

Ministère de l'enseignement supérieur et de la recherche scientifique

UNIVERSITE SAAD DAHLEB DE BLIDA 1

Faculté des Sciences

Département d'informatique



MÉMOIRE DE MASTER INFORMATIQUE
Spécialité : Ingénierie du Logiciel Ingénierie du
Logiciel

*Génération automatique de corpus de réponses
modèles dans un système d'évaluation automatique des
réponses courtes*

Réalisé par :

Abassi Selma

Boulhouache Houda

Proposé et encadré par : Mme OUAHRANI Leila

Composition de jury :

Président : Mme Mezzi Melyara

Examineur : Mme Bacha Siham

Soutenu le : 29/09/2019

المخلص

أصبح التقييم التلقائي واحدا من أهم الأسلحة في مجال التعليم، لكفاءته من حيث توفير الوقت لذلك لم يعد التعلم التقليدي قائما. ومع ذلك، فإن التعلم الإلكتروني هو أكثر بكثير من مجرد بديل له العديد من المزايا مثل أداره الوقت والحد من عبء العمل للمعلمين (مكاسب الجهد). فبالنالي تقييم هذا التعلم و جعله تلقائي أمر ضروري تتركز دراستنا علي التقييم التلقائي. حيث نقترح أسلوبا من شأنه أن يولد قاموسا من المرادفات ومجموعه من الردود النموذجية لتصميم نظام تقييم تلقائي للإجابات القصيرة المتكيفة مع اللغة العربية استنادا إلى نهج الفضاء الدلالي وكلمه تضمين. وتقتراح طرق حساب التشابه مع مجموعه من القياسات النحوية. وقد أدى هذا النهج إلى تحسين نتائج العمل السابق.

الكلمات المفتاحية: التقييم التلقائي للردود القصيرة، تضمين الكلمات، الفضاء الدلالي، مقاييس التشابه، قاموس مرادف، إعادة صياغة

Résumé

L'évaluation automatique est devenue une des armes les plus importantes dans le champ de l'éducation, pour son efficacité en termes de gain de temps. Il est souhaitable que l'apprentissage traditionnel n'existe plus. Toutefois, l'apprentissage en ligne est bien plus qu'une simple alternative : c'est une excellente alternative qui a de nombreux avantages tels que la gestion de temps (gains de temps) et la réduction du volume de travail des enseignants (gains d'efforts). Cependant, l'évaluation de cet apprentissage est nécessaire et son automatisation est plus appréciée.

Notre étude s'intéresse à l'évaluation automatique. Nous proposons une méthode qui va générer un dictionnaire des synonymes et un corpus de réponses modèles pour la conception d'un système d'évaluation automatique des réponses courtes adapté à la langue arabe basée sur l'approche espace sémantique et les word embedding. Des méthodes de calcul de similarité sont proposées avec une combinaison des mesures syntaxiques. L'approche a permis d'améliorer des travaux qui sont déjà réalisés dans le même contexte.

Mots clés : Evaluation automatique des réponses courtes, Word Embedding, espace sémantique, Mesures de similarité, dictionnaire des synonymes, paraphrases

Abstract

Automatic evaluation has become one of the most important weapons in the field of education, for its efficiency in terms of saving time. We do not wish that traditional apprenticeship no longer exists. Nonetheless, e-learning is much more than a simple alternative: it is a great alternative that has many benefits such as time management (time savings) and the reduction of the workload of teachers (effort gain). Yet, the evaluation of this learning is necessary and its automation is more appreciated.

Our study focuses on evaluation; we propose a method that will generate a thesaurus and a corpus of model responses for the design of a system of automatic assessment of short responses adapted to the Arabic language based on the semantic space approach and word embedding. Similarity calculation methods are proposed with a combination of syntactical measures. The approach has improved the work that is already done in the same context

Keywords: Automatic evaluation of short responses, Word Embedding, semantic space, Similarity measures, synonym dictionary, paraphrases

Remerciements

Nous tenons tout d'abord à remercier Allah, qui nous a donné la force, la capacité et la patience pour accomplir ce travail.

Nos profondes gratitudee et sincères remerciements vont à notre promotrice Madame Ouahrani Leila pour l'aide qu'elle nous a fournie et les connaissances qu'elle a su nous transmettre. Nous La remerions également pour sa disponibilité et la qualité de ses conseils.

Il est naturel que nos pensées les plus fortes aillent vers nos parents. Qu'ils sachent que l'amour qu'ils nous donnent continue à nous animer et nous permet d'envisager l'avenir comme un défi. Ce travail est dédié à nos grands parents qui étaient toujours devant nous.

Nous tenons à exprimer nos sincères remerciements à nos frères et sœurs, et les familles : Boulhouache, Abassi, kisserli benelkadi, zitouni et tous ceux qui nous ont aidés tant moralement que matériellement, qui par leurs paroles, leurs écrits, leurs conseils et leurs critiques ont fait de nous ce que nous sommes aujourd'hui.

Nous remercions nos amies pour leur soutien et encouragement, Sans oublier HENNICHE Adel Nassim et BENAYAD Asma pour leurs aides. Merci à vous.

Nous souhaitons adresser nos remerciements les plus sincères au corps professoral et administratif de l'Université Saad Dahleb, Blida, pour la richesse et la qualité de leur enseignement et qui déploient de grands efforts pour assurer à leurs étudiants une formation actualisée.

Nous tenons à remercier toute personne qui a participé de près ou de loin à l'exécution de ce modeste travail.

Liste des Tables

Tableau 1: Exemple de stemmer.....	27
Tableau 2: Exemple de génération des synonymes avec le modèle Aravec CBOW.....	39
Tableau 3: Exemple de génération des synonymes avec le modèle Aravec SkipGram.....	39
Tableau 4: Exemple génération des synonymes avec le modèle Zahran CBOW.....	39
Tableau 5: Exemple de génération des synonymes avec le modèle Zahran SkipGram.....	40
Tableau 6 : matrice générée avec le modèle CM.....	50
Tableau 7: sélection des nombre dans la matrice générée avec le modèle CM.....	51
Tableau 8 : la matrice après élimination des mots en commun.....	51
Tableau 9: la matrice après élimination du maximum.....	51
Tableau 10: évaluation des synonymes selon les mots de sketch engine.....	64
Tableau 11: Signification des valeurs la corrélation de pearson.....	65
Tableau 12: Aperçu du dataset de CCASAG.....	67
Tableau 13: Résultats du modèle CM avec l'espace sémantique CNN Tashapyne stem.....	72
Tableau 14: Résultats du modèle CM avec l'espace sémantique Khaleej avec Tashapyne stem].....	73
Tableau 15: Résultats du modèle CM avec l'espace sémantique CNN Tashapyne stem.....	73
Tableau 16: Résultats du modèle CM avec l'espace sémantique BBC+CNN avec Khoja stem.....	73
Tableau 17: Résultats du modèle CM avec l'espace sémantique Khaleej avec Khoja stem.....	74
Tableau 18: Résultats du modèle CM avec l'espace sémantique Khaleej avec tashapyne stem combiné avec dice.....	75
Tableau 19: Résultats du modèle CM avec l'espace sémantique cybercriminalité avec Tashaphyne stem.....	75
Tableau 20: Résultats du modèle CM avec l'espace sémantique cybercriminalité avec Khoja stem.....	76
Tableau 21: Résultats du modèle CM avec l'espace sémantique cybercriminalité avec Stanford_core_nlp stem.....	76
Tableau 22: Résultats du modèle CM avec l'espace sémantique cybercriminalité avec tashapyne stem combiné avec dice.....	77

Tableau 23: Résultat du modèle CM en utilisant les WE en avec stanford stem.....	77
Tableau 24: Résultat du modèle CM en utilisant les WE Combiner avec Dice.....	78
Tableau 25: Résultat du modèle CM en utilisant les WE en avec stanford stem.....	78
Tableau 26: Résultat du modèle CM en utilisant les WE Combiner avec Dice.....	79
Tableau 27: Résultats du modèle SV avec l'espace sémantique Khaeej Tashapyne stem	79
Tableau 28: Résultats du modèle SV en utilisant la moyenne avec l'espace sémantique Khaeej Tashapyne stem.....	80
Tableau 29: Résultats du modèle SV en utilisant le max avec l'espace sémantique Khaeej Tashapyne stem.....	80

Liste des Figures

Figure II-1 : Organisation historique des systèmes ASAG [3].....	8
Figure II-2: Les approches de similarité	9
Figure II-3: Les mesures syntaxiques	11
Figure II-4: Mesures de similarité sémantique	12
Figure II-5:L'architecture de modèle CBOW	15
Figure II-6: L'architecture de modèle SkipGram	16
Figure III-1: schéma de représentation des étapes de notre système.....	23
Figure III-2: phase de création du dictionnaire avec l'espace sémantique	28
Figure III-3: Phase de création de l'espace sémantique	29
La figure III-4 présente un exemple de traitement d'un corpus.	Erreur ! Signet non défini.
Figure III-5: la représentation du corpus par le modèle Bac of words.....	30
Figure III-6: Exemple de traitement d'un corpus	32
Figure III-7: Un espace sémantique.....	34
Figure III-8: La matrice de profils.....	35
Figure III-9: une matrice de profils	37
Figure III-10: phase de création du dictionnaire avec les WE	38
Figure III-11: La génération des paraphrases	40
Figure III-12: Vue globale sur le fonctionnement du modèle CM.....	45
Figure III-13: Algorithme de calcul de similarité	49
Figure III-14: Vue globale sur le fonctionnement du modèle SV en utilisant la moyenne	52
Figure III-15: Vue globale sur le fonctionnement du modèle SV en utilisant le max.....	54
Figure III-16: Matrice des similarités max	56
Figure III-17: Vue globale sur le fonctionnement du modèle SV en utilisant les paraphrases.....	57
Figure III-18: Classification selon le K-means	61
Figure IV-1: Outils de normalisation.....	68

Figure IV-2: Outils de stemming.....	69
Figure IV-3: Outils de création du dictionnaire	70
Figure IV-4: Outils d'évaluation automatique des réponses courte.....	71

Glossaire

Terme	Définition
ASAG	Système d'évaluation automatique des réponses courtes (Automatic Short Answer Grading System)
RM	Réponse Modèle
RE	Réponse étudiant
WE	Word Embedding
CM	Calcul matriciel
SV	Somme des vecteurs
CCASAG	Cyber Criminality Arabic Short Answer Grading Dataset

Table des matières

chapitre I. Contexte et problématique	1
1. <i>Introduction</i>	1
2. <i>Problématique</i>	1
3. <i>Objectifs du travail</i>	2
4. <i>Importance du travail</i>	2
5. <i>Portée et limites du travail</i>	3
chapitre II. Etat de l'art	4
1. <i>Le système d'évaluation automatique des questions à réponse courte (ASAG)</i>	4
1.1. Une vue historique sur les systèmes ASAG.....	5
2. <i>Les approches de similarité d'ASAG</i>	8
2.1. Similarité syntaxique :.....	9
2.2. Similarité sémantique :.....	11
2.3. Similarité hybride.....	13
3. <i>La représentation des mots</i>	13
3.1. Le modèle d'espace vectoriel.....	13
3.2. Le modèle sac de mots.....	13
3.3. Les Word Embedding.....	14
4. <i>Les travaux connexes sur l'évaluation automatique des réponses courtes</i>	16
4.1. Les paraphrases :.....	18
4.2. La génération des paraphrases :.....	19
4.3. La traduction automatique statistique :.....	19
4.4. Méthodologie de génération typique (logique des prédicats).....	20
4.5. Génération à l'aide des réseaux de neurones récurrents.....	20
4.6. Les travaux liés au même Projet entamés durant l'année 2017/2018.....	20
chapitre III. Système de génération automatique des réponses modèles dans un système ASAG 22	
1. <i>Méthodologie</i>	22
1.1. La normalisation.....	25

1.2.	Le stemming	25
2.	<i>Construction du dictionnaire des synonymes</i>	27
2.1.	Dictionnaire basé sur l'espace sémantique	28
2.2.	Dictionnaire basé sur une représentation par Les word Embedding	38
3.	<i>Génération des paraphrases pour la réponse modèle</i>	40
3.1.	La combinaison des phrases.....	40
4.	<i>Modèles du calcul de similarité sémantique entre deux réponses courtes</i>	44
4.1.	Le modèle calcul-matriciel (CM).....	45
4.2.	Le modèle somme-vecteurs (SV).....	52
5.	<i>Pondération des mots basée sur la fréquence de mots dans le corpus</i>	57
5.1.	Calcul des poids	57
5.2.	Pondération des mots basée sur l'étiquetage morphosyntaxique	59
5.3.	Combinaison entre le POS et le TF_MIN_MAX.....	60
6.	<i>Passage au score</i>	60
6.1.	Classifieur k-means [42] [43].....	60
	chapitre IV. Résultats expérimentaux et évaluation	62
1.	<i>Démarche expérimentale</i>	62
1.1.	L'impact de l'élaboration de corpus des réponses modèle dans l'évaluation automatique des réponses par composition à base de synonymes	62
1.2.	L'impact du stemmer	63
1.3.	La qualité du dictionnaire (évaluation du dictionnaire).....	63
2.	<i>Métriques d'évaluation</i>	64
2.1.	Coefficient de Pearson	65
2.2.	Erreur quadratique RMSE (Root Mean Squared Error (RMSE))	65
3.	<i>Jeux de données (Datasets)</i>	66
3.1.	Gomaa Dataset (Cairo university dataset)[46][47].....	66
3.2.	Cyber criminality Arabic Short Answer Grading Data set (CCASAG) :	66
4.	<i>Outils développés</i>	67
4.1.	Outils de Normalisation et stemming	67
4.2.	Outils de la génération du dictionnaire des synonymes pour les réponses modèle	69

4.3. Outils d'évaluation automatique des réponses courte	70
5. <i>Ressources matérielles et logicielles utilisées lors du développement</i>	71
6. <i>Résultats et discussion</i>	72
6.1. Résultat en utilisant les espaces sémantique Pour le modèle CM.....	72
6.2. Résultat en utilisant les WE	77
6.3. Résultat en utilisant l'espace sémantique Pour le modèle SV.....	79
chapitre V. Conclusion et Perspectives	82

chapitre I. Contexte et problématique

1. Introduction

L'évaluation est un outil essentiel tout au long de l'apprentissage. Dans l'évaluation traditionnelle, chaque heure consacrée à l'évaluation est pratiquement perdue pour l'apprentissage, des chercheurs ont concentré leurs travaux sur l'automatisation de l'évaluation afin de réduire la charge de travail des enseignants et fournir aux étudiants un apprentissage plus efficace en intégrant la culture d'évaluation au travail quotidien dans un environnement E-Learning. Depuis plusieurs années des systèmes d'évaluation automatique sont en pratique dans le domaine de l'enseignement en ligne, tels que les réponses à choix multiple (QCM) ou à réponses vrai/faux qui sont jugées insuffisantes ne permettant pas la saisie des connaissances acquises, à l'opposé des questions à réponses courtes présentées par l'apprenant en langage naturel qui sont jugées plus correctes dans l'évaluation de ses connaissances acquises. La recherche en question-réponse se développe tous les jours et depuis les années 1990, on commence à se pencher et à s'intéresser à la langue arabe. A notre connaissance, il y a beaucoup de systèmes qui ont été développés pour de nombreuses langues du monde (ex. anglais, français, chinois, japonais, etc.). Cependant, il existe peu de systèmes qui ont été développés pour la langue arabe. Ces systèmes fournissent des réponses sous forme de passages courts, des extraits de collection de documents. Par conséquent, la performance de ces systèmes est limitée par la difficulté de traitement de la langue arabe et le manque considérable d'outils de TAL efficaces [1]. En revanche, notre travail consiste à automatiser la création d'un corpus de réponses modèle en utilisant les synonymes.

2. Problématique

Le principe général de l'évaluation consiste à comparer la réponse de l'apprenant avec la réponse modèle (de référence) formulée par l'enseignant et d'attribuer un score. Les diverses approches de correction automatique se basent sur un ensemble de corpus de réponses contenant un nombre important de réponses à une question. Ces réponses sont associées à différentes formulations possibles que l'enseignant pourrait imaginer. La mise en place d'un corpus efficace n'est pas une tâche facile et demande un temps très important puisque réalisée souvent

manuellement ou par le biais de grammaires que même l'enseignant informaticien ne maîtriserait pas toujours.

3. Objectifs du travail

L'objectif principal consiste à automatiser la création d'un corpus de réponses modèles. Avec cette automatisation l'enseignant serait déchargé de la création d'un corpus de réponses modèles possibles. Une réponse modèle écrite par l'enseignant constitue l'embryon du corpus des réponses modèles. Le reste du corpus sera généré automatiquement en appliquant les observations faites par l'enseignant.

L'automatisation du corpus de réponses modèles doit passer par la proposition d'un modèle de spécification de la réponse modèle qui permettrait de générer automatiquement le corpus de toutes les réponses possibles avec les contraintes d'évaluation imposées par l'enseignant.

4. Importance du travail

L'importance de notre travail réalisé dans le cadre de ce projet de fin d'étude se résume dans les points suivants :

- Réduire la charge de travail des enseignants en automatisant une partie de la tâche d'évaluation
- Intégrer la culture d'évaluation au travail quotidien des apprenants dans un environnement d'e-Learning
- Améliorer la continuité de l'apprentissage pour tous les étudiants.
- Mettre en œuvre un dictionnaire distribué de synonymes
- Générer des paraphrases à partir du dictionnaire distribué de synonymes

5. Portée et limites du travail

- ✓ L'exigence de la disponibilité d'une réponse modèle (réponse de référence)
- ✓ Les réponses doivent être rédigées dans un langage naturel.
- ✓ La disponibilité d'un corpus lié au domaine de la recherche
- ✓ Le système développé est dédié pour l'évaluation en la langue arabe, toute autre ressource, mot ou paraphrase d'une langue autre que l'arabe sera ignoré
- ✓ Les questions doivent être des questions à réponses courtes non pas d'autres types tels que les questions de sélection, les essais, la classification, les codes (programmes à évaluer) ou les formules mathématiques

chapitre II. Etat de l'art

Dans ce chapitre nous allons situer notre travail par rapport aux fonctionnements des ASAGS, les différentes approches d'évaluation automatique de réponses courtes, ainsi que les travaux connexes déjà réalisés par rapport à notre travail.

1. Le système d'évaluation automatique des questions à réponse courte (ASAG)

Le principe de l'évaluation automatique des questions à réponses courtes consiste à comparer la réponse de l'apprenant avec la réponse modèle et à mesurer la similarité entre les deux réponses pour l'attribution d'un score. La recherche dans ASAG a également été considérée comme un problème de reconnaissance des paraphrases [2] Par conséquent, la comparaison des réponses des enseignants et des élèves dans ASAG pourrait être étayée par la similitude textuelle sémantique et la paraphrase des communautés. En plus de la similitude textuelle sémantique et de la paraphrase, le domaine des systèmes intelligents de tutorat peut également être considéré comme étant lié comme une forme plus interactive de systèmes ASAG. Les questions à réponse courte sont traditionnellement utilisées tout au long du processus d'apprentissage. Elles renforcent l'apprentissage et aident à développer des compétences cognitives, et elles sont l'instrument préféré de l'examineur parce qu'elles évaluent efficacement la compréhension sans offrir des instructions ou des indices.¹ D'après [3] elles peuvent être considérées comme suit:

- La question doit exiger une réponse qui rappelle les connaissances externes au lieu d'exiger que la réponse soit reconnue à l'intérieur de la question.
- La question doit exiger une réponse donnée en langage naturel.
- La longueur de réponse doit être à peu près entre une phrase et un paragraphe

¹www.intelligentassessment.com

- L'évaluation des réponses devrait se concentrer sur le contenu plutôt que sur le style d'écriture.
- Le niveau d'ouverture dans les réponses ouvertes par rapport aux réactions closes devrait être restreint par une conception objective des questions.

1.1. Une vue historique sur les systèmes ASAG

Bien qu'il se trouve dans le temps plusieurs systèmes ASAG exploités par le système pédagogique exploité, il en ressort un des plus récent et plus complet qu'est « The Eras and Trends of Automatic Short Answer Grading » [3] où les auteurs regroupent les systèmes d'évaluation automatiques des réponses courtes dans 5 ères différentes représentés dans la figure II-1 avec les outils ASAGS les plus connus de chaque ère:

1.1.1. Les méthodes basées sur le mappage de concepts (Concept Mapping)

En 1996 la technique de mappage de concept est apparue. Cette technique est exprimée au niveau de la phrase, L'idée est de considérer les réponses des étudiants composées de plusieurs concepts et pour détecter la présence ou l'absence de chaque concept lors de la notation. Nous prenons exemples des systèmes ASAG :

Burstein : Burstein et al. (1996)[4] utilisent la technique de la représentation de la structure conceptuelle lexicale selon laquelle un lexique conceptuel et une grammaire conceptuelle doivent être développés à partir d'un ensemble d'apprentissage avant de classer les hypothèses dans les réponses des étudiants.

c-rater : [5] c'est un moteur de notation automatisé qui a été développé pour marquer les réponses aux questions de réponse courte, il est basé sur le contenu. Il ne s'agit pas simplement d'un programme d'appariement des chaînes, mais plutôt d'un ensemble des règles et une représentation canonique des textes en utilisant la variation syntaxique, anaphore, morphologique variation, synonymes et correction orthographique pour attribuer un score complet ou partiel à une question de réponse courte. C-rater a été utilisé dans deux études: National Assessment for Educational Progress (NAEP) et une évaluation à l'échelle de l'État dans l'Indiana. Dans les deux études, c-rater était d'accord avec les personnes humaines environ 84 % du temps.

1.1.2. Les méthodes basées sur l'extraction d'information (Information Extraction)

C'est une série d'opérations de correspondance de modèle telles que des expressions régulières ou des arbres d'analyse. Elles peuvent extraire des données structurées à partir de sources non structurées, tels que le texte libre et représentent les données structurées sous forme de n-uplets utilisables dans de nombreuses applications. Nous prenons exemples des systèmes ASAG :

eMax:[6] Le système eMax est conçu pour prendre en charge divers scénarios d'examen, il fournit une évaluation semi-automatique des textes courts librement formulés ainsi que des problèmes mathématiques partiellement résolus. L'approche de la notation est une approche combinatoire, où toutes les formulations possibles sont prises en compte lors de l'appariement de modèles.

FreeText Author : [7] Le système a été utilisé pour écrire et marquer les tâches d'évaluation de texte libre à réponse courte, il fournit une interface utilisateur graphique pour l'entrée de réponse de l'enseignant et le classement des réponses des élèves. Les réponses de l'enseignant sont composées comme des modèles syntaxiques-sémantiques pour que les réponses de l'élève soient comparées. Ces modèles sont automatiquement générés à partir de l'entrée de langage naturel des réponses des enseignants.

1.1.3. Les méthodes basées-corpus (Corpus Based Methods)

En exploitant les propriétés statistiques de grands corpus de documents. Bien que ces méthodes sont souvent utilisées pour des applications avec des textes plus longs. Ces méthodes peuvent également être utiles lors de l'interprétation des synonymes dans les réponses courtes, utiliser uniquement le vocabulaire de réponse original et de limiter les réponses correctes qui peuvent être identifiées. Une technique typique pour augmenter le vocabulaire consiste à utiliser corpus parallèles bilingues pour analyser la fréquence de résolution des paires de termes en une même traduction en langue seconde. Ensuite, des synonymes avec des traductions particulièrement courantes peuvent être incorporés dans la réponse de l'enseignant. Nous prenons comme exemples des systèmes ASAG :

SAMText : (Short Answer Measurement of TEXT) applique une variante de LSA basée sur une structure de données inversées d'index, qui est estimée par le contenu d'un analyseur Web utilisant des documents topiques pertinents. En revanche, LSA utilise normalement une structure de données matricielle basée sur de grands corpus pour modéliser la parenté sémantique. L'index inversé et l'idée rampante est plus approprié pour des réponses courtes comparées aux réponses longues parce que les explorations de web peuvent être adaptées à chaque sujet au lieu d'essayer de modéliser toute la langue à la fois [8].

Mohler : [9] utilise des techniques non supervisées pour la tâche de classement automatique de réponse courte, en comparant un certain nombre de mesures fondées sur la connaissance et le corpus de la similitude de texte et évaluant l'effet du domaine et de la taille sur les mesures basées sur le corpus, et introduisant également une technique pour améliorer les performances du système en intégrant commentaires des réponses de l'élève.

1.1.4. Les méthodes basées sur l'apprentissage automatique (Machine learning)

En 2008 une nouvelle technique est devenue célèbre, La technique d'apprentissage automatique. Les systèmes d'apprentissage par machine utilisent généralement un certain nombre de mesures extraites des techniques de traitement du langage naturel similaires, qui sont ensuite combinés en une seule note ou un score en utilisant un modèle de classification ou de régression. Cela peut être pris en charge par une boîte à outils d'apprentissage machine telle que [10]. Les caractéristiques impliquant le sac-de-mots et les n-grammes sont typiques de cette catégorie, comme sont des arbres de décision et supportent des machines vectorielles comme des algorithmes d'apprentissage représentatifs. Nous prenons des exemples des systèmes ASAG Les systèmes les plus représentatifs sont :

CAM : [11] Utilise une approche pour l'évaluation du contenu des réponses des apprenants de langue anglaise aux questions de compréhension de la lecture à réponse courte, il utilise un classificateur k-plus proche voisin (KNN) et des caractéristiques qui mesurent le pourcentage de chevauchement du contenu sur différents niveaux linguistiques entre les réponses de l'enseignant et celles de l'élève. Il atteint une précision de 87% pour la détection d'erreurs sémantiques qui est amélioré jusqu'à 88% sur un ensemble de données de test retenu.

Madnani 13 : [12] il mesure les compétences de compréhension de la lecture des élèves par le marquage automatique des résumés rédigés par les élèves et compare les résultats par rapport à la notation humaine. L'approche d'apprentissage automatique comprend huit caractéristiques (BLEU, ROUGE, mesures concernant différentes dimensions de la copie de texte, nombre de phrases, et nombres de mots de connecteur de discours couramment utilisés) comme entrée à un classificateur de régression logistique.

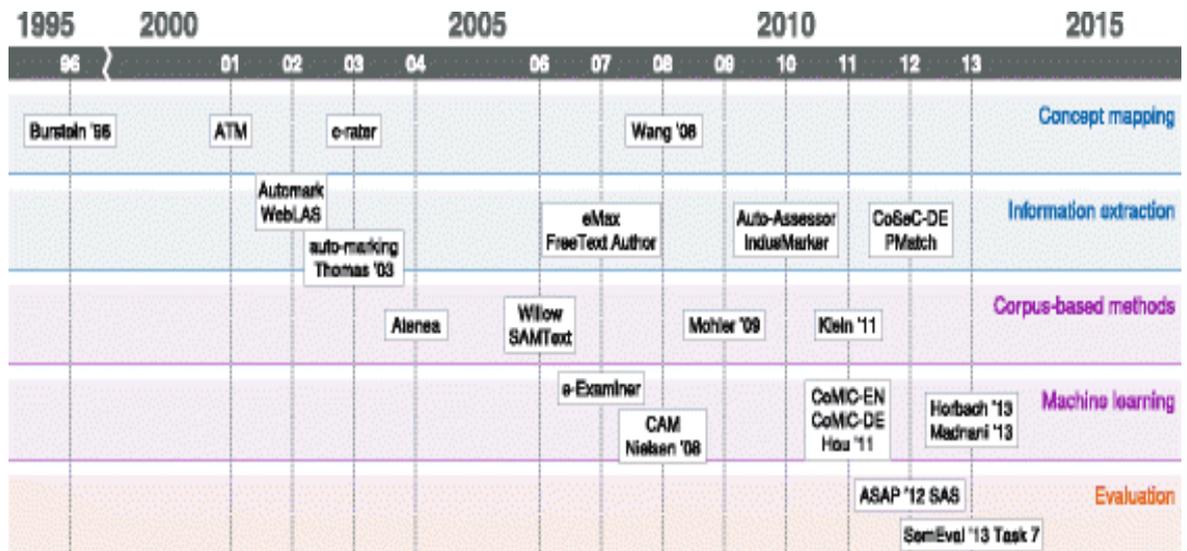


Figure II-1 : Organisation historique des systèmes ASAG [3]

2. Les approches de similarité d'ASAG

Évaluer des similarités entre entités textuelles est un des problèmes centraux dans plusieurs disciplines comme l'évaluation automatique des réponses, l'analyse de données textuelles, la recherche documentaire ou l'extraction de connaissances à partir de données textuelles (Text Mining)... Dans chacun de ces domaines, les similarités sont en effet utilisées pour une large variété de traitements. Nous citons trois catégories principales des approches de similarité (Figure II-2) :

1. Similarité syntaxique (String-based similarity).
2. Similarité sémantique (Semantic similarity).
3. Similarité hybride (hybrid similarity)

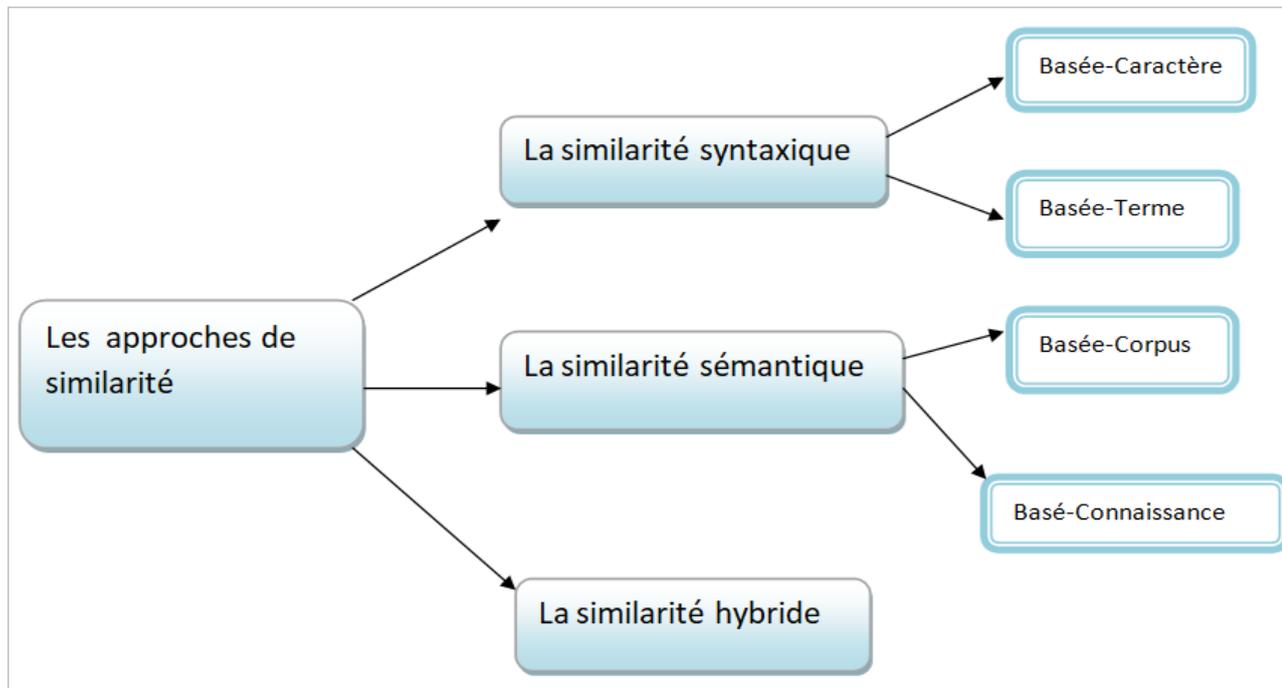


Figure II-2: Les approches de similarité

2.1. Similarité syntaxique :

Elle consiste à comparer les documents textuels, c'est une métrique qui mesure la similarité entre 2 chaînes de caractères. Par exemple, les chaînes de caractère "يدخل" et "دخل" peuvent être considérées comme très proches, alors que "دخل" et "ولج" pourront être considérées comme très différents [13]. La figure II-3 montre les mesures de similarité syntaxique. Il existe des mesures qui sont basées sur des caractères tandis que les autres sont des mesures de distance basée sur les termes. Parmi ces mesures les plus utilisées :

2.1.1. Similarité Cosinus

La similarité cosinus est fréquemment utilisée [14] en tant que mesure de ressemblance entre deux texte $d1$ et $d2$. Il s'agit de calculer le cosinus de l'angle entre les représentations vectorielles des textes à comparer. La similarité obtenue $Sim_{cosinus}(d1, d2) \in [0, 1]$.

$$sim_{cosinus}(d1, d2) = \frac{\vec{d1} \cdot \vec{d2}}{\|\vec{d1}\| \|\vec{d2}\|}$$

2.1.2. Coefficient de corrélation de Pearson

Calculé la similarité entre deux textes d1 et d2 comme le cosinus de l'angle entre leurs représentations vectorielles centrées-réduites. La similarité obtenue $Sim_{pearson}(d1, d2) \in [-1, 1]$.

$$sim_{pearson}(d1, d2) = sim_{cosinus}(d1 - \bar{d1}, d2 - \bar{d2})$$

Où d1 (resp. d2) représente la moyenne de d1 (resp. d2).

2.1.3. Distance euclidienne

Calcule la similarité entre deux textes d1 et d2 comme la distance entre leurs représentations vectorielles ramenées à un seul point.

$$sim_{euclidienne}(d1, d2) = \|\bar{d1} - \bar{d2}\| = \sqrt{\sum_{i=1}^n (d1_i - d2_i)^2}$$

Où n est le nombre total de termes représentés.

2.1.4. Coefficient de Jaccard

Le rapport entre la cardinalité de l'intersection des ensembles considérés et la cardinalité de l'union des ensembles. Il permet d'évaluer la similarité entre les ensembles. Les documents d1 et d2 sont donc représentés non pas comme des vecteurs, mais comme des ensembles de termes. La similarité obtenue $(d1, d2) \in [0, 1]$. [15]

$$sim_{jaccard}(d1, d2) = \frac{\|d1 \cap d2\|}{\|d1 \cup d2\|}$$

2.1.5. L'indice de Dice

Mesure la similarité entre deux textes d1 et d2 en se basant sur le nombre de termes communs à d1 et d2.

$$sim_{dice}(d1, d2) = \frac{2N_c}{N1 + N2}$$

Où Nc est le nombre de termes communs à d1 et d2, et N1 (resp. N2) est le nombre de termes de d1 (resp. d2).

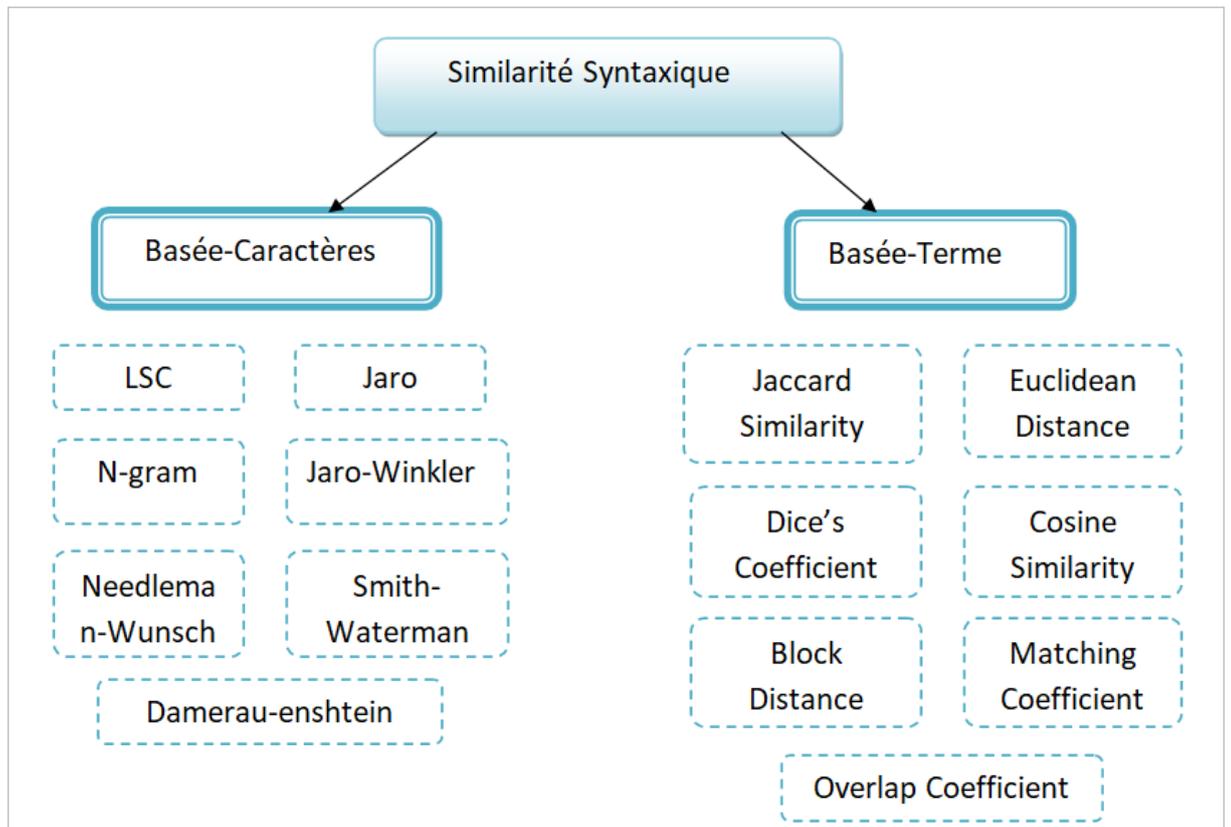


Figure II-3: Les mesures syntaxiques

Afin de mesurer la similarité syntaxique il faut construire le modèle vectoriel (vector space model). Il réduit la complexité des documents et facilite leur manipulation c'est-à-dire sa version textuelle intégrale, en un vecteur qui décrit le contenu du document, La représentation d'un document sous forme vectorielle se déroule en 2 étapes :

Extraction des termes pertinents : Il s'agit de pré-traiter le texte des documents textuels en supprimant les mots-vides, la ponctuation et les éventuels 'retours-chariots', de lemmatiser le texte et de le segmenter.

Calcul des poids : Le poids de chaque terme dans un document peut être obtenu de différentes manières : booléenne, fréquence des termes, Tf-Idf (Term frequency - Inverse Document Frequency). [16] traité en détail dans le chapitre 3

2.2. Similarité sémantique :

La similarité sémantique est un concept selon lequel un ensemble de documents ou de termes se voient attribuer une métrique basée sur la ressemblance

de leur signification (contenu sémantique) [13]. La figure II-4 montre les mesures de similarité sémantique les plus répandues.

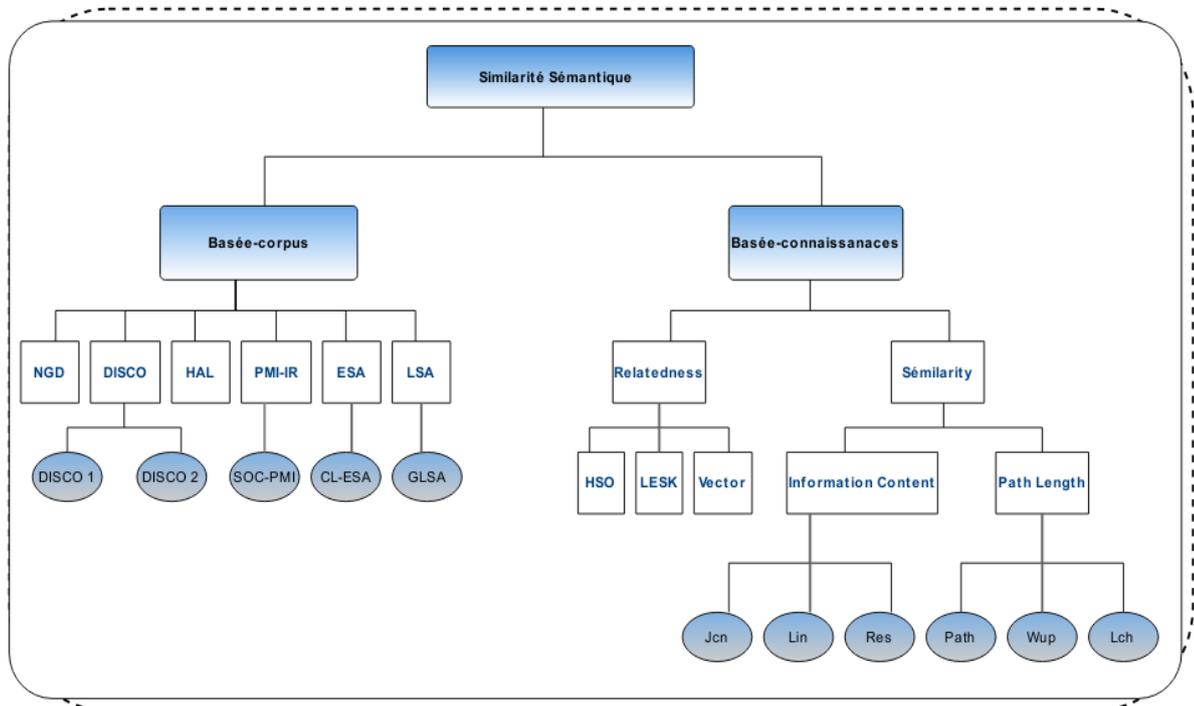


Figure II-4: Mesures de similarité sémantique

D'après la figure II-4 nous distinguons deux types de similarité sémantique :

2.2.1. Les mesures basées sur des corpus

C'est une mesure statistique qui détermine la similarité entre les mots en fonction de l'information obtenue d'un corpus volumineux. Elle consiste à créer un espace sémantique à partir des cooccurrences de mots. Une matrice mot à mot est formée dont chaque élément de la matrice est la force d'association entre le mot représenté par la ligne et le mot représenté par la colonne. Les valeurs matricielles sont accumulées en pondérant la cooccurrence de manière inversement proportionnelle à la distance de focalisation du mot. Les mots voisins les plus proches sont considérés comme reflétant davantage la sémantique du mot cible et sont donc pondérés plus haut c'est-à-dire attribuer une importance. Les mots sont représentés par des vecteurs sémantique d'où la notion d'espace vectoriel VSM (Vector Space Model) [17].

2.2.2. Les mesures basées connaissances (Knowledge-Based Similarity)

Elle situe sur l'identification du degré de similitude entre les mots à l'aide d'informations dérivées de ressources externes tel que le réseau WordNet² ou les dictionnaires de synonymes [18]

Dans la Figure II-4 nous pouvons reconnaître quelques mesures les plus répondues.

2.3. Similarité hybride

Les méthodes hybrides utilisent une combinaison de mesures de similarité, soit des méthodes sémantiques et syntaxiques à la fois où bien plusieurs méthodes de même type. Elle donne de supérieures valeurs du facteur de corrélation et donc de meilleurs résultats par rapport à l'application d'une seule mesure de similarité. [19]

3. La représentation des mots

3.1. Le modèle d'espace vectoriel

Le modèle d'espace vectoriel est un modèle mathématique pour représenter des documents textuels en tant que vecteurs d'identificateurs. La représentation d'un document sous forme vectorielle se déroule comme suit. Pour chaque document dans un ensemble, on compte le nombre d'occurrences de chaque mot (fréquence d'occurrence du terme, notée « tf »). Ce décompte ou cette vérification des occurrences forme un vecteur pour chaque document et l'ensemble des vecteurs forme une matrice parfois appelée matrice d'occurrences (nous allons l'expliquer par la suite dans le chapitre 3). Au lieu de compter le nombre d'occurrences, on peut aussi simplement vérifier la présence du terme et attribuer la valeur 1 lorsque le terme est présent, et la valeur 0 quand le terme est absent.

3.2. Le modèle sac de mots

Le modèle de sac de mots est un moyen d'extraire des caractéristiques du texte pour les utiliser dans des algorithmes de l'apprentissage automatique, plus précisément dans le traitement automatique du langage, la représentation sac de mot (Bow) consiste à décrire un document comme un ensemble de sous document chacun

² <https://wordnet.princeton.edu>

étant représenté par un sac de mots .Les textes sont transformés simplement en vecteurs dont chaque composante représente un mot .

3.3. Les Word Embedding

C'est une représentation savante pour le texte où les mots qui ont le même sens ont une représentation similaire. Cette approche de la représentation des mots et des documents qui peut être considéré comme l'une des percées clés du deep learning sur les problèmes de traitement du langage naturel difficile. Les incorporations des mots sont en fait une classe de techniques où les mots individuels sont représentés en tant que vecteurs à valeur réelle dans un espace vectoriel prédéfini. Chaque mot est mappé à un vecteur et les valeurs vectorielles sont apprises d'une manière qui ressemble à un réseau neuronal, et donc la technique est souvent localisée dans le domaine du deep learning.

La représentation distribuée est acquise en fonction de l'utilisation des mots. Cela permet aux mots qui sont utilisés de façon similaire de se traduire par des représentations similaires, capturant naturellement leur signification. Cela peut être contrasté avec une représentation fragile dans un modèle de sac de mots où, à moins de gestion explicite, des mots différents ont des représentations différentes, indépendamment de la façon dont ils sont utilisés. Nous distinguons deux modèles neuronaux: CBOW et Skip-Gram [19]

3.3.1. Le modèle Continuous Bag-of-Words(CBOW)

L'architecture du modèle CBOW tente de prédire le mot cible actuel (le mot central) en fonction des mots de contexte source (mots environnants).D'autre façon La façon dont CBOW fonctionne, c'est qu'elle tend à prédire la probabilité d'un mot donné un contexte. Un contexte peut être un mot unique ou un groupe de mots La figure II-5 montre l'architecture de ce modèle. Le but du réseau de neurones CBOW est de maximiser l'équation suivante [20] :

$$\frac{1}{V} \sum_{t=1}^V \log p(m_t | m_{t-\frac{c}{2}} \dots m_{t+\frac{c}{2}})$$

Où V (et C) est la taille du vocabulaire (contexte).

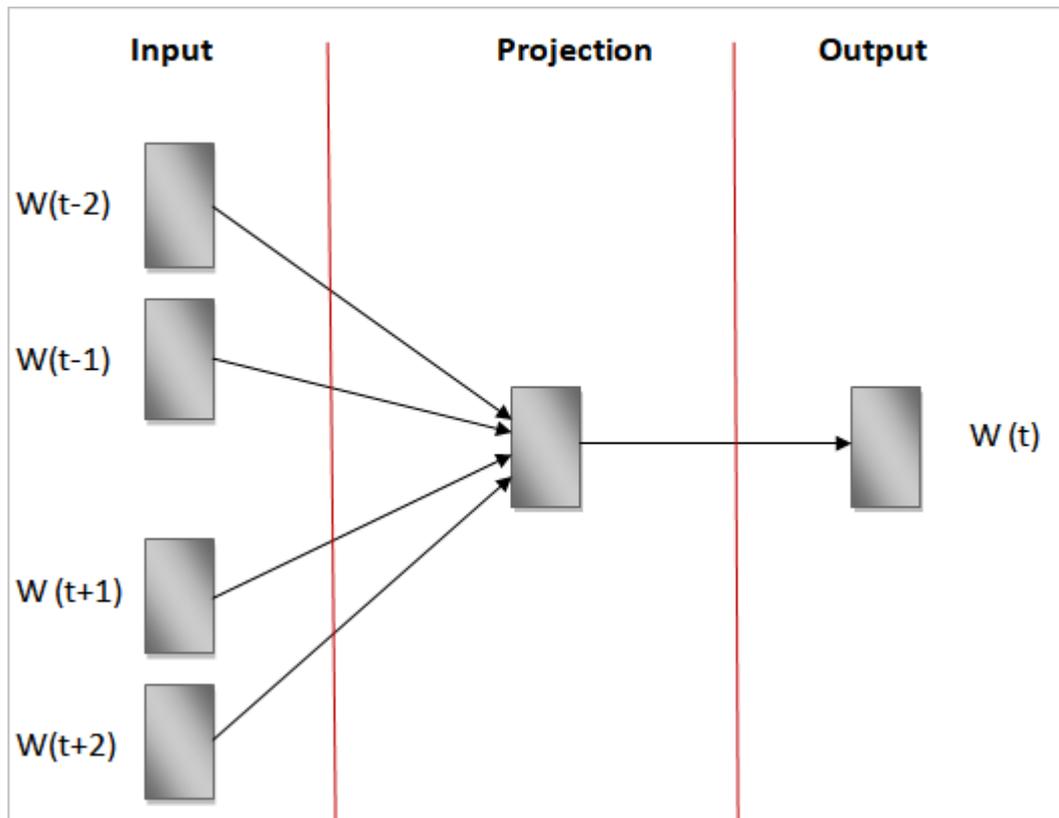


Figure II-5: L'architecture de modèle CBOW

3.3.2. Le modèle Continuous Skip-Gram

L'objectif de formation du modèle Skip-Gram est de trouver des représentations de mots qui sont utiles pour prédire les mots environnants dans une phrase ou un document plus formellement. La figure II-6 montre l'architecture de ce modèle. Le but du réseau de neurones SkipGram est de maximiser l'équation suivante [20] :

$$\frac{1}{V} \sum_{t=1}^V \sum_{j=t-c, j \neq t}^{t+c} \log p(m_j \setminus m_t)$$

Où V (et C) est la taille du vocabulaire (contexte).

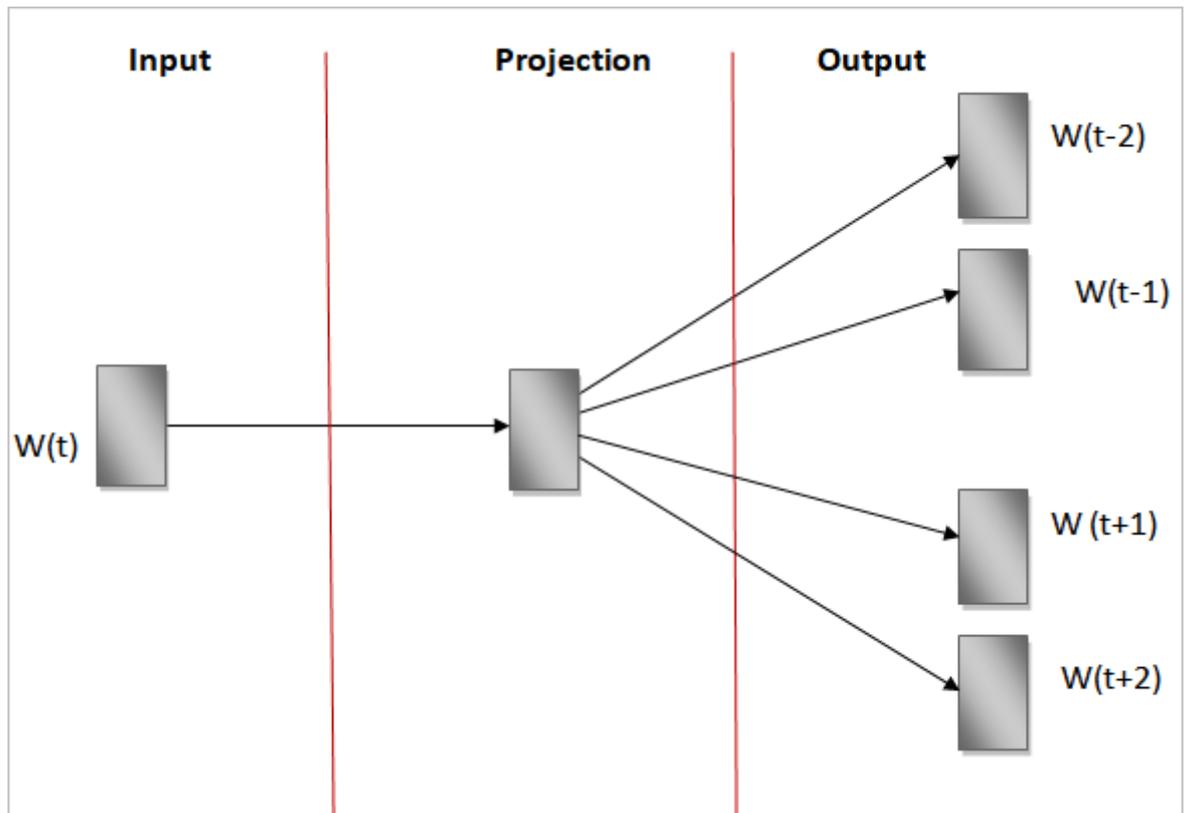


Figure II-6: L'architecture de modèle SkipGram

4. Les travaux connexes sur l'évaluation automatique des réponses courtes

El Moatez Billah Nagoudi et al. [21] ont construit un système pour SemEval 2017 Task1: similarité sémantique textuelle (Track1). Le système nommé LIM-LIG propose un modèle d'intégration de mots consacrés à la mesure de la similarité sémantique dans les phrases arabe, L'idée principale est d'exploiter le mot représentations en tant que vecteurs dans un espace multidimensionnel pour capturer la sémantique et les propriétés syntaxiques des mots. Les pondérations IDF et Part-of-Speech tagging sont appliquées sur les phrases examinées à soutenir l'identification des mots qui sont hautement descriptifs dans chaque phrase. Le système LIM-LIG obtient une corrélation de Pearsons de 0.746, qui lui a permis de se classer en deuxième place parmi tous les participants à la tâche STS des couples monolingues arabes organisée dans le cadre de la campagne d'évaluation SemEval 2017.

L'approche LInSTSS proposée par Bojan Furlan et al. [22] présente une méthodologie pour la création d'un système capable à déterminer la similarité

sémantique entre deux textes courts donnés. Cette approche est particulièrement appropriée pour une application dans des situations où aucune grande ressource linguistique électronique accessible au public ne peut être trouvée pour la langue désirée ce qui rend LInSTSS indépendante de la langue. LInSTSS est adaptée à la langue serbe afin d'évaluer et analyser les résultats obtenues en appliquant cette approche. Le système a marqué 76.6 d'exactitude et 82.93 de précision, qui sont jugées meilleures en les comparant avec d'autres méthodes STS.

Mihalcea et al. [23] ont proposé une méthode pour mesurer la similarité sémantique de deux textes courts (phrases ou paragraphes) en utilisant des mesures basées sur le corpus et fondées sur la connaissance de la similarité des mots. Pour chaque mot du texte, cette méthode identifie la meilleure correspondance du texte opposé et l'inclut ensuite dans la mesure globale de la similarité sémantique. Cette approche donne un score de mesure F élevé, mais elle est exigeante en termes de calcul et nécessite l'utilisation d'un modèle de données de mots. Islam et Inkpen [24] ont amélioré la mesure de similarité en combinant un algorithme d'appariement de chaînes modifié avec une mesure basée sur le corpus de similarité sémantique. La similarité sémantique joue également un rôle important dans la reconnaissance de l'implication textuelle (RTE pour Recognizing Textual Entailment) et partage de nombreuses caractéristiques avec celle-ci. [21]

Mohamed A. Zahran et al. [25] ont comparé les différentes techniques pour construire des représentations spatiales vectorisées pour l'arabe et tester par la suite ces modèles via des évaluations intrinsèques et extrinsèques. L'évaluation intrinsèque évalue la qualité des modèles à l'aide d'un ensemble de données sémantiques et syntaxiques de référence, tandis que l'évaluation extrinsèque évalue la qualité des modèles en fonction de leur impact sur deux applications de traitement du langage naturel: recherche d'information et notation à réponse courte. Enfin, les auteurs cartographient l'espace vectoriel arabe avec l'homologue anglais en utilisant le réseau neuronal de régression d'erreur Cosinus et montrent qu'il surpasse les réseaux neuronaux de régression d'erreur quadratique moyenne standard dans cette tâche.

Notre travail est connexe aux travaux menés par Gomaa & al. [26]. Les auteurs ont utilisé des mesures de similarité syntaxiques et des mesures basées sur le corpus pour développer leur système de notation à réponse courte. Ils ont testé les mesures

sur le dataset (GOMAA dataset) qu'ils ont construit eux-mêmes. Leurs résultats ont montré que les meilleures valeurs de corrélation obtenues en utilisant des mesures syntaxiques ont été obtenues en utilisant respectivement les approches de distance de n-gramme et de distance de Manhattan. Dans la deuxième étape, ils ont mesuré la similarité en utilisant des mesures de similarité basées sur le corpus [27]: DISCO1 (Calcule la similarité du premier ordre entre deux mots basés sur leurs ensembles de collocation) et DISCO2 (Calcule la similarité du second ordre entre deux mots basés sur leurs ensembles de distribution des mots similaires). Les résultats ont montré que DISCO1 atteint des valeurs de corrélation plus efficaces. Dans la troisième étape, la similarité a été évaluée en combinant des mesures basées sur la syntaxe et le corpus. La meilleure valeur de corrélation a été obtenue en mélangeant n-gramme avec les techniques de similarité DISCO1.

Le travail de [28] et [29] sont tous aussi intéressants en considérant leurs résultats par rapport à une approche basée sur le calcul vectoriel et les word Embedding. [28] ont évalué leur approche sur Gomaa dataset alors que [29] a obtenu le 2ème meilleur score du SemEval 2017 d'où l'intérêt que nous portons pour ces travaux utilisant les mêmes datasets que nous.

La génération des réponses modèles consiste à générer des phrases qui sont similaires à cette réponse.

Concernant la notion de paraphrase nous citons les travaux suivants :

4.1. Les paraphrases :

Dans un sens général, paraphraser signifie reformuler. Il s'agit d'une opération qui consiste à changer les mots (la forme, la syntaxe) d'un texte, tout en préservant sa signification (la sémantique). En effet, il est possible d'exprimer une idée de différentes manières en fonction de l'objectif du locuteur. Egalement, nous dirons que pour deux phrases qu'ont le même sens P1 et P2, l'une est la paraphrase de l'autre et vice versa, puisque la relation de paraphrase est symétrique.

D'après « Madnani and Dorr » [30] Le concept de paraphraser est généralement défini sur la base du principe de l'équivalence sémantique, c'est une forme de surface alternative dans la même langue exprimant le même contenu sémantique que la forme originale, Les éléments lexicaux individuels ayant la même

signification sont généralement appelés paraphrases lexicales ou synonymes. Cependant, les paraphrases lexicales ne peuvent être strictement restreintes au concept de synonyme, Il existe plusieurs autres formes telles que l'hyperonyme, où l'un des mots dans la relation paraphrastique est soit plus général ou plus spécifique que l'autre.

4.2. La génération des paraphrases :

La génération de paraphrase est une tâche importante dans le TAL, qui peut être une technologie clé dans de nombreuses applications telles que la génération des réponses modèles, le paraphrasant de réponse est également utile lors de la cartographie des réponses des utilisateurs. Cependant, en raison de la complexité du langage naturel, générer automatiquement des paraphrases exactes et variées est encore très difficile. Les générateurs des paraphrases on une expression (ou un modèle) de langue unique comme entrée, et ils sont tenus de produire autant d'expressions de langage de sortie (ou de modèles) que possible, de sorte que les expressions de sortie sont des paraphrases ou elles constituent, avec l'entrée, des paire phrases correctes. Cette approche est similaire à l'utilisation de représentations de sens indépendantes de la langue.

On site quelque approches :

4.3. La traduction automatique statistique :

La traduction automatique statistique (SMT) permet d'émettre des hypothèses de traduction dans une langue cible à partir d'une phrase dans une langue source. Cette approche est basée sur l'observation d'exemples de traductions, les corpus parallèles, permettant l'apprentissage automatique de correspondances bilingues. Les corpus parallèles bilingues (exemple un corpus pour la langue arabe et l'autre pour la langue française) (Koehn et al.)[31] a étendu la notion d'alignements entre segments de taille quelconque .Des techniques d'alignement permettent d'apprendre des probabilités de traduction de segments. Brown et al. [32] proposent un modèle probabiliste En utilisant ces modèles et à partir d'un corpus parallèle annoté, il est possible d'apprendre des relations statistiques entre deux langues données.

4.4. Méthodologie de génération typique (logique des prédicats)

L'entrée de ce processus est un prédicat (argument hiérarchique) et à la sortie devrait être un ensemble de phrases grammaticales dont la signification correspond à l'entrée.

4.5. Génération à l'aide des réseaux de neurones récurrents

Les réseaux récurrents (ou RNN pour Recurrent Neural Networks) sont des réseaux de neurones dans lesquels l'information peut se propager dans les deux sens, y compris des couches profondes aux premières couches. En cela, ils sont plus proches du vrai fonctionnement du système nerveux, qui n'est pas à sens unique. Les modèles RNN ont la capacité d'apprendre des structures de texte en formant un jeu de données à l'entrée, et ensuite pour produire (générer) un texte acceptable (plus ou moins) dans la sortie, autrement le processus d'apprentissage est basé sur la signification des mots (ou des unités de texte, en arabe une unité de texte peut être une lettre, un mot ou une phrase). Les RNN ne sont capables de mémoriser que le passé dit proche, et commencent à « oublier » au bout d'une cinquantaine d'itérations environ. Et ce n'est que récemment que des méthodes efficaces ont été mises au point comme les LSTM (Long Short Term Memory). Ces réseaux à large « mémoire court-terme » LSM peut apprendre les séquences d'un texte donné, puis génère de nouvelles séquences possibles en faisant des prédictions.

4.6. Les travaux liés au même Projet entamés durant l'année 2017/2018

Notre travail rentre dans le cadre de la continuité des travaux entamés l'année passée et menés dans l'objectif d'élaboration de méthodologies et d'outils pour l'évaluation automatique des réponses courtes destinée à la langue arabe.

En effet,

[34] présente une synthèse expérimentale des différentes approches de mesures de similarité syntactique appliquées à des ensembles de données (Data Sets) exprimées dans la langue arabe.

[16] présente une synthèse expérimentale de plusieurs approches de mesures de similarité sémantique appliquée à des ensembles de données (Data Sets) exprimées dans la langue arabe.

[35] donne une synthèse expérimentale de l'utilisation des Word Embedding appliquée à des ensembles de données (DataSets) exprimés dans la langue arabe.

A synthèse a permis d'un côté, de retenir les meilