaucoup d'encre scientifique durant les deux dernières décennies. Le TAL (Traitement Automatique de Langue) arabe ou ANLP pour Arabic Natural Language Processing en anglais, a connu un engouement des scientifiques dans les grands laboratoires de recherche et les grandes universités à l'instar de l'université de Stanford et l'université de Pennsylvanie aux Etats Unis notamment après les événements du 11 septembre 2001. Cet engouement est intensifié davantage avec l'apparition des réseaux sociaux qui peuvent être considérés comme un grand média d'échange entre le grand public sauf que ce grand public utilise pour la communication des langues vernaculaires non standards (les dialectes de l'arabe). En plus des points cités ci-dessus s'ajoute l'intérêt que donnent ALESCO (Arab League Educational, Cultural and Scientific Organization) et tous les pays arabes et particulièrement les pays de Golf à l'enrichissement et l'arabisation du contenu web et à l'informatisation de leurs gouvernements (gouvernement électronique). Un autre facteur à cet engouement, c'est la place qu'occupe la langue arabe dans le classement mondial des langues les plus utilisées sur Internet (la quatrième place devant le français par deux positions¹). Ce classement montre d'une façon claire que cette langue est en pleine expansion en termes d'utilisateurs sur Internet. Le monde arabe en tant que marché représente une grande opportunité aux grandes entreprises internationales de recherche et de développement et les pousse à adopter cette langue dans leurs produits informatiques.

L'extraction des informations et des connaissances sont des applications les plus connues et anciennes de Traitement Automatique des Langues. Malgré leurs fins différentes, elles

utilisent les mêmes approches et techniques pour réaliser ces fins. L'objet de l'extraction d'information est la création de représentation structurée d'information bien sélectionnée à partir de sources non structurées tels que des textes en langage naturel.

L'expression temporelles par ses différentes formes et types constitue un pourcentage considérable dans les textes du web, notamment les articles journalistiques. La détection et repérage et extraction de cette information relève du domaine de l'extraction d'information qui trouve actuellement son intérêt dans plusieurs domaines de technologie d'information et de communication telles que les systèmes de navigation, les moteurs de recherche, les systèmes question-réponse etc.

L'organisation des chapitres

Ce mémoire est organisé en deux parties et en cinq chapitres. La première partie regroupe les trois premiers chapitres dont l'objectif est d'introduire les notions, définitions, approches et en général le bagage théorique nécessaire à la compréhension des solutions proposées dans la deuxième partie.

Dans le premier chapitre, on détaille les notions en relation avec les relations sémantiques entre les mots arabe et on essaye de donner une synthèse sur l'état de l'art des différents niveaux de traitement, des outils, des ressources et corpus développés jusqu'à aujourd'hui.

Dans le deuxième chapitre, on introduit la notion des expressions temporelles en général tout en discutant le cas de l'arabe. A chaque fois si nécessaire on donne des exemples d'éclaircissement en arabe standard moderne.

- Dans le troisième chapitre, on entame en détail la reconnaissance et l'extraction des expressions temporelles. A la fin de ce chapitre on fait une synthèse d'état de l'art sur les approches et sur les systèmes de reconnaissances et d'extraction des expressions temporelles arabes.

La deuxième partie représente notre contribution et résume notre travail dans le cadre de ce mémoire. Il est organisé en deux chapitres (quatrième et cinquième chapitre).

Dans le quatrième chapitre, on expose la conception. L'approche utilisée pour construire l'outil de détection et d'extraction des expressions temporelles.

La mise en œuvre de notre travail et les résultats sont l'objet du dernier chapitre de ce mémoire.

Les relations sémantiques entre les mots

1.1 Introduction

La sémantique est la théorie du sens. C'est une discipline qui englobe plusieurs niveaux et conceptions du sens : lexicaux, formels, structurels et fonctionnels.

Elle étudie la relation entre signes et, en ce sens, la signification d'éléments langagiers. Il n'est pas facile de réaliser des traitements profonds des textes sans informations suffisantes sur la sémantique des termes et les relations sémantiques entre les mots constitutifs des textes, pour cela les chercheurs du traitement automatique de la langue naturelle n'ont pas écarté les questions de Synonymie et d'Antonymie et d'autres relations sémantiques, en raison de leurs valeurs ajoutées pour résoudre plusieurs problèmes comme l'analyse automatique des textes, la compréhension des textes, la traduction automatique,...

Dans ce chapitre, nous allons prêter attention aux importants points de la sémantique. Nous allons définir la sémantique et les types de la sémantique de manière générale. Puis nous allons nous approcher du principal sujet à savoir les relations sémantiques. Nous allons classer des types des relations. Ensuite, nous allons donner un exemple pour chaque relation pour bien comprendre.

1.2 Travaux antérieurs

La sémantique occupe une position importante dans le traitement de la langue naturelle. Quelques études récentes sur termes et relations sémantiques en corpus spécialisés : rapport entre patrons de relations sémantiques (PRS) et types sémantiques (TS) à Université de Montréal Faculté des études supérieures, L'objectif principal consiste à élaborer un

modèle d'association des patrons de relations sémantiques (PRS) aux types sémantiques (TS) auxquels appartiennent les termes[1].

Un autre travail sur l'extraction des relations sémantiques à partir du Wiktionnaire Arabe à Laboratoire LRI/Equipe SRF, Université Badji Mokhtar, Annaba, Algérie, son objectif est d'exploiter le Wiktionnaire arabe pour extraire les relations sémantiques et comment il peut contribuer à la création ou l'enrichissement d'un réseau lexical du domaine public pour l'arabe[2].

1.3 Définition de la sémantique

L'étude du langage et des signes linguistiques (mots, expressions, phrases) du point de vue du sens[3].

L'étude générale de la signification des signes conçue comme une relation entre les signes et leurs référents [4].

L'étude de la sémantique peut être philosophique (Sémantique pure) ou linguistique (Sémantique descriptive ou théorique)[5].

1.4 Les types de la sémantique

1.4.1 Sémantique du lexique

Comme les mots sont des petites unités de langage distinctes et facilement reconnaissables, beaucoup d'efforts ont été faits pour les définir voire, si possible, d'« imposer » leur sens de manière étymologique. Des quantités énormes de dictionnaires, d'encyclopédies et de méthodes de classification en sont témoins. Un point de départ naïf consiste à assigner à chaque mot un référent dans le monde réel (non linguistique), mais cela ne marche que pour les noms propres ou, à un niveau plus abstrait, les mots déictiques tels que je, tu, elle, ici, là, maintenant, hier. Cependant, la plupart des mots ne correspondent pas à des choses ou à des points de l'espace-temps, mais à des concepts. Et comme les concepts n'existent que dans les esprits, il arrive facilement que deux personnes comprennent le même mot de manière légèrement différente[6].

1.4.2 Sémantique formelle et traits sémantiques

La sémantique formelle s'efforce de réduire les énoncés à des propositions concernant des situations au moyen de valeurs de vérité définissables, de quantifications, etc. Dans l'idéal, ces propositions doivent pouvoir être interprétées selon un raisonnement logique et mathématique [7].

1.4.3 Sémantique fonctionnelle

[8] Selon J. R. Firth .1957 « on reconnaît un mot à ceux qui lui tiennent compagnie ». D'autres termes, la sémantique est une question d'emploi des mots et de système linguistique. Des mots adjacents s'associent sous forme de collocations. Certaines collocations sont figées à 100 % et peuvent être considérées comme une seule unité sémantique (pomme de terre, avoir lieu), tandis que d'autres sont composées mais pas figées (haute montagne, haute température, haute mer, haute trahison). Mais les associations de mots ne se limitent pas aux collocations. Les verbes en particulier ont des emplacements sémantiques pour leurs arguments syntaxiques . Par exemple, caresser dispose d'un emplacement de sujet pour un agent (humain) et d'un emplacement d'objet pour une entité « animée » (personne, animal ou partie du corps).

1.4.4 Sémantique structuraliste

Le structuralisme a fortement influencé la sémantique moderne. Il considère les langues comme des réseaux de relations systématiques. En partant du principe de composition les linguistes essaient de faire correspondre les relations sémantiques à des relations syntaxiques entre les sens des mots[9]. Les langues romanes, dont le français, exploitent la position des adjectifs dans un but sémantique : « grand » a un sens différent dans un homme grand et dans un grand homme.

L'ordre des mots joue aussi un rôle sémantique dans d'autres langues : par exemple, en

arabe, la différence entre حمزة أسد شجاع et حمزة أسد شجاع peut être déduite de l'ordre des mots : أسد dans la première phrase est une métaphore mais dans la deuxième est un animal. Toutes les sciences descriptives commencent par classer, et la sémantique ne fait pas exception. Il existe plusieurs systèmes de classification sémantique, dont les thésaurus et les index thématiques des bibliothèques. Les ressources linguistiques modernes telles que WordNet et les ontologies distinguent différentes relations sémantiques[9].

1.5 Définition des relations sémantiques

Une relation sémantique est une relation mettant en évidence les liens de signification entre les termes ou les classes d'un langage documentaire[10]. Nous pouvons trouver deux types de relations, des relations d'équivalence et d'opposition qui se rapportent aux unités de même rang, ce sont des synonymes et des antonymes. Ensuite, des relations de hiérarchie et d'inclusion qui se rapportent aux unités qui n'occupent pas le même rang et ce sont des hyponymes, hyperonymes, il existe aussi la relation partie-tout. Nous ne devons pas oublier la polysémie (monosémie) et les homonymes, les holonymes[11].

en résumé ces relations dans le schéma suivant (Figure 1.1)

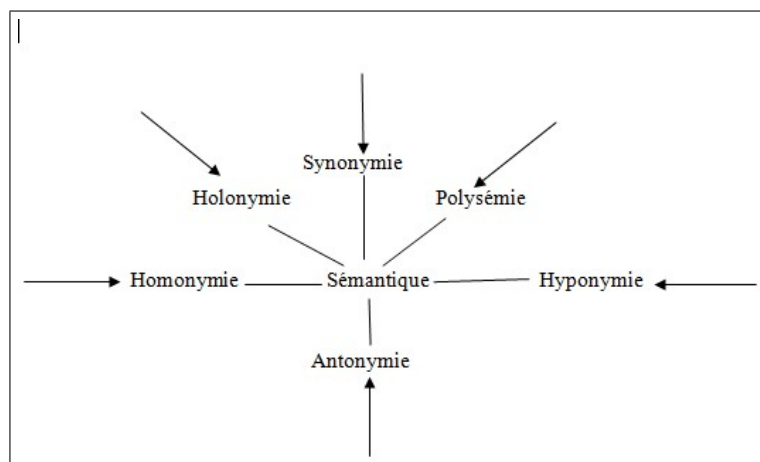


FIGURE 1.1 – les relations sémantiques

1.6 Les types des relations sémantiques

1.6.1 Les relations d'équivalence et d'opposition

1. La Synonymie : est une relation d'équivalence sémantique. Elle se distingue par la différente forme (ayant le même sens) dans plusieurs unités lexicales. Les synonymes prennent un signifié et des signifiants différents. Nous pouvons dire que nous remplaçons un mot par un autre dans un même contexte. Nous parlons des synonymes si le sens n'est pas modifié [12].

La synonymie présente les relations entre les signes et les choses mais aussi quand un objet a plusieurs noms par exemple : *أم* et *والدة* .

Il existe deux types de synonymie :

- Synonymie totale ou absolue :(le contexte n'intervient pas)
- Synonymie partielle ou contextuellet : dans le cas où les synonymes concernent des unités polysémiques, la synonymie ne porte que sur une acception. C'est le contexte qui permet de savoir quelle acception est à prendre en considération.
 - (a) La synonymie totale elle est très rare, concerne surtout les nomenclatures scientifiques ex : les doublets médicaux avec emprunts au latin et au grec. Par exemple : Le vocabulaire de la médecine présente des doublets, les uns souvent empruntés ou latins, les autres au grec, Par exemple : ictère, hépatite.
 - (b) La synonymie partielle est la plus courante Les synonymes peuvent varier en fonction des différents sens des mots polysémiques, Par exemple : Remplir = occuper une fonction Remplir = compiler un formulaire

2. L'antonymie

L'antonymie est le contraire de la synonymie. Alors, l'antonymie est une relation entre deux mots avec le sens contraire. Nous devons accentuer que ces deux termes doivent avoir quelques traits qui les mettent en relation pertinente. Nous montrons

un petit exemple : ordinateur et chanter : les deux mots ne sont pas synonymes ni antonymes, ils n'ont pas une relation pertinente entre eux. Ils existent les antonymes des qualités, des valeurs, des quantités, des dimensions, des déplacements, des rapports chronologiques. Mais pas chaque mot n'a son contraire.

Classement des antonymes

- Les antonymes complémentaires ou non gradables par exemple : *حي* ، *ميت* et *متزوج* ، *أعزب* . Ces types de paires n'ont que la négation de l'un ressort l'affirmation de l'autre et à l'envers.
- Les antonymes gradables par exemple : *كبير* ، *صغير* et *فقير* ، *غني* . Ce type des antonymes se présente par des expressions entre lesquelles nous pouvons interposer d'autres termes par gradation : *كبير* ، *متوسط* ، *صغير* . Nous pouvons aussi constater que la gradation est fondée sur la comparaison par exemple : *برج خليفة أكبر من برج إيفل* .
- Les antonymes réciproques par exemple : *اشترى* ، *باع* et *زوجة* ، *زوج* . Ce type des antonymes nous interprète qu'un mot est le contraire de l'autre mot. Ces paires des mots signifient la même relation, mais l'ordre des arguments est inversé. Nous montrons un exemple pour compréhension : *عبد الغني باع منزلا : لعلي فيعني هذا أن عليا اشترى منزلا من عبد الغني*.
- Un autre type d'antonyme est appelé opposition directionnelle. Son exemple est la relation entre des mots tels que : *يذهب* ، *يأتي* et *يغادر* ، *يصل*, et de va-et-vient, qui combinent le mouvement dans l'une des deux directions opposées pour un lieu.

1.6.2 Les relations hiérarchiques et d'inclusion

1. **L'hyponymie et l'hyperonymie** : Ces deux relations lexicales sont très semblables. L'hyperonymie est caractérisée par sa supériorité. Alors, nous comprenons

l'hyponymie comme un terme supérieur. Nous donnons un exemple : *زهرة* qui est un hyperonyme de *خزامى*. Par contre, l'hyponymie est un terme subordonné alors *التوليب* est un de plusieurs hyponymes de *زهرة* tel que : *وردة، بنفسج* : Alise Lehmann définit l'hyponymie et l'hyponymie de cette façon : «La relation d'hyponymie est une relation hiérarchique qui unit un mot spécifique (sous-ordonné), l'hyponyme, à un mot plus général (super-ordonné) nommé l'hyperonyme» [11].

2. Homonymie, polysémie et holonymie

(a) **L'homonymie, l'homophonie** : Les homonymes sont des mots qui ont la forme identique, mais leurs significés sont différents. Par exemple : *المغرب البلد* et *المغرب وقت الصلاة*.

Les homophones se prononcent de la même façon mais ces homophones n'ont pas la même graphie. Par exemple : *كرسي*, « Polguère » a pris une autre définition pour décrire les homonymes : «Deux lexies sont des homonymes si elles sont associées aux même signifiants, mais ne possèdent aucune intersection des sens notable» [13].

(b) **La polysémie** : La polysémie est étroitement en rapport avec l'homonymie et nous pouvons les confondre. Cette expression signifie qu'un mot peut avoir plusieurs sens. Donc, nous pouvons exprimer qu'un signe a plusieurs significés. Mais nous devons aussi mentionner la monosémie qui est opposée à la polysémie. Ce terme indique qu'un mot a un signifiant et un signifié unique. Par exemple :

الحظيظ اسفل الجبل et *الحظيظ الرجل المحظوظ*. La monosémie fait partie des vocabulaires de spécialité dont les expressions n'ont qu'un seul sens. Par rapport à la monosémie, des expressions polysémiques apparaissent plus fréquemment.

Ils se trouvent entre les vocabulaires communs. Par exemple : *والعين : العين*
الأولى للنظر، والثانية من الإصابة بالعين

أما الثالثة فهي العين التي هي الميل في الميزان

أما العين الرابعة فهي العينُ السحابة أو المطر متواصل الهطل أياماً

Dans le dictionnaire de linguistique et des sciences du langage "Larousse" est la polysémie définie comme : «Une propriété d'un signe linguistique qui a plusieurs sens. Le concept de polysémie s'inscrit dans un double système d'opposition : l'opposition entre polysémie est homonymie et l'opposition entre polysémie et monosémie" [14].

- (c) **L'holonymie et la méronymie** La relation de méronymie est la relation qui s'établit entre une partie et son tout. Elle est asymétrique et sa réciproque, la relation entre un tout et l'une de ses parties, est l'holonymie. C'est une relation qui opère principalement entre deux noms, bien que Winston et al. (1987) proposent une relation FEATURE/ACTIVITY pour les couples désignant une étape dans un processus comme paying/shopping. La définition que donne Cruse (1986) de la méronymie est la suivante " X is a meronym of Y if and only if sentences of the form A Y has Xs / an X and An X is a part of a Y are normal when the noun phrases an X, a Y are interpreted generically". Ainsi, La main est une partie du bras est vrai, et ce même s'il existe des bras dont la main a été coupée. La relation de méronymie est prise en compte dans la construction de thésaurus et d'ontologies [13]. Par exemple : يد est un méronyme de جسم , de même que سقف est un méronyme de منزل .

1.7 Le champ sémantique

Le champ sémantique est l'ensemble des significations que peut prendre un mot ou un groupe de mots dans des contextes déférents. Par exemple, on dit que les mots pou, puce, punaise et tique appartiennent au champ sémantique des parasites corporels de l'homme. Plus exactement, le champ sémantique est un système clos défini comme l'association d'un champ lexical et d'un champ conceptuel ; dans la mesure où le champ conceptuel peut être nommé par un mot (plutôt que par une paraphrase), le champ sémantique est l'association d'un ensemble de termes spécifiques et d'un terme générique [15]. Par exemple :

اندلعت الحرب العالمية الأولى في يوم الثامن والعشرون من حزيران يونيو عام الف وتسعمائة (تاريخ، وقت، حدث). cette phrase appartient au champ sémantique وأربعة عشر ميلادي.

1.8 La relation entre la sémantique et les expressions temporelles

L'information temporelle peut être normalisée, car quelles que soient les modalités ou la langue utilisée, chaque expression temporelle se référant à la même sémantique utilisée peut être normalisée à la même valeur dans un format standard. Cette propriété rend le terme information temporelle et la langue indépendante [16]. L'expression temporelle est un ensemble de mots, ces mots sont des informations temporelles, le champ sémantique de ces expressions est la temporalité.

1.9 Conclusion

Les textes, supports d'échanges au sein d'une communauté technique ou scientifique, sont en effet des sources de connaissances stabilisées dont l'exploitation en partie automatisée permet de recueillir rapidement un noyau de connaissances structurées. Ils sont un point de départ pour identifier les termes utilisés dans un domaine et le sens de ces termes à travers les relations lexicales qui les associent.

Dans notre travail, nous avons présenté un état de l'art sur la sémantique mettant l'accent sur les types de la sémantique. Nous avons également présenté un état de l'art sur les relations sémantiques, en mettant l'accent sur les relations sémantiques dans la langue arabe. Ensuite, nous avons donné des exemples en arabe pour chaque relation. Afin d'expliquer la relation entre la sémantique est les expressions temporelles. Les relations sémantiques jouent donc un rôle capital pour accéder aux définitions des concepts tels que les acteurs du domaine les expriment dans les textes.

Les aspects des expressions temporelle arabes

2.1 Introduction

Le traitement automatique de l'information temporelle est défini pour la connaissance des éléments qui explique ce type d'information dans les textes en langage naturel, à les interpréter afin de leur donner une valeur univoque dans l'espace du temps et à en fournir une représentation normalisée. Le résultat de ce traitement est essentiel pour une exploitation aisée et complète des informations temporelles. Plusieurs applications peuvent recourir à la temporalité dans différents domaines, comme la recherche ou l'extraction d'informations, afin de satisfaire leur objectif.

Les expressions temporelles sont des éléments intéressants pour l'indexation de textes. Si l'aspect thématique représente une dimension très changeante de l'information, la dimension temporelle est plus stable que ce soit du point de vue du son expression qu'en ce qui concerne sa distribution dans les différents textes[17]. Les informations thématiques demandent une méthode adaptable au domaine exploité tandis que le traitement de l'information temporelle reste une tâche assez peu variable.

2.2 Le rôle des entités temporelles dans les applications TAL

L'extraction d'informations à partir de textes est un processus non trivial qui construit un modèle de connaissances valide, nouveau, potentiellement utile et au final compréhensible à partir de textes bruts. Elle est un processus à l'intersection de plusieurs domaines dont notamment celui l'apprentissage, mais également celui du traitement automatique des

langues, ou encore de la recherche d'informations et de l'extraction d'information[18].

Dans ce travail, on penche sur l'extraction d'entités temporelles qui est un sujet de recherche de longue date dans le traitement du langage naturel et a été utilisé pour aider, entre autres, l'acquisition de connaissances, la recherche d'informations et de réponse aux questions.

2.2.1 Recherche d'information

Il s'agit d'identifier et de récupérer des documents pertinents à partir d'un ensemble de données en fonction d'une requête d'entrée. La recherche d'information peut bénéficier de la reconnaissance et classification des entités temporelles en deux phases : premièrement, la reconnaissance des ET dans la requête d'une part ; et reconnaître les ET dans les documents recherchés d'autre part, puis extraire les documents pertinents en tenant compte de leurs ET classées et comment elles sont reliées à la requête. Par exemple : si on recherche la date 1962 جويلية 05. Le résultat de cette recherche est : l'extraction des documents pertinents contenant cette date[19].

2.2.2 Système de question-réponse

C'est très similaire à la recherche d'informations mais avec des résultats plus sophistiqués. Un système de question-réponse prend en entrée des questions en langage naturel et donne en retour des réponses concises et précises. La tâche de RET peut être utilisée dans la phase d'analyse de la question afin d'y reconnaître les ET qui aideront plus tard à identifier les documents pertinents et à construire la réponse à partir des passages pertinents.

Les systèmes de question-réponse pourraient bénéficier substantiellement de la RET, parce que la réponse à de nombreuses questions. Par exemple, la réponse aux questions « متى؟ » « Qui concerne la date, « كم دام؟، كم استمر؟ » concerne la durée[19].

2.2.3 Affectation automatique des périodes

Les timbres du contenu d'un document numérique plutôt que de ses métadonnées, telles que la dernière modification et les dates de création d'un fichier, présente un intérêt à la communauté des chercheurs. Li et al de Microsoft a déposé un brevet pour leur application qui extrait les dates d'auteur. Ils supposent que la dernière date de modification d'un document n'est pas représentative de sa date de production. Les documents peuvent être téléchargés ou copiés sur des sites Web collaboratifs et la méta-date est remplacée par la date de téléchargement, ce qui est rarement significatif [20].

L'ouvrage [5] aborde le problème de l'affectation automatique des périodes dans des documents tels que des articles de journaux, des rapports médicaux ports et documents légaux pour faciliter la récupération de la documentation. Périodes des événements examinés dans un document peuvent revêtir une importance plus grande, et aussi la date à laquelle le document a été rédigé.

2.3 Information temporelle

2.3.1 Information

L'information est un concept ayant plusieurs sens. Il est étroitement lié aux notions de contrainte, communication, contrôle, donnée, formulaire, instruction, connaissance, signification, perception et représentation. Au sens étymologique, l'information est ce qui donne une *forme* à l'esprit. Elle vient du verbe latin *informer*, qui signifie « donner forme à » ou « se former une idée de ».

L'information désigne à la fois le message à communiquer et les symboles utilisés pour l'écrire, elle utilise un code de signes porteurs de sens tels qu'un alphabet de lettres, une base de chiffres, des idéogrammes ou pictogrammes. Hors contexte, elle représente le véhicule des données comme dans la théorie de l'information et, hors support, elle représente un facteur d'organisation.

2.3.2 La temporalité linguistique [22]

Le temps peut être considéré comme une localisation temporelle, il indique une référence temporelle déictique par rapport au temps de l'énonciation, une référence temporelle ana-

phorique par rapport à un moment donné dans le discours. Chaque événement est lié à des informations temporelles, figuré dans un agenda ou un horaire, les informations temporelles sont véhiculées par les temps verbaux dans la datation et la chronologie.

Dans la vie courante, nous manipulons des concepts temporels, qui en résultent d'un système de représentation. Il existe bien alors une relation entre le monde réel et ces concepts temporels.

Beaucoup de travaux de recherche ont abordé la temporalité dans le langage naturel comme l'extraction d'informations.

Le temps peut être exprimé par de nombreux moyens, comme des expressions adverbiales, des connecteurs qui ont une dimension temporelle « et » pour expliquer la suite de deux actions. Enfin, les événements ont aussi une dimension temporelle, implicite, auxquels une granularité naturelle leur est associée.

Un autre moyen qui véhicule la dimension temporelle : les dates dites absolues qui sont invariables et les dates dites relatives. Egalement, on peut recourir à des expressions temporelles imprécises, cependant ce type d'expression peut conduire à une ambiguïté, l'exemple suivant illustre cette ambiguïté :

-En mars, les prix des véhicules vont baisser ou *في شهر مارس ، سوف تنخفض أسعار السيارات*, dont la période n'est pas claire.

L'organisation d'un texte dans le domaine temporel, nécessite des moyens véhiculant la dimension temporelle, cependant l'étude de la temporalité dans les textes connaît un certain nombre de problèmes en fonction du type du moyen qui véhicule la temporalité.

Il existe des mécanismes de raisonnement qui permettent de calculer l'extension temporelle des périodes. Ce raisonnement permet d'effectuer un changement des instances selon un modèle linguistique sous forme d'intervalles de temps qui représente un modèle calendaire, pour lesquels un statut d'accessibilité est attribué[21].

La localisation temporelle

La localisation temporelle peut être exprimée par des représentations qui peuvent être groupées en trois groupes en fonction de l'information de base considérée pour former des informations temporelles :

- **Le point** : fait référence à une suite ordonnée d'instants ponctuels, l'associer à un fait revient à lui associer tous les instants pour lesquels le fait est vrai.
- **L'intervalle** : la valeur d'un événement est connue par un ensemble d'intervalles.
- **L'occurrence** : dans ce cas, le temps est issu des occurrences qui se produisent, chaque occurrence définit un instant.

Traitement de la temporalité

Tout traitement de temporalité prend en compte des étapes permettant l'aboutissement à des objectifs bien déterminés. Avant de faire l'extraction, un prétraitement est nécessaire pour réaliser le traitement, c'est pourquoi l'interprétation des informations temporelles est réalisée. De l'autre, les événements eux aussi véhiculent la temporalité, ce qui est inclus dans le traitement de la temporalité. Les informations temporelles que l'on souhaite extraire doivent être interprétées afin de leur donner une valeur univoque, c'est-à-dire lui attribuer une valeur temporelle relative à un système calendaire.

Mais avant de les interpréter, il faut d'abord les reconnaître en parcourant l'ensemble des informations temporelles disponibles, permettant ainsi une exploration complète.

Plusieurs auteurs ont consacré leurs efforts à ces opérations dont le but de les approfondir. Nous pouvons citer les travaux de Bittar [20], dont il explique clairement la procédure d'annotation ainsi que la prise en compte des événements qui eux aussi entraînent la temporalité, cependant les événements englobent un large éventail et correspondent à des aspects différents, mais les systèmes d'extraction d'informations temporelles se limitent à un type bien défini selon les besoins ainsi que les objectifs visés par le système.

Parmi les événements considérés par les différents systèmes, nous pourrions distinguer les verbes, ainsi que les événements qui sont véhiculés par des verbes finis, alors que d'autres considèrent les verbes et les groupes nominaux.

Selon le format d'annotation Time ML, les événements considérés sont ceux qui engendrent des verbes, des états, des noms événementiels ou encore des adjectifs. Il existe aussi des relations entre les événements et les expressions temporelles, qui consistent à

établir des relations entre les événements et le temps calendaire[23].

2.3.3 Le temps et l'information

La spécification de l'information temporelle constitue un enjeu au niveau des systèmes d'annotation automatiques, du fait qu'ils doivent être en mesure de savoir quelles sont les unités qui expriment l'information temporelle, ainsi que les niveaux de représentation qui permettent une appréhension de la sémantique du temps. Les informations temporelles peuvent regrouper des expressions de dates comme le 16 janvier, des expressions d'heure comme midi, des expressions de durée comme 2 minutes, elles peuvent regrouper aussi des événements sous forme de verbes, de noms déverbaux, ou autres noms, ou encore les états sous forme d'adjectifs ou verbes. Ces informations ont une importance qui réside dans le fait qu'elles assurent une bonne compréhension du texte[18].

Trois aspects importants de l'information temporelle peuvent être dégagés :

- Des groupes de mots véhiculant l'information temporelle qui facilitent la reconnaissance et l'annotation.
- L'interprétation de groupes de mots pour donner une valeur relative.
- Le dernier aspect, la structuration du discours par les valeurs temporelles.

D'après Turenne, l'information temporelle peut aussi être vue selon trois aspects[18] :

- Un aspect linguistique dans lequel, l'accent est mis sur la sémantique du fait qu'elle permet de résoudre le problème d'ambiguïté qui se trouve dans une expression ou dans une phrase.
- Un aspect statistique dans lequel un document contient des champs classiques à savoir le titre, le résumé et la date de création. Ces informations calendaires peuvent ordonner chronologiquement les documents.
- Aspect ontologique qui prend en compte les modèles de temps en intervalles, tout en négligeant la notion de temps. proposent des intervalles temporels qui ont été pris par l'intelligence artificielle.[18]

Les expressions qui expriment le temps

Nous distinguons les expressions suivantes :

1- Les adverbiaux temporels : Les adverbiaux temporels véhiculent des informations importantes concernant les relations qui existent entre les événements, un problème

se pose lors de la reconnaissance de l'évènement auquel ils sont rattachés et la déduction des informations temporelles qu'ils expriment.

Le temps peut être exprimé par des expressions qui ont une valeur temporelle. Les expressions qui sont regroupées sous le nom d'adverbiaux temporels, selon Laurent sont [17] :

- Les adverbiaux de référence temporelle comme aujourd'hui
 - Les adverbiaux de référence temporelle comme aujourd'hui
 - Les adverbiaux de durée : en cinq semaines
 - Les adverbiaux de fréquence : tous les jours
 - Les adverbiaux itératifs : plusieurs fois
 - Les adverbiaux de quantification : toujours
 - Les adverbiaux de présuppositions : encore
 - Adverbes simples ou composés : demain, plus tard
 - Syntagmes prépositionnels : en mars
 - Formes nominales : le premier jour, mardi dernier

2- Les connecteurs temporels

Les connecteurs temporels sont des éléments qui possèdent des caractéristiques d'adverbes temporels et de connecteurs de discours. on distingue trois catégories d'adverbes qui peuvent être considérées comme des connecteurs, qui sont [24] :

- Les adverbes relationnels, dont le rôle est de lier deux événements, et qui, véhiculent la temporalité ou l'ordre chronologique existant entre deux événements.
- Les adverbes de référence anaphorique, dont le rôle est de faire entrer la temporalité concernant une situation connue par rapport au moment de l'énonciation.
- Les adverbes aspectu-temporels, qui font référence à la modalité, tout en prenant une dimension temporelle.

3- Le temps

Il y a plusieurs types ou niveaux de temps, qui est le temps physique « un continu uniforme, infini, linéaire. Il a pour corrélat dans l'homme une durée infiniment variable que chaque individu mesure au gré de ses émotions et au rythme de sa vie intérieure », chronique qui est « le temps des événements, qui englobe aussi notre propre vie en tant que suite d'évènements » et grammatical. **4- L'aspect**

Il fait référence à la structure temporelle des actions ou des situations décrites par les phrases. C'est une interprétation des temps grammaticaux, il y a deux notions d'as-

pects[17], l'aspect lexical qui exprime le procès tel qu'il est conçu et l'aspect grammatical tel qu'il est montré. Le temps et l'aspect expriment tous les deux la temporalité.

5- La structure du discours

La structure de discours exprime elle aussi la temporalité, elle concerne l'ordre d'apparition des événements qui composent le texte, cet ordre permet une interprétation dans le domaine temporel. Ces événements sont généralement liés par marqueurs.

6- Les cadres de discours

Les cadres de discours correspondent à plusieurs types d'univers : spatiaux, de connaissances, de représentation, dénonciation et temporels. Ce dernier type structure et organise les expressions temporelles, en effet les introduire en début de phrase, permet une structuration du texte[17].

Représentation d'informations temporelles

La communauté Intelligence Artificiel IA distingue deux catégories de formalismes de représentation par rapport au type de raisonnement envisagé : - Les formalismes algébriques[17], telles que les algèbres d'intervalle. L'algèbre des intervalles s'adapte à la représentation des phénomènes qui se répètent souvent. Cette algèbre s'adapte à la représentation des phénomènes les plus fréquents. Elle met en évidence les relations qui existent entre ces intervalles sous forme de disjonction des relations de bases[25]. Cette algèbre a été étendue en considérant d'autres types d'intervalles qui expriment des actions sporadiques, et d'autres qui expriment des actions répétées. Les Relations temporelles sont explicitement marquées par des prépositions temporelles (par exemple avant, sur ou par). Les sept relations temporelles suivantes : before, after, incl, at, starts, finishes, excl. La préposition on comme on Friday, désigne la relation incl rapport, alors que la préposition by comme by Friday est représentée comme une relation finishes.

- Représentations proposées par la linguistique informatique[17], dans laquelle l'extraction et l'annotation d'informations temporelles peuvent être représentées par des formalismes proposés par des linguistes informaticiens. Parmi les langages d'annotation pour la normalisation des informations temporelles, nous citons le Time ML qui est utilisé pour rendre le raisonnement et l'inférence sur le temps plus facile. Ce schéma prend en compte l'annotation des événements, des expressions temporelles, il met en évidence les relations entre les événements et les expressions[23].

2.4 Les classes d'information temporelle [26]

Il existe trois classes d'informations temporelles, celles qui dénotent des expressions temporelles comme les dates et les durées, celles qui indiquent les événements et celles qui montrent les relations temporelles :

2.4.1 Les expressions temporelles

Les expressions temporelles contenues dans un texte se présentent souvent sous forme d'expressions calendaires, car elles sont ancrées au niveau du calendrier, ce qui conduit à les détecter ensuite leur associer une valeur selon la ligne de temps, ce qui fait émerger des systèmes sémantiques, qui sont utilisés dans le résumé multi documents, pour gérer l'ordonnancement temporel des données extraites afin de faciliter la compréhension. Elles sont définies comme étant un élément dans un langage qui lexicalise le concept du temps en matière de reconnaissance et généralement unités temporelles quantifiables. Les expressions temporelles se distinguent par rapport au moment de l'énonciation comme demain, par rapport à un moment de référence comme 5 jours avant, par une date relative comme le 5 juillet, et aussi plus souvent.

Il existe plusieurs catégories d'expressions temporelles qui sont les dates, les heures, les durées, les fréquences, des quantifications temporelles et les expressions non marquables.

1. **Dates** : elles font référence au calendrier , nous pouvons citer les exemples suivants :

- في السابع عشر من رمضان في العام الثاني من الهجرة, الجمعة ١٥ مارس ٢٠١٩ sont des dates calendaires complètes.
- الثلاثاء est une date partielle
- الفصل الأول، السداسي الثاني sont des sous-intervalles calendaires.
- الثلاثاء القادم، العام التالي sont des expressions déictiques.
- بعد غد، غدا sont des expressions anaphoriques.

2. **Heures** : subdivision particulière de la journée, par exemple :

- ساعة، ٥١، ٢٠ دقيقة، ٣٠ يوما٥١ sont des temps alphanumériques.
- الصباح المساء الفجر العصر العشاء sont des parties de la journée.
- هذه الليلة est une expression déictique.

2.4.2 Les expressions absolues

Elles correspondent à des dates explicites ou concrètes, elles désignent clairement une zone du calendrier, leur identification est simple contrairement à leur délimitation. L'interprétation des expressions absolues se fait généralement en les positionnant sur un calendrier. Les exemples suivants illustrent ce concept :

- في الثاني من مارس، المكتب سيكون مفتوحا
- المكتب سيكون مفتوحا في مارس

Ces deux exemples font référence à un jour précis et à un mois entier, qui peuvent être placés sur un calendrier.

2.4.3 Les expressions relatives

Une expression relative est normalisée en calculant la valeur effective de sa référence calendaire, par exemple, normaliser une expression telle qu'aujourd'hui consiste à lui attribuer une valeur calendaire, par exemple le 16 janvier 2015. Elles sont dites aussi des expressions déictiques, qui représentent des unités linguistiques. Nous prenons les deux exemples suivants pour illustrer les expressions temporelles déictiques :

- غدا، المكتب سيكون مفتوحا
- المكتب سيغلق في ٠٣٠

L'interprétation des deux expressions nécessite de savoir le moment de l'énonciation, pour le premier exemple, nous devons connaître la date d'aujourd'hui, et pour le deuxième le mois et l'année en cours.

2.4.4 Les itératifs

Les itératifs sont considérées comme des fréquences telles que "Une fois par semaine", un autre point de vue les considère comme un ensemble d'expressions ayant un ancrage multiple sur le calendrier. Nous observons quatre classes d'itérateurs :

- Les itérateurs calendaires : كل يوم أحد
- Les itérateurs fréquentiels : بعض
- Les itérateurs événementiels : عندما يصل عمر الى العمل، يقوم أولاً بركن السيارة

2.4.5 Les événements

Les événements véhiculés par des verbes sont plus faciles à détecter et à être relié aux informations temporelles, alors que les événements véhiculés par des noms indiquent son importance. Les événements peuvent être :

- Duratifs ou ponctuels
- Accomplis ou en cours
- Factuels ou non

Ils marquent ainsi une influence sur l'ordre des événements. Les événements véhiculés par des noms peuvent être classés en trois groupes :

- Un verbe peut indiquer l'action de l'évènement, ce qui est désigné par des noms déverbaux.
- D'autres événements sont construits à partir des noms qui ne sont pas déverbaux mais ambigus.
- Un autre moyen d'exprimer les événements est les syntagmes nominaux, qui n'ont pas de valeurs événementielles, mais font référence à un événement qui mentionné par un lieu ou une date.

Les événements peuvent être duratifs ou ponctuels, accomplis ou en cours, factuels ou non (ou peut-être). Time ML estime l'annotation des verbes conjugués, des adjectifs et des noms concernant des événements ou des états, en leur attribuant des propriétés constitue l'évènement.

2.4.6 La circonstance du temps

Un nom qui vient indiquer le moment où l'acte s'est produit par exemples :

- جاء محمد من السفر ليلاً
- استيقظت اليوم باكراً
- عاد الطالب من المدرسة ظهراً
- حفظتُ الدرس مساءً اليوم

D'après les exemples précédents, nous trouvons que les noms (ليلاً، باكراً، ظهراً، مساءً) sont en état de sédition et indiquent l'heure à laquelle l'acte a eu lieu, si nous demandions quand Mahomet est arrivé? La réponse ليلاً et si nous demandions quand nous sommes réveillés? La réponse sera باكراً, et si nous demandons quand l'étudiant est revenu? La réponse serait ظهراً, et si nous demandions quand la leçon était sauvegardée? La réponse sera مساءً. Donc les circonstances du temps ont aidé à déterminer le moment où les actes se sont produits et ont montré leur temps.

2.5 Les différentes formules de l'expression temporelle arabe

La vision et l'évaluation du temps ont bien changé depuis l'aube de l'humanité. L'expression du temps, elle, varie selon la langue parlée : L'expérience humaine du temps se manifeste à travers la langue. Or la langue n'est pas un calque de la réalité; chaque langue offre sa propre construction du réel et élabore son système temporel, qui n'est ni le temps du monde (temps physique) ni celui du calendrier (temps chronique)[27]. Donc, suivant la langue et la vision du monde qu'elle fonde, la façon d'exprimer le temps change. Nous verrons dans ce chapitre que l'expression de la temporalité est différente en arabe. Puisque notre outil a été conçu pour la langue arabe, nous allons tenter de décrire plus précisément la formulation du temps dans la langue arabe. L'expression temporelle arabe comprend les sections suivantes : Le jour, le mois et l'année

2.5.1 Le jour

Se compose du nom du jour de la semaine :

Exemple : الجمعة، يوم الأربعاء

Il comprend aussi le numéro de jour dans le mois. Ce tableau 2.1 résume les différentes formules de jours

Les composants du jour	les jours de la semaine	السبت، الاحد، الاثنين، الثلاثاء الأربعاء، الخميس، الجمعة
	les numéros des jours	1-31
	les numéros ordinaux des jours	الأول، الثاني، الثالث، الرابع، الخامس، السادس، السابع، الثامن، التاسع، العاشر، الحادي عشر
Les Préfixes de jours	يوم، ليلة	ليلة الجمعة، يوم الجمعة يوم الحادي والعشرون
Les suffixes de jours	On ajoute la préposition «من» après le numéro ordinal s'il été écrit en lettres	السابع من مارس ال ١٢ من رجب
Les adjectifs des jours	الجارى، الفائت، الماضى القادم، الحالى، المنصرم	السبت الفائت الجمعة القادم

TABLE 2.1 – Les différentes formules du jours

2.5.2 Le mois

il se compose du nom de mois ou de son numéro, et le mois peut être Hijri ou grégorien .quand le mois est grégorien, son nom peut varier d'une région à l'autre ou être écrit en deux mots synonymes : Par exemple :

- شهر رمضان
- رمضان ١٩٣٧

Ce tableau 2.2 résume les différentes formules de mois

Les composants du mois	Les noms des mois	Les mois hijri	محرم، صفر، ربيع الأول ... رجب، شوال، رمضان
		Les mois syriaques	كانون الثاني، شباط، آذار نيسان، أيار، حزيران،
		Les mois grégoriens	جانفي، فيفري، مارس أفريل، ماي، جوان، ...
		Numéros des mois	1-12
		Les numéros des mois ordinaux	الأول، الثاني، الثالث، الرابع، الخامس، السادس، السابع، التاسع، العاشر، ...
Préfixes des mois	شهر أو الشهر	الشهر السادس، شهر مارس	
Les suffixes des mois	On ajoute la préposition «من» après le numéro ordinal s'il été écrit en lettres	الشهر السادس من سنة ٢٠١٥ أو شهر ٦ من سنة ٢٠١٦	

TABLE 2.2 – Les différents formules du mois

2.5.3 L'année

Se compose du numéro de l'année où précédés de mot : عام أو سنة :

Exemple : العام الفائت، العام المنصرم،

Ce tableau 2.3 résume les différentes formules de l'année

Numéros d'année	1-3000	
Numéro d'année ordinal	écrit sous forme d'un terme numérique	سنة ألف وتسعمئة وأربعة وخمسين
Les préfixes d'année	عام أو سنة السنة أو العام	سنة ١٥ هجرية، عام ١٥٠
Les suffixes d'années	Nous ajoutons une distinction à la date hijri ou grégorienne en plusieurs versions	١٥ للهجرة، ٤ قبل الهجرة، ١٥ قبل الميلاد، ١٥ قم ٤ قه ٥١ م، ١٥ ميلادي
Les adjectifs des années	الجاري، الفائت، الماضي القادم، التالي، المقبل، الماضية، القادمة، الجارية	أوائل العام الماضي السنة الماضية، العام القادم

TABLE 2.3 – Les différents formules d'année

2.6 Les propriétés spécifiques des expressions temporelles en langue arabe

La langue arabe est plus riche que les langues latines en raison de différences morphologiques des entités liées aux expressions temporelles par exemple :

— آخِر، آخِر، أو آخِر، أخير —

— بدء، بدأ، بدئى، ابتداء

Considérez également le mot الاثنين ce qui signifie lundi et le numéro deux. Et aussi l'utilisation des temps de prière pour exprimer le temps par exemple :

— استيقظت قبل صلاة الفجر

— سندهب الى العمل بعد صلاة الظهر

— تناولنا طعام العشاء بعد المغرب

Le mot الثانية ce qui signifie seconde et le numéro ordinal. Exemple :

• سقط المتسابق في الثانية الأولى من السباق

• انعقد مؤتمر الصومام في السنة الثانية بعد اندلاع الثورة

2.7 Les difficultés pour la RET arabe

La reconnaissance des entités temporelles arabes rencontre des difficultés qui rendent cette tâche plus difficile par rapport aux langues indo-européennes. Et ces difficultés sont à cause des particularités de cette langue. Car c'est une langue fortement fléchie, avec une morphologie riche et une syntaxe complexe. On peut citer quelques difficultés :

2.7.1 Agglutination

La nature agglutinative de l'arabe mène à des motifs différents qui créent des variations lexicales. Chaque mot peut consister en un ou plusieurs proclitiques/préfixes, une base ou une racine, et un ou plusieurs suffixes/enclitiques dans des combinaisons différentes, aboutissant à une morphologie très systématique mais compliquée. Dans d'autres langues comme le français ou l'anglais les clitiques seraient traités comme des mots séparés qui n'agglutinent pas aux mots. L'arabe a un ensemble de clitiques qui sont attachés à une EN, y compris :

— des conjonctions telles que : و ، ف .

- des prépositions telles que ب ، ك ، ل.
- une combinaison des deux telles que و ، ل.

2.7.2 Les étiquettes des mois

Les mois dans la langue arabe ont plusieurs étiquettes ce qui pose un problème lors de la reconnaissance des ET. Par exemple : le mois janvier/ يناير، محرم، كانون الثاني/جانفي nommé

2.7.3 Analyse morphologique

La morphologie considère la composition d'un mot en plusieurs morphèmes. Un morphème est une tige ou un affixe, un affixe est un suffixe ou un préfixe ou un infixé. Les préfixes et les suffixes sont attachés au début ou fin d'un mot respectivement, tandis que les infixes introduisent des changements dans la tige. La figure (2.1) représente un exemple de l'analyse morphologie

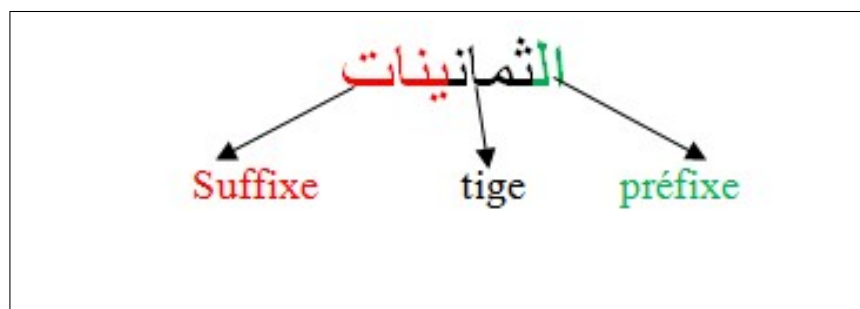


FIGURE 2.1 – La morphologie d'un mot arabe

2.8 Conclusion

Dans ce chapitre, on a essayé d'exposer le rôle des entités temporelles dans les applications TAL. Nous avons également présenté un état de l'art sur Les aspects des expressions temporelle arabes, en mettant l'accent sur les différentes classes de l'information temporelle en général. En suite nous avons présenté les différentes formules d'une expression temporelle dans la langue arabe, les propriétés spécifique des expressions temporelles. En

fin on conclut le chapitre avec la problématique de reconnaissance des entités temporelles arabes.

Reconnaissance et extraction des expressions temporelles

3.1 Introduction

Le temps est l'un des aspects les plus importants de l'information, car elle est généralement centrée sur les événements et organisée par ordre chronologique. Pour comprendre les progrès d'événements, les individus doivent capturer des informations temporelles les concernant, y compris des expressions temporelles et des relations temporelles. Pour cette raison, l'extraction d'informations temporelles en tant que tâche fondamentale du traitement du langage naturel a toujours attiré beaucoup d'attention. Un grand nombre de systèmes utilisant diverses méthodes, notamment à base de règles et d'apprentissage automatique, ont été développés pour extraire des informations temporelles sous forme de texte dans différentes langues (anglais, chinois, français, etc.), dans différents domaines. Les approches d'extraction des événements et des expressions temporelles sont vraiment rares, surtout en ce qui concerne la langue arabe. Les approches qui ont attirés notre attention sont ceux cités en introduction générale. On essaiera dans ce qui suit, de les présenter afin de choisir une approche pour notre système.

3.2 Les approches pour la Reconnaissance des Entités Temporelles RET

La reconnaissance d'informations temporelles consiste à l'analyse du texte de manière partielle, permettant ainsi une extraction d'information spécifique. Généralement la tâche

d'extraction regroupe trois sous-tâches[17] :

- La reconnaissance des entités nommées.
- La reconnaissance des relations entre les entités nommées.
- La reconnaissance d'évènements.

La tâche de reconnaissance des entités nommées est une tâche assez complexe et difficile. cependant la reconnaissance des relations qui existent entre les entités et les événements représente une tâche souvent prise en compte par les outils d'extraction, sans oublier la tâche d'extraction des événements qui peut être considérée comme un sous ensemble de relation. Ces tâches sont importantes pour le développement d'outils d'extraction, d'où l'apparition de diverses campagnes d'évaluations, qui ont fait émerger plusieurs approches qui font appel à des prétraitements.

L'extraction d'informations temporelles est une tâche qui prend en compte des caractéristiques selon le type d'information que l'on souhaite extraire. Plusieurs méthodes ont été développées sur ce sens, afin de permettre une extraction complète de l'information temporelle. Ces méthodes regroupent généralement des prétraitements qui assurent le traitement de la temporalité.

La reconnaissance des informations temporelles est un traitement permettant de parcourir un texte en détectant ces informations. Ce traitement est suivi de l'annotation qui lui attribue une valeur aux informations extraites. Ces valeurs sont données selon un système calendaire.

L'extraction d'informations temporelles considère deux tâches essentielles, la tâche de repérage des informations ainsi que la tâche d'interprétation. L'extraction d'informations temporelles a été élaborée selon plusieurs approches, qui peuvent être regroupées en deux classes : l'approche à base de règles (basée sur la linguistique) et l'approche basée sur l'apprentissage automatique[17].

3.2.1 Approche à base des règles

L'extraction est fondée sur les techniques linguistiques. Le principe de cette approche est d'établir des règles manuelles, afin d'assurer le repérage et l'extraction des informations souhaitées. Cette approche se focalise sur les systèmes de règles et les ressources linguistiques. Elle vise à atteindre un bon niveau de précision, tant dit que le rappel est lié aux ressources développées. Plusieurs auteurs se sont penchés vers cette approche dont

le but de décrire les étapes de reconnaissance et d'interprétation. Des automates à état fini ont été utilisés pour décrire ces étapes qui prennent en compte des types d'informations bien précises comme les dates et les adverbes. Cette approche fondée sur les automates à état fini représente une base pour d'autres travaux, comme pour Hagege et Tannier. Dans cette approche, l'étape d'interprétation n'est pas aussi précise à cause des automates, car elle nécessite des ressources indispensables à cette interprétation, c'est pourquoi des règles sont manipulées pour la reconnaissance, ainsi que l'interprétation qui peut être mise en œuvre de plusieurs façons[29].

3.2.2 Approche d'apprentissage

Cette approche considère un texte annoté, et utilise des caractéristiques plus ou moins linguistiques. Les informations extraites par cette méthode doivent être mises en relation. Elle permet de mettre en évidence l'étape de reconnaissance d'informations temporelles, contrairement à l'interprétation de celles-ci, qui s'annonce moins efficace, et nécessite des règles pour une meilleure interprétation. En effet, à une expression temporelle peut correspondre plusieurs classes selon une classification, ce qui rend son interprétation difficile. L'extraction d'informations temporelles est basée aussi sur les méthodes d'apprentissage, ces méthodes sont aussi utilisées pour la détection des informations temporelles, en se focalisant sur des caractéristiques bien déterminées de ces informations, ses caractéristiques varient selon les divers systèmes existants. Sur ceux, nous pouvons avoir une liste de mots comme caractéristique, cette liste comporte les mots qui sont fréquemment exploités par les expressions temporelles. Son utilisation a permis d'augmenter le rappel[30].

3.3 Système à base d'apprentissage

3.3.1 ZamAn and Raqm : Extraction des expressions temporelles et numériques en arabe

ZamAn est le premier système robuste et précis de reconnaissance et d'extraction des expressions temporelles de l'Arabe. ZamAn, une méthode d'apprentissage automatique développée pour étiqueter les temporels arabes, en traitant le dashtag -TMP fonctionnel utilisé dans la banque d'arbres arabe pour marquer un modificateur temporel qui

représente une référence à un point dans le temps.

Raqm, une méthode d'apprentissage automatique appliquée à l'identification de différentes formes d'expressions numériques afin de les normaliser en chiffres.

L'évaluation du système ZamAn & Raqm obtient des résultats de pointe de la mesure F1 de 88,5% (resp. 96%) pour le bracketing et de 73,1% (resp. 94,4%) pour la détection [31].

3.4 Systèmes à base de règles

Les systèmes à base de règles sont mis en place à partir de ressources linguistiques (dictionnaires, grammaires, etc...). Le système s'appuie sur ces ressources développées à la main par un expert. Ce type de système est défini comme la manière suivante :

"Rule based or knowledge-driven methods exploits human knowledge about the contents of the text to be processed in addition to linguistic and lexicographic knowledge. This knowledge is encoded in the form of patterns that express rules which are used to extract desired information from the text. Information is mined from corpora by using predefined or discovered linguistic patterns, which can be either lexico-syntactic patterns or lexico-semantic patterns. While lexico-syntactic exploits lexical representation with syntactical information, lexico-semantic uses semantic or meaning of the information along with lexical representations"[32].

3.4.1 HeidelTime

HeidelTime a été créé par Jannik Strötgen et Michael Gertz dans le cadre du défi TempEval-2. Il a été conçu pour l'extraction et la normalisation des expressions temporelles en anglais dans le but d'utiliser l'information temporelle dans d'autres applications[33]. C'est un système à base de règles, défini comme un *"multilingual, domain-sensitive temporal tagger developed at the Database Systems Research Group at Heidelberg University. It extracts temporal expressions from documents and normalizes them according to the TIMEX3 annotation standard"*. HeidelTime a été élaboré en UIMA (Unstructured Information Management Architecture) 3 afin de proposer un système utilisable dans différents contextes et domaines sans (ou avec peu de) modifications. Le système est aujourd'hui développé et abouti dans quelque treize langues et des recherches sont menées afin de l'étendre automatiquement à toutes les langues du monde, Moriceau & Tannier

ont élargi le système au français[34].

3.5 Quelques systèmes d’annotation des expressions temporelles

3.5.1 Extraction d’événements dans le cadre de TempEval 2

Extraction d’événements dans le cadre de TempEval 2 utilise trois balises TimeML : `TIMEX 3`, `EVENT` et `TLINK`.

TIMEX 3 marque les expressions temporelles dans le texte et est identique à la balise `TIMEX3` dans TimeML. Les expressions temporelles peut être exprimée syntaxiquement par adverbial ou phrases prépositionnelles, comme indiqué dans la suite :

- a. jeudi
- b. 15novembre 2004
- c. jeudi soirée
- d. plus tard cet après-midi

Les deux attributs principaux de la balise `TIMEX 3` sont `TYPE` et `VAL`, tous deux montrés dans l’exemple :

```
22 novembre 2004 type = "DATE" val = "2004-11-22"
```

Pour TempEval-2, nous distinguons quatre valeurs temporelles types : `HEURE` (à 14h45), `DATE` (27 janvier, 1920, hier), `DURATION` (deux semaines) et `SET` (tous les lundi matin). Il y a eu à ce jour trois éditions des campagnes d’évaluations TempEval (conjointes aux conférences SemEval3).

TempEval-14 : s’intéressait à la détection de relations entre entités préalablement annotées. Cette première édition ne proposait que des textes en anglais[35].

TempEval-25 : s’est concentré sur la détection d’événements, d’expressions temporelles et de relations temporelles. Cette campagne est multilingue (comprenant notamment l’anglais, le français, l’espagnol) et les tâches ont été plus précisément définies par rapport à TempEval-1[36].

TempEval-36 : s’inscrit dans la continuité des éditions précédentes. Il s’agit là aussi d’évaluer l’extraction d’événements et de relations temporelles mais seuls l’anglais et l’espagnol sont proposés. Par ailleurs, un des focus de ce nouveau challenge était d’évaluer

l'impact de l'ajout de données d'entraînement annotées automatiquement aux données annotées manuellement[37].

Plusieurs systèmes ont été proposés dans le cadre de la compétition TempEval-2, travaillant pour la plupart sur l'anglais. Le système TIPSEM repose sur un apprentissage artificiel par CRF et des attributs de différentes natures : lemmes, parties-du-discours, informations syntaxiques obtenues automatiquement ou par des règles écrites à la main, informations sémantiques tirées de WordNet. Ce système a obtenu les meilleurs résultats à TempEval-2 et a été utilisé comme référence pour TempEval-3[17].

3.5.2 TimeML - un langage de balisage

TimeML est un langage de balisage conçu pour la reconnaissance et la normalisation des expressions temporelles et des événements présents dans un texte. Ce langage propose un guide d'annotation initialement prévu pour un texte libre en anglais [38]. Cependant, TimeML a par la suite été développé dans d'autres langues, dont le français . TimeML est spécialisé dans les expressions temporelles, les événements et les relations qu'ils entretiennent. Il a été développé dans le but d'annoter et de standardiser ces éléments dans les textes écrits en langue naturelle. L'annotation est prévue comme la figure(3.1) :

<EVENT>	permet de marquer les événements, ponctuels ou ayant une certaine durée, qui se produisent ou qui se sont produits
<TIMEX3>	permet de marquer les heures, les dates, les intervalles et les durées. Ils peuvent être complètement spécifiés (ex : « le 3 juin 1950 ») ou non (ex : « le 3 juin »)
<SIGNAL>	permet d'annoter les mots indiquant comment les informations temporelles sont reliées (ex : avant, après, durant, ...)
<TLINK>	Temporal link : permet de décrire les relations temporelles entre les balises <EVENT>, entre les balises <TIMEX3>, ou entre les balises <EVENT> et <TIMEX3>
<ALINK>	Aspectual link : permet de décrire les relations entre les balises <EVENT> correspondant à un événement aspectuel (commencer, finir, ...) et les balises <EVENT> correspondant à l'événement argument
<SLINK>	Subordination link : permet de décrire le contexte introduisant les relations entre deux balises <EVENT>

FIGURE 3.1 – Les balises TimeML utilisées en français[39].

TimeML permet à l'information temporelle d'être exprimée de manière structurée. Ce langage a été développé dans le but de répondre à quatre attentes[1] :

- 1 Identifier un événement et déterminer son espace temps.
- 2 Ordonner les événements les uns par rapport aux autres.
- 3 Permettre de contextualiser les expressions temporelles non spécifiées (du type « la semaine passée »).
- 4 permettre de calculer la durée d'un événement et les répercussions qu'a cet événement.

Par exemple :

L'expression « avant septembre 2017 » serait analysée comme suit :

<TIMEX3> <SIGNAL> avant </SIGNAL> septembre 2017 </TIMEX3>

Une extension a été apportée au langage TimeML, l'ISO-TimeML. C'est un langage de spécification formel utilisé pour marquer l'information temporelle dans le langage naturel. La norme ISO-TimeML ne diffère pas tellement du langage TimeML, mais elle a été développée dans un esprit multilingue et ne contient pas certaines redondances de TimeML (par exemple, la balise makeinstance, utilisée pour les différentes instances d'un événement, est supprimée)[39].

Un autre changement important entre les deux langages est le passage d'une annotation par ligne (TimeML) à une annotation "en confrontation" (ISO-TimeML). La raison première d'étendre le TimeML à l'ISO est la production de ressources annotées durables du langage pour des applications pratiques. Trois éléments doivent être précisés : la position de l'intervalle par rapport aux autres intervalles (*order*), la taille de l'intervalle (*measure*) et le nombre d'intervalles (*quantity*). La représentation temporelle dans ce type de corpus est cependant difficile, à cause de la diversité des expressions de temps, la complexité à déterminer les relations temporelles entre les événements, et la difficulté à gérer les granularités temporelles d'un événement. De manière plus générale, elle est également ardue dans le traitement de la langue naturelle[38].

3.6 Conclusion

Dans ce chapitre on a essayé d'exposer les différentes formules de l'expression temporelle, Le rôle de ces entités dans les applications TAL et la problématique de la reconnaissance des entités temporelles arabes et surtout les difficultés rencontrées, les approches utilisées et les systèmes réalisés dans ce domaine. Au terme de ce chapitre, on conclut que la reconnaissance des entités temporelles arabes est encore en stade de la recherche scientifique. Après l'étude des approches de reconnaissance des ET, On choisira donc l'approche à base de règle avec un choix d'indice linguistique. Cette dernière sera bien détaillée dans le chapitre suivant.

Conception

4.1 Introduction

L'arabe est une langue très riche. Cependant, cette richesse nécessite une manipulation particulière qui rend les systèmes de traitement de la langue naturelle, conçus pour d'autres langues ne sont pas en mesure de le gérer. Dans ce chapitre, nous présentons les étapes du développement de notre outil, l'approche et les méthodes utilisées pour mettre en œuvre la détection et l'extraction des expressions temporelles arabe. Cela veut dire parcourir le texte brut et appliquer toutes les règles pour :

- prétraiter le texte
- détecter les entités temporelles
- extraire les expressions temporelles

Ce travail a requis la conception et la mise en œuvre de plusieurs programmes, ressources linguistiques et transducteurs. La figure (4.1) montre les différents niveaux nécessaires à la réalisation de l'approche. Comme il est montré dans l'approche, le traitement se fait en pipeline comme c'est le cas pour toutes les applications TAL. A chaque niveau ou étape dans un niveau on utilise des programmes et des ressources linguistiques qui sont majoritairement développés dans le cadre de ce travail. Les différentes étapes seront expliquées au long des sections qui vont suivre.

4.2 L'approche proposée

En général, pour la langue arabe, les approches servant à extraire les événements et les expressions temporelles dans la littérature ne sont implémentées que dans les tweets ou

pour des petites phrases où l'on ne trouve qu'un seul évènement. Par conséquent, dans le cas des articles arabes où l'on trouve plusieurs évènements et plusieurs types d'expressions temporelles, les approches sont rares voire inexistantes.

Dans notre cas, on s'est penchés pour une méthode « à base de règles ». Ce choix s'est fait principalement par rapport à son efficacité à extraire les évènements et les expressions temporelles dans les articles arabes est généralement considéré comme déclencheur d'évènements ; En plus de la disponibilité d'outils d'indices linguistiques.

4.3 Description des données

Cette section est consacrée à la description des données utilisées en entrée et à celles des résultats produits.

4.3.1 Données en entrée

Nous avons choisi, comme mentionné précédemment, de nous concentrer sur les expressions qui exprime le temps dans la langue arabe. Donc l'entrée est un texte en arabe.

4.3.2 Données en sortie

La sortie est un texte (des phrases qui exprime des expression temporelles). L'algorithme traite l'entrée et fait des opération pour détecté et extraite les entités temporelles, puis appliquer les règles sur ces entités afin de construire le résultat final.

4.4 L'algorithme de l'application

Le tableau ci-dessous représente l'algorithme de notre application.

<p>Entrées : texte arabe</p> <p>Sortie : Expression temporelles</p> <hr/> <p>Début :</p> <ol style="list-style-type: none"> 1. Le Pré-traitement <ul style="list-style-type: none"> — L'acquisition : insérer les données — Encodage unique 2. La segmentation <ul style="list-style-type: none"> — La segmentation en mots — La segmentation des clitiques 3. Le traitement linguistique <ul style="list-style-type: none"> — Supprimer les signes diacritiques — La détection des entités temporelles arabes — La tagination 4. Application des règles <p>Fin</p>

TABLE 4.1 – L'algorithme générale de système.

4.5 Protocole de système

L'architecture globale de notre système d'extraction des expressions temporelles, est représentée dans la figure 4.1. Le processus de l'application de ce système suit une série de traitements englobe en 4 étape, chaque étape contient des processus, en commençant par un pré-traitement puis une segmentation des mots et des clitiques suivie d'un traitement qui contient deux étapes le traitement linguistique et enfin l'application des règles.

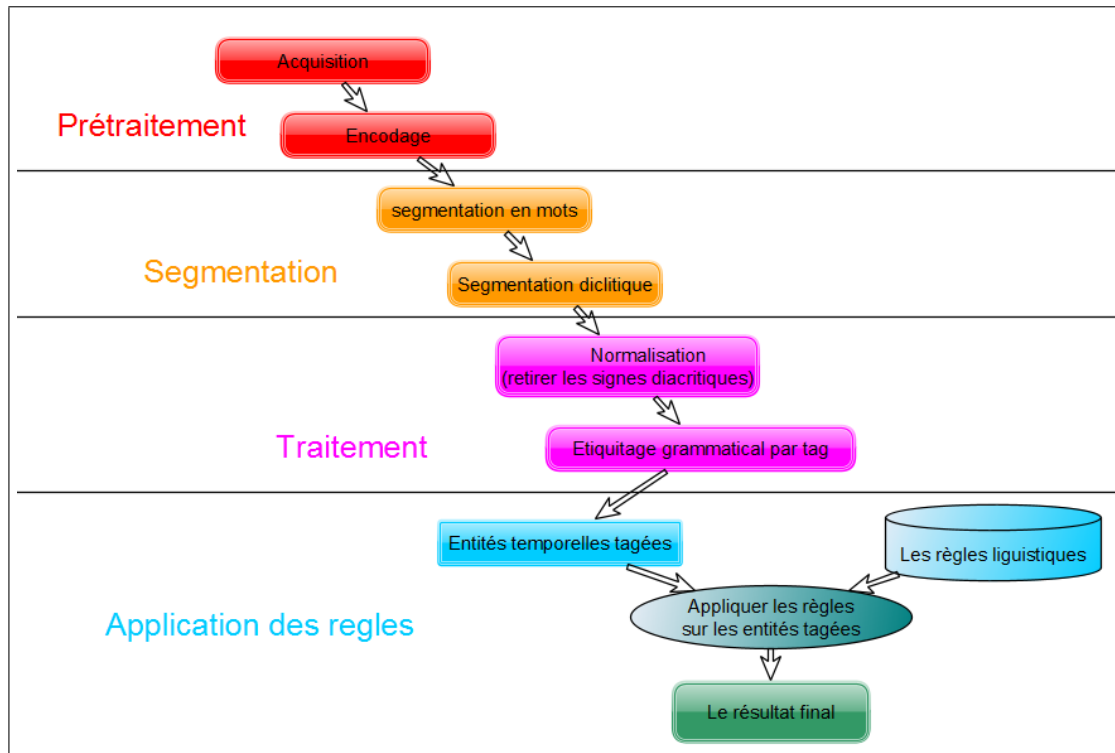


FIGURE 4.1 – Processus général de la méthodologie du système.

4.6 Description détaillée du protocole de développement

4.6.1 Le pré-traitement

Le pré-traitement (*preprocessing*) est une étape très importante qui consiste en plusieurs techniques visant à traiter les textes pour les structurer et faciliter leurs utilisations. Elles convertissent les données textuelles originales dans une structure d'exploration de données prêtes et en passant par les étapes détaillées ci-après.

Étape d'acquisition : Ce processus est chargé d'insérer des textes soit manuellement ou à travers une base de données (dataset). Notre système fonctionne uniquement avec des textes en arabe.

Encodage unique : Tous les inputs de notre système doivent être représentés dans un fichier.csv avec encodage UTF-8. Pour qu'il n'est pas de déformations au niveau des

caractères lors de sa lecture. Cette étape permet donc de convertir un article de type (.csv) fichier excel a un fichier texte (.txt).

4.6.2 La segmentation

La segmentation en mots : Cette étape permet de découper les termes du texte en segments, de sorte qu'ils puissent être introduits dans un capteur morphologique ou dans un étiqueteur de position pour un traitement ultérieur.

La segmentation des clitiques : Le phénomène de l'agglutination dans la morphologie arabe augmente le degré d'ambiguïté lexicale et accroît la complexité de l'analyse d'un mot en ses composants. Comme il a été discuté dans le chapitre 3, un mot arabe est grossièrement composé de pré-base, base et post-base. Ces derniers composants eux même composés des affixes et des clitiques arabes qui peuvent s'agglutiner en plusieurs niveaux. La Figure 2 montre les composants d'un mot arabe et la Figure 3 donne un exemple de l'analyse d'un mot arabe en affixes et en clitiques. (voir Figure 4.2)

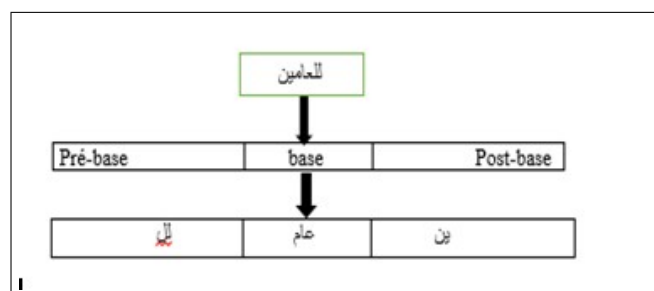


FIGURE 4.2 – Exemple d'agglutination.

4.6.3 Le traitement linguistique

Supprimer les signes diacritiques : fait que les signes diacritiques sont ajoutés à un mot pour résoudre une éventuelle ambiguïté ; la suppression de ces caractères apporte un degré d'ambiguïté supplémentaire au texte. Mais pour une simplicité de traitement l'utilisateur peut choisir ou non la suppression des signes diacritiques (facultatif).

Etiquetage gramatical par TAG : Pour détecter les entités temporelles arabes on utilise un ensemble de dictionnaires, tel que le système doit vérifier l'appartenance de chaque mot dans un de ces dictionnaires.

Un tagueur des entités temporelles est un traitement qui lit un texte et donne un signe à chaque mot : NB, NI, 0.

NB : désigne le premier mot qui exprime une entité temporelle dans la phrase.

NI : on donne ce signe aux mots qui expriment des entités temporelles et qui sont classées après le NB.

0 : désigne le mot n'est pas une entité temporelle.

4.6.4 Application des règles

Pour extraire les expressions temporelles, on a implémenté grâce au logiciel Yed qui facilite l'implémentation des grammaires (ou bien des graphes) des automates (Automate d'État Fini) AEF qui traitent les différentes variétés de l'expression temporelle.

4.7 Construction des règles :

En langue arabe, les expressions temporelles sont divisées en plusieurs classes : dates, heures, durées, fréquences, quantification temporelle et expressions non marquables (voir chapitre 02). Ces classes peuvent être absolues ou relatives. La figure 4.3 ci-dessus montre les classes et sous-classes de l'expression temporelle.

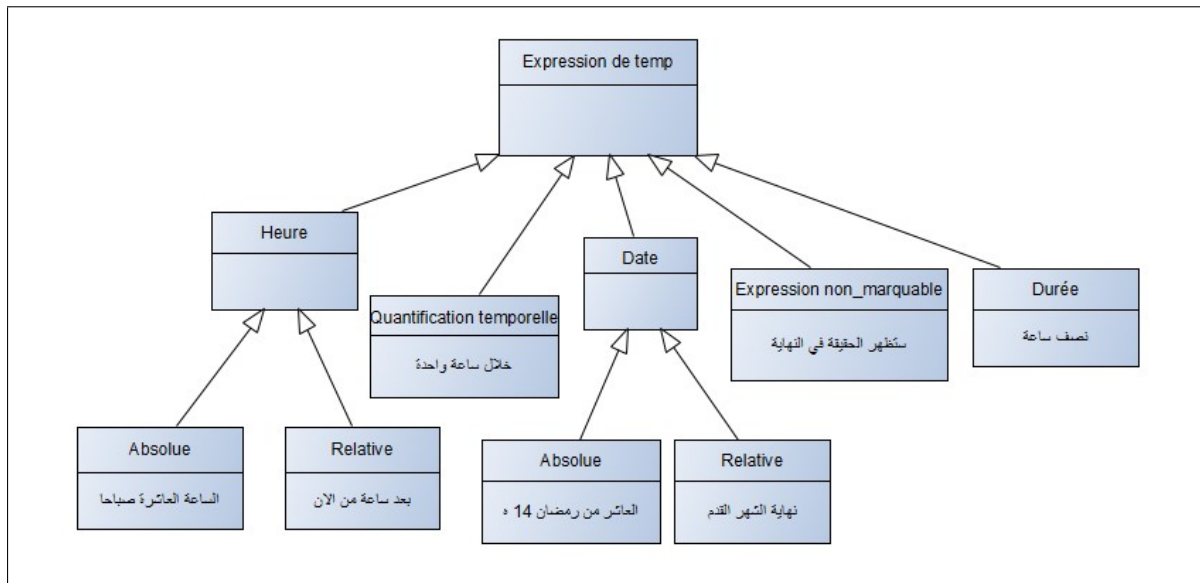


FIGURE 4.3 – Les classes et sous classes de l'expression temporelle.

Certain des entités utilisées dans les règles de conception sont les suivantes : Les numéros ordinaux, Les noms du mois, les noms de jour, les adjectives, les calendriers. Nous avons établi un ensemble de dictionnaires (voir Annex A).

Pour extraire les différents types d'ET, on a construit des AEF qui traitent les différentes variétés de chaque type d'Et en la langue arabe.

4.7.1 Les dates

Les dates dans un texte arabe peuvent apparaître sous différents types, absolues et relatives. Elles peuvent également apparaître sous différentes formes en chiffres et en lettres, et les noms des mois peuvent se trouver en hijri, en grégorien ou en syriaque. Tous les mots ou groupes de mots faisant référence aux jours de la semaine.

Date absolue

Une date est considérée absolue lorsque la date du document annoté n'est pas nécessaire pour la déterminer précisément.

Annotation des dates absolues en chiffres

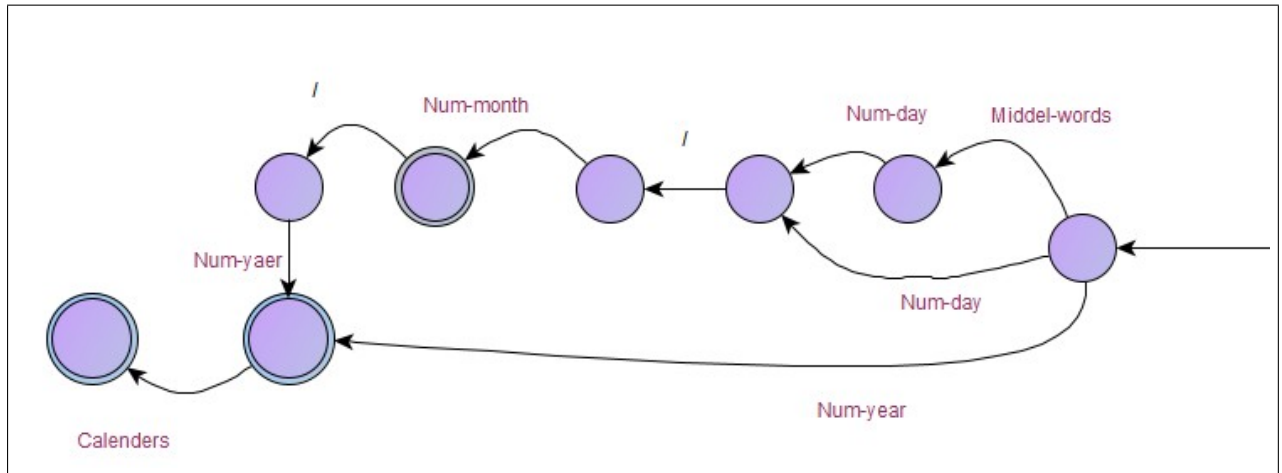


FIGURE 4.4 – AEF d’annotation des dates absolues en chiffre

Les formes reconnues par le graphe (4.4) sont :

- 05/07/1962
- الموافق ل ١٥ / ١٢ ميلادي

Annotation des dates absolues en lettres

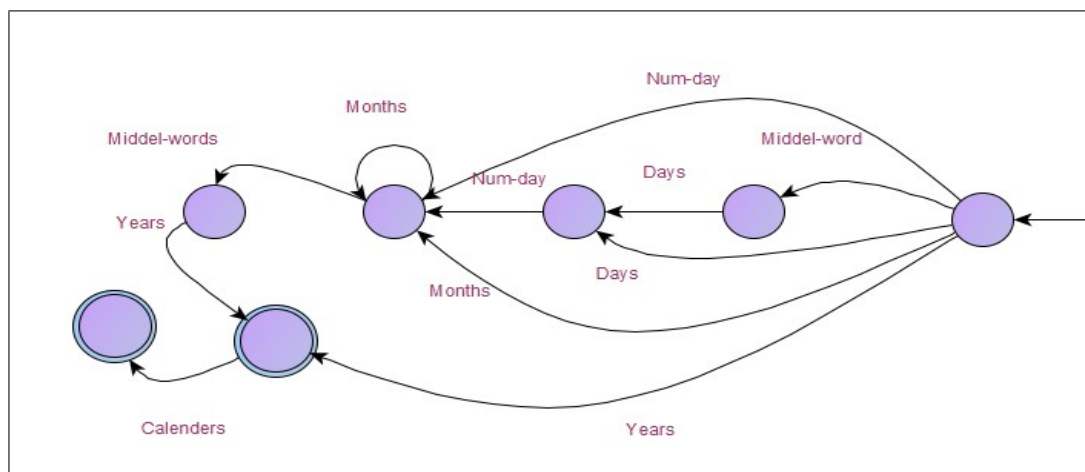


FIGURE 4.5 – AEF d’annotation des dates absolues en lettres.

Exemple d'une forme reconnue par le graphe (4.5) :

ليلة الاحد الخامس عشر من شهر نوفمبر الجاري

Annotation des dates absolues en lettres et en chiffre

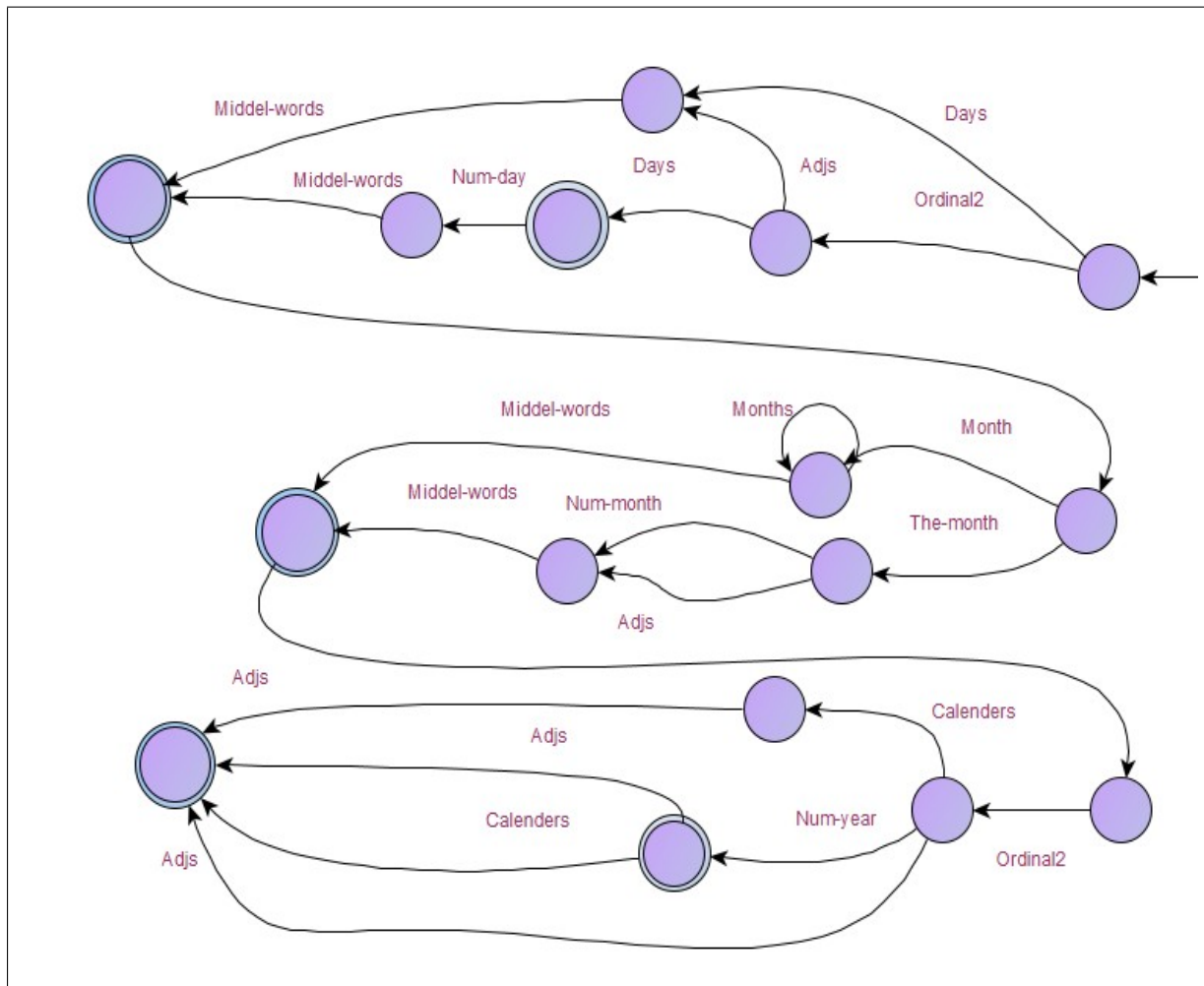


FIGURE 4.6 – AEF d'annotation des dates absolues en lettres et en chiffres

Exemple de forme reconnue par le graphe (4.6) :

— بتاريخ الاحد ١٢ نوفمبر من السنة الميلادية 1995

Annotation des dates absolues sous des formes irrégulières

On note que les dates peuvent prendre des formes non habituelles. voir la figure (4.7).

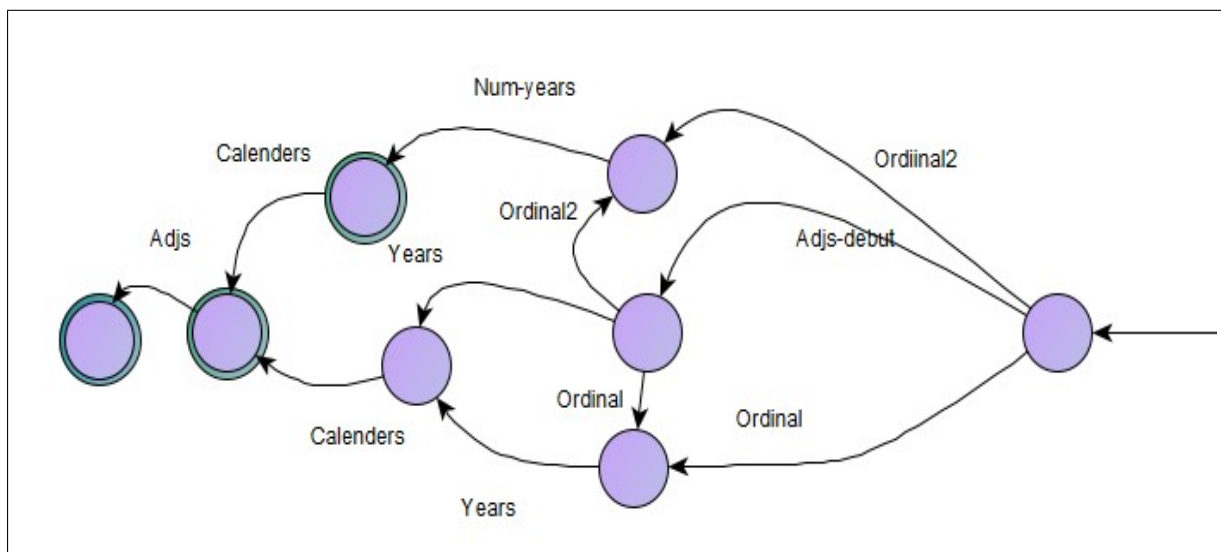


FIGURE 4.7 – AEF d'annotation des dates absolues sous des formes irrégulières

Quelques formes reconnues par ce graphe (4.7) :

- أوائل خمسينات القرن الماضي
- نهاية القرن الثاني عشر ميلادي

4.7.2 Date relative

Une date est considérée comme relative lorsque la date du document annoté est indispensable pour la déterminer précisément. Dans notre travail on s'intéresse beaucoup plus aux dates relatives car elles se trouvent dans les textes journalistiques sous formes d'expressions temporelles pures.

4.7.3 Les quantificateurs temporels :

Une période est un intervalle dont le début et la fin ne sont pas précis.

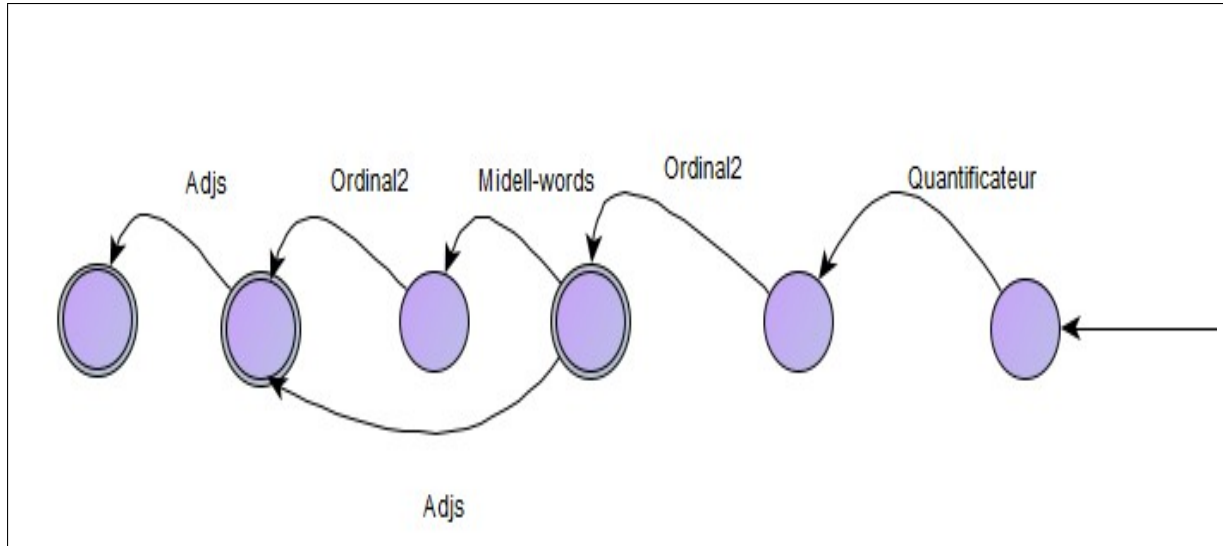


FIGURE 4.9 – AEF d'annotation des quantificateurs temporels.

Quelques formes reconnues par ce graphe (4.9) :

بضعة أيام في الشهر

4.7.4 Les heures

Les horaires sont construits par concaténation d'un nombre pour l'heure, un séparateur, et, éventuellement, un nombre pour les minutes

L'annotations des heures relative

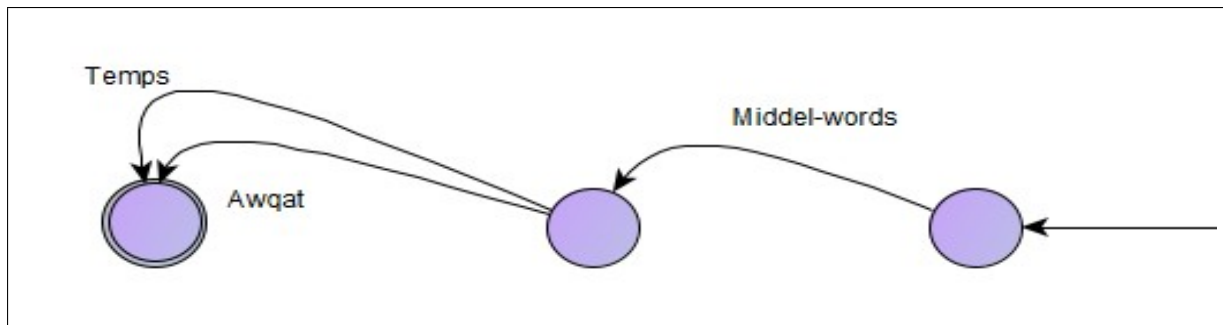


FIGURE 4.10 – AFE d'annotations des heures relative

Quelques formes reconnues par ce graphe (4.10) :

هذا المساء

منذ ساعة

L'annotations des heures absolue

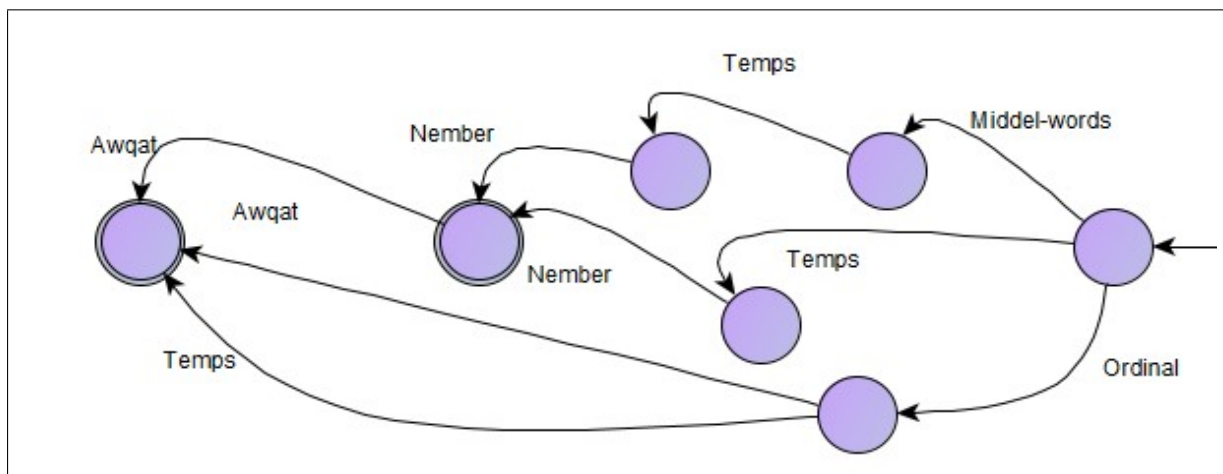


FIGURE 4.11 – AEF d'annotations des heures absolue

Quelques formes reconnues par ce graphe (4.10) :

الساعة العاشرة صباحا

بعد الدقيقة الخامسة

Tout les règles représentés programme a été implémenté (avec une autre interface présentée dans le chapitre qui suit) afin d'extraire toutes les expressions temporelles d'un texte donné.

4.8 Construction du corpus :

Face a la croissance de l'information, il existe une demande croissante d'applications rebustes a utiliser pour la recherche des informations dans des textes et pour créer des modèles permettant d'extraire des connaissances sémantiques de textes. à partir de ca explorer les expressions temporelles et historiques particulier dans un texte donné, un texte historique ou une nouvelle, afin de mieux comprendre les textes et d'extraire les relations les plus importantes entre les mots et les phrases.

Le développement d'un système d'extraction d'entités temporelles nécessite, au préalable, de rassembler un nombre suffisant de textes qui serviront non seulement de corpus d'observation et d'analyse (pour construire les règles) mais également de corpus de test (extraction des entités).

Pour couvrir le domaine de recherche, il nous a fallu recueillir des exemples liés aux expressions temporelles et constituer un dataset pour la création de notre système.

La méthodologie suivie consistait donc à analyser une dataset composée de divers phrases, paragraphes qui contient des expressions temporelles. 80% sont utilisés pur l'entraînement et 20% pour le test.

4.9 Conclusion

Dans ce chapitre, on a décrit notre système, dont l'objectif était de concevoir un système capable d'extraire automatiquement les expressions temporelles à partir des textes arabes. La conception de système est divisée en 4 étapes. Cette dernière sera mise en fonction dans le chapitre qui suit.

Implémentation

5.1 Introduction

Dans ce chapitre, nous allons parler des concepts de base de tous les outils qui ont été utilisés dans notre implémentation. Nous allons également mentionner les résultats des tests qui ont été faits sur le dataset et leur évaluation. Avant d'entrer dans les détails, nous décrivons d'abord notre environnement de travail.

5.2 Environnement de Travail

D'abord, Dans cette section nous allons mentionner et décrire les outils et l'environnement utilisés dans notre implémentation :

5.2.1 Environnement matériel

Quand il s'agit de l'exigence de matériel, vous n'avez pas vraiment besoin de quelque chose de fantaisie. N'importe quel ordinateur qui marche correctement est suffisant qui marche sous n'importe quel système d'exploitation (Windows, Linux, Mac Os), vu que le langage utilisé est un langage interprété. Pour notre cas, nous avons utilisé un ordinateur portable php équipé d'un processeur I3.

5.2.2 Environnement logiciel

Nous avons utilisé le langage de programmation Python la version 3.6.4. Python est un langage de script de haut niveau, interprété, interactif et orienté objet. Python est conçu pour être très lisible. Il utilise fréquemment des mots-clés anglais alors que d'autres

langues utilisent la ponctuation, et il a moins de constructions syntaxiques que d'autres langues[40].

5.3 Déroulement

On présente dans cette section le déroulement de travail de notre système d'extraction, on affichera les résultats d'extraction à partir d'une partie de notre dataset dans le tableau(5.1) .

Text	Résultat1	Résultat2	Résultat3	Le résultat sélectionné par le système
غزوة بدر أو بدر الكبرى أو بدر بدر القتال أو يوم الفرقان هي غزوة حدثت في السابع عشر من رمضان بين المسلمين وقبيلة قريش وحلفائها	السابع عشر من رمضان في العام الثاني من الهجرة			السابع عشر من رمضان في العام الثاني من الهجرة
غزا النبي صلى الله عليه وسلم وسلم وصحابته رضي الله عنهم غزوات كثيرة، وقد اختلف العلماء والمؤرخون في عدد عدد هذه الغزوات؛ ف قيل خمسة وعشرون، وقيل سبع وعشرون، وقيل تسع و عشرون غزوة، وكانت كلها بعد الهجرة، واستمرت لعشرة أعوام	لعشرة أعوام	بعد الهجرة		خمسة وعشرون سبع وعشرون تسع وعشرون الهجرة لعشرة أعوام

اندلعت الحرب العالمية الأولى في يوم الثامن والعشرون من حزيران يونيو عام / 1914 واتتهت بعد استمرار أربعة أعوام متتالية، أي في يوم الحادي عشر من تشرين الثاني / نوفمبر عام 1918	يوم الثامن والعشرون من حزيران يونيو عام يونيو ١٩١٤	أربعة أعوام متتالية	يوم الحادي عشر من تشرين الثاني / نوفمبر	الثامن والعشرون من حزيران عام يونيو، ١٩١٤ متتالية. الحادي عشر من تشرين الثاني / نوفمبر عام ١٩١٨ أربعة أعوام
حدثت غزوة حنين في السابع من شوال من العام الثامن للهجرة	السابع من شوال من العام الثامن			السابع من شوال من العام الثامن للهجرة
وفي مساء السابع عشر من تشرين الأول أو كما يُسمى بالأحد الدامي تم فتح النيران على المتظاهرين	مساء السابع عشر من تشرين الأول			وفي مساء السابع عشر من تشرين الأول، بالأحد

TABLE 5.1 – Les résultats de déroulement de quelques textes

Ce tableau 5.1 représente quelques textes en arabe qui contient des expressions temporelles, la deuxième, la troisième et la quatrième colonne représente le résultat sélectionné, la cinquième colonne représente le résultat sélectionné par notre système.

On remarque que la plupart des résultats sont sélectionnés correctement, quelques résultats sont relativement corrects comme le résultat de la dernier exemple, et rarement le résultat sélectionné est incorrect 100%.

5.4 Evaluation du système

Pour évaluer la performance de notre système, nous avons fait une évaluation semi-automatique de ce dernier sur un ensemble de paragraphes qui contient les différents types

différentes expression temporelle.

L'évaluation sert à comparer les expressions temporelles sélectionnées par le système et celles qui étaient définies préalablement. En utilisant les métriques Rappel et Précision afin de calculer F- mesure.

Les mesures Rappel, Précision et F-mesure sont calculées par les équations suivantes :

$$-Rappel = Corrects / (touteslesinformations)$$

$$-Précision = Corrects / (Corrects + Incorrects)$$

$$Fmesure = 2 * (Rappel * Précision) / (Rappel + Précision)$$

Corrects : Nombre d'expressions temporelles sélectionnés correctement par le système.

Incorrects : Nombre d'expressions temporelles sélectionnés incorrectement par le système.

Toutes les informations : nombre d'expressions temporelles sélectionnées préalablement.

	Précision	Rappel	F-mesure
Les données d'entraînement	0,88	0,90	0,89
Les données de test	0,82	0,86	0,84

TABLE 5.2 – Les résultats de la mesure

Ce tableau 5.2 montre l'efficacité de notre système de reconnaissance. Par ailleurs, son évaluation est généralement subjective. Car le style de rédaction des textes diffère d'un auteur à un autre, et les formes ainsi que les types des expressions temporelles ne peuvent pas toutes être implémentées, à cause de la richesse de la langue arabe.

5.5 Discussion

Notre système d'extraction d'expressions temporelles est fait après une analyse linguistique et l'utilisation des "dictionnaires" pour trouver les informations correspondantes, il y a peu de bruit, mais toutes les informations ne sont pas extraites. Nous pensons que ces résultats peuvent être améliorés :

- Le rajout de termes utilisés dans les dictionnaires.
- Le rajout de règles linguistiques.

5.6 Conclusion

Dans ce chapitre nous avons mentionné les principaux outils et paquets qui ont été implémentés dans notre programme, afin de savoir quel niveau d'exactitude le programme donne. Notre programme prend en entrée un ensemble de données qui contient des entités temporelles. Pour notre cas, nous avons testé notre programme avec une dataset. La fiabilité du programme dépend de la précision de notre analyse .mais ce n'est pas toujours le cas, car les données à analyser jouent un grand rôle sur la performance de résultat obtenu. Car si ces données contiennent trop d'ambiguïté, c'est vraiment difficile de détecter les entités qui exprime le temps. En fin nous avons présenté quelques exemples d'utilisation de notre outil.

Conclusion générale et perspectives

Les travaux présentés dans ce mémoire s'inscrivent dans le contexte de l'extraction des informations temporelles arabes. Les informations temporelles intégrées dans les documents sous la forme d'expressions temporelles offrent un moyen intéressant pour faire avancer et améliorer la fonctionnalité des applications de la recherche de l'information. Le traitement de la temporalité est crucial pour la compréhension de textes en langue naturelle, ce qui a motivé le développement d'outils pour la détection de ce type de donnée ainsi que l'extraction d'informations afin de permettre une tâche complète de l'extraction d'informations temporelles.

L'expression de la temporalité est complexe et variée, et sa compréhension l'est tout autant : une certaine connaissance du monde, un événement porteur de temporalité, des expressions temporelles, des marqueurs linguistiques,... tous ces aspects doivent être pesés et réfléchis dans le développement d'un système d'extraction d'expressions temporelles. Par ailleurs, un domaine spécifique peut avoir sa propre façon d'exprimer le temps, comme c'est le cas en médecine. Ces éléments multiples doivent être pris en considération. La dimension temporelle dans une langue précise est relativement stable et fait preuve d'une certaine régularité. Un système à base de règles peut donc être développé pour efficacement repérer et extraire ce type de données dans un texte, dans un document. C'est ce que nous avons démontré avec notre outil.

Dans notre travail, nous avons présenté un état de l'art sur la sémantique entre les mots en arabe. Nous avons également présenté un état de l'art sur les aspects des expressions temporelles arabe. Nous avons aussi présenté un état de l'art sur la détection et l'extraction des ET, en mettant l'accent sur les classes d'information temporelle arabe. Ensuite,

nous avons proposé une conception pour la reconnaissance des entités temporelles temporelles. Afin d'évaluer la faisabilité de notre proposition, nous avons développé outil de détection et d'extraction des ET dans les textes en arabe. Pour ce faire, nous avons utilisé l'environnement de développement intégré python.

Une perspective à court terme de ce travail est d'étendre l'annotation et la reconnaissance des expressions temporelles à la normalisation des expressions, selon la norme TimeML. L'enrichissement du corpus avec la normalisation des expressions est en cours. Et nous intéresser à la détection des événements et des relations entre événements et expressions temporelles.

Bibliographie

- [1] Claudine Bodson, Termes et relations sémantiques en corpus spécialisés : rapport entre patrons de relations sémantiques (PRS) et types sémantiques (TS), Département de linguistique et de traduction Faculté des arts et des sciences, 2004
- [2] Bakhouche Abdelali, Yamina Tlili-Guiassa. ,” Extraction des relations sémantiques à partir du Wiktionnaire Arabe ”, Laboratoire LRI/Equipe SRF Université Badji Mokhtar Annaba Algérie,2006.
- [3] [https ://www.Larousse.com/dictionnaire](https://www.Larousse.com/dictionnaire). DV :Août 2019.
- [4] Rey,Alain et DELESALLE. ”Problèmes et conflits lexicographiques ”, Langue Française, 4-26, 1979.
- [5] Vincent Nyckees , « La sémantique »,Paris :Belin, 1998.
- [6] Pierre cadiot, François Nemo « Sémantique lexicale »,p127-p146, 1997 .
- [7] Mark-Aronoff ., lanie Rees-Miller. ; « the handbook of linguistics, Wiley-blackweel», chapter 15 : An introduction to formal semantics,2003.
- [8] Claude Germain , « La sémantique fonctionnelle », Paris,1981.
- [9] A.J.Greimas, « Sémantique structurale »,1996.
- [10] [http ://dictionnaire.sensagent.leparisien.fr](http://dictionnaire.sensagent.leparisien.fr)(18.02.2019).
- [11] Lehmann, Alise. Martin-Berthe, Françoise , « Introduction à la lexicologie. Sémantique et morphologie », 2e édition. Paris. 2005. p. 53.
- [12] Niklas Salminen, La lexicologie, 2008 (page 110,123).
- [13] Polguère A, « Lexicologie et sémantique lexicale. Notions fondamentales », Troisièmes éditions, Les Presses de l’Université de Montréal, Montréal,2016.
- [14] Dubois, Giacomo, « Larousse-Le dictionnaire de linguistique et des sciences du langage », 2012, p. 369.

- [15] B. SCHWISCHA , « Syntaxe du français », 2013.
- [16] : J. Allen, N. Uzzaman, « TRIPS and TRIOS System for TempEval-2 : Extracting Temporal Information from Text. In Proceedings of the 5th International Workshop on Semantic Evaluation (SemEval '10) », pages 276–283, 2010.
- [17] Laurent Kevers. Accès sémantique aux bases de données documentaires : Techniques symboliques de traitement automatique du langage pour l'indexation thématique et l'extraction d'informations temporelles. Thèse de doctorat en en Langues et lettres, université catholique de louvain, louvain-la-neuve, 2011.
- [18] N. Turenne. Apport de l'extraction d'information temporelle à la, modélisation des réseaux biologiques. INRA - Unité Mathématique, Informatique et Génome (MIG), 2004.
- [19] Shaalan. K, A. Survey of Arabic Named Entity Recognition and Classification. Computational Linguistics, 40(2), 469-510, doi :10.1162/COLI a 00178, (2014).
- [20] Li, H., Hu, Y., Gao, G., Shnitko, Y., Meyerzon, D., Mowatt, David : Techniques for extracting authorship dates of documents (December 2009).
- [21] C. Teissedre. Analyse sémantique automatique des adverbiaux de localisation temporelle : application à la recherche d'information et à l'acquisition de connaissances. Thèse de doctorat en sciences du langage et traitement automatique des langues, université paris ouest nanterre la défense, 2012.
- [22] A. Bittar. Annotation temporelle de textes en français. Séminaire de linguistique de l'Université de Marne la Vallée, 2009.
- [23] G.Katz et I.Mani. The specification language TimeML. In I. Mani, J. Pustejovsky et R. Gaizauskas, Eds., The Language of Time : A Reader. Oxford University Press, 2005.
- [24] A.borillo., M.Bras, A.Ledraoulec, L.Vieu, A.Molendijk, H. deswart,H.Verkuyl, C.Vet, C.Vetters. Tense, connectives and discourse structure. In CORBLIN, F. de SWART, H., éditeurs : Handbook of French Semantics, CSLI Lecture Notes, pages 309à348. CSLI Publications, Standford, 2004.
- [25] J. Allen, N. Uzzaman. TRIPS and TRIOS System for TempEval-2 : Extracting Temporal Information from Text. In Proceedings of the 5th International Workshop on Semantic Evaluation (SemEval '10), pages 276–283, 2010.
- [26] Roser Saurí et al, Adapted from the TimeML Annotation Guidelines version

1.2.1, 2006.

[27] Tomassone, R., Grands repères culturels pour une langue : le français. Hachette Education, Paris, 2001.

[28] Noureddine Doumi. Extraction d'information à partir d'un texte arabe : extraction des entités nommées et leurs relations sémantiques. Intelligence artificielle [cs.AI]. Université Djillali Liabes de Sidi Bel Abbès, 2017.

[29] C. Habel et H. Schilder. From temporal expressions to temporal information : Semantic tagging of news message. In Proceedings of ACL'01 workshop on temporal and spatial information processing, pages 65–72, Toulouse, France. Association for Computational Linguistics, 2001.

[30] S-F. Adafre, D. AHN, M. de Rijke. Extracting temporal information from open domain text : A comparative exploration. Journal of Digital Information Management, 3(1) :14–20, 2005.

[31] Iman Saleh, Lamia Tounsi, and Josef van Genabith., ZamAn and Raqm : Extracting Temporal and Numerical Expressions in Arabic, Faculty of Computers and Information, Cairo University, 2011.

[32] Gupta, N. Temporal Information Extraction, Extracting Events and Temporal Expressions, A Literature Survey, (2015).

[33] Strötgen, J. Gertz, M. ., A Baseline Temporal Tagger for all Languages, 2015.

[34] Moriceau, V., Tannier, X, French resources for extraction and normalization of temporal expressions with heideltme, (2014).

[35] Marc Verhagen, Robert Gaizauskas , Frank Schilder, Mark Hepple , Graham Katz, "TempEval Temporal Relation Identification ", Stanford University, 2007.

[36] Marc Verhagen, Roser Saur, Tommaso Caselli , James Pustejovsky, " TempEval2", 2010.

[37] Naushad. UzZaman, Hector .Llorens, Leon Derczynski, Marc Verhagen, James Allen, Evaluating Time Expressions, Events, and Temporal Relations, University of Rochester, USA : University of Alicante, Spain : Department of Computer Science, University of Sheffield, UK : Computer Science Department, Brandeis University, USA : Nuance Communications, 2013.

[38] J. Pustejovsky et M. Verhagen. SemEval-2010 Task 13 : Evaluating Events, Time Expressions, and Temporal Relations (TempEval-2). In Proceedings of the Work-

shop on Semantic Evaluations, pages 112–116. ACL, 2009.

[39] A .Bittar. Annotation des informations temporelles dans des textes en français.

In Actes de la 12e édition de RECITAL, pages 11–20, Avignon, France, 2008.

[40] [https ://docs.python.org/fr/3/tutorial/](https://docs.python.org/fr/3/tutorial/) (consulté le 15 octobre 2019)

Les dictionnaires des mots qui expriment le temps

Les dictionnaires utiliser dans notre système sont présentes dans le latbleau B.1

Dictionnaire	Les mots
Ordinal	الاول، الأولى، الثانية، الثالثة، الرابعة، ثلاثة، ثلاث، اثنا، اثنا، ست، ستة، عشر، الأربع، الخمس، الست، أربعة، خمس، خمسة عشر، ثلاثون، مائة، السبع، الثمان، التسع، العشر، عشرين، ثلاثين، مائتين، ثلاثمئة، أربعمئة، خمسمائة، ستمائة، سبعمائة، اربعين، أربعون، خمسون، خمسين، ستون، ستين، تسعمائة، ألف، ألفين، مائة، مئة
Ordinal2	للعام، للسنة، لسنة، لقرن، للقرن، السنة، سنة، القرن، قرن، عام، اعوام، اشهر، ايامة يوم، شهر، سنين، سنوات عاما، شهر، يوما ليل، نهار، الليل، النهار، ساعة، دقيقة، ليلة، أسبوعا، أسبوع الماضي، المستقبل، حاضر، ماضي، الليلة، تمام، للعام، الحاضر مستقبل.....

Months	محرم، صفر، ربيع الأول، ربيع الثاني، جمادى الأولى، جمادى الآخرة، رجب، شعبان، كانون الثاني، شباط، آذار، نيسان، أيار، يناير، جانفي، فبراير، فيفري، مارس، أبريل، أفريل، مايو، ماي، يونيو، جوان، حزيران، تموز، آب، أيلول، تشرين
Calenders	قه، قم، قبل الميلاد، للميلاد، ميلادي، ميلادية، هـ، هـ، هـ، للهجرة، لهجرة، الهجرة، بعد الهجرة، هجري، هجرية، الهجرة الميلادي، الهجري، الميلادي
Adjs	الجارية، الماضية، القادمة، الجاري، الماضي، التالي، التالية، المنصرم، المقبل، المقبلة، الفائت، الفائتة، الحالي، الحالية، اوائل، اواخر، أواسط، معدودة، معدودة، محدودة المعدودة، أواسط، أواخر، متتالية، المتتالية، السابق، السابقة، المنصرمة،
Temps	ثانية، الساعة، الدقيقة، الثانية، دقيقة، ساعة، ساعتين، ساعات، دقيقتين، دقائق، ثانيتين، ثواني
Awqat	صبح، ظهر، عصر، مغرب، عشاء، صلاة، صباحا، مساء، فجرا، صباح، صبيحة، صبيحة، الصباح صباحا، ظهرا، عصرا، صباحا، مساء، فجر، المساء، الفجر، الصبح، الظهر، العصر، المغرب، العشاء
Middel-words	من، في، تمام، بعد، قبل، الى، بين، الموافق، بتاريخ

TABLE A.1 – Les dictionnaires utilisés.

Titre de l'annexe

Le tableau suivant B.1 présente un partie de dataset :

Input	Output1	Output2	Output3
غزوة بدر أو بدر الكبرى أو بدر القتال أو يوم الفرقان هي غزوة حدثت في السابع عشر من رمضان في العام الثاني من الهجرة بين المسلمين بقيادة الرسول محمد صلى الله عليه وسلم وقبيلة قريش وحلفائها	السابع عشر من رمضان في العام الثاني من الهجرة		
غزا النبي صلى الله عليه وسلم وصحابته رضي الله عنهم غزوات كثيرة، وقد اختلف العلماء والمؤرخون في عدد هذه الغزوات؛ ف قيل خمس وعشرون، وقيل سبع وعشرون، وقيل تسع وعشرون غزوة، وكانت كلها بعد الهجرة، واستمرت لعشرة أعوام	لعشرة أعوام	بعد الهجرة	
حدثت غزوة حنين في السابع من شوال من العام الثامن للهجرة	السابع من شوال من العام الثامن للهجرة		
أسباب الحرب، تتلخص في ظهور الحزب النازي	عام ١٩٣٣ م	عام ١٩٢٢ م	

في دولة ألمانيا عام ١٩٣٣ م ، وظهور الحزب الفاشي في دولة إيطاليا عام ١٩٢٢ م			
بلغ رسول الله - صَلَّى الله عليه وسلّم - ما كان من أمر انصراف غير قريش من الشام في رمضان على رأس ثلاثة عشر شهراً من الهجرة	رمضان على رأس ثلاثة عشر شهراً من الهجرة		
تمّ إعلان استقلال الجزائر في الخامس من تموز لعام ١٩٦٢ م	الخامس من تموز لعام ١٩٦٢ م		
اجتمع رأيه على رأي المختار الثقفي بالخروج ليلة الخميس في الرابع عشر من ربيع الأول لعام ٦٦ من الهجرة لقتال من قتل الحسين	ليلة الخميس في الرابع عشر من ربيع الأول لعام ٦٦ من الهجرة		
توالت الأحداث بسرعة، إذ تم إعلان انتحار وزير الدفاع في اليوم التالي	اليوم التالي		
في الفترة الممتدة منذ عام ثمانمائة وتسعين للميلاد وحتى عام ألف وأربعة للميلاد.	عام ثمانمائة وتسعين للميلاد	عام ألف وأربعة للميلاد	
تم إعدام عمر المختار في صبيحة يوم الأربعاء، الذي يوافق يوم السادس عشر من عام ١٩٣١ م، وذلك في تمام الساعة التاسعة صباحاً	يوم السادس عشر من عام ١٩٣١ م	صبيحة يوم الأربعاء	تمام الساعة التاسعة صباحاً

TABLE B.1 – Une partie de dataset.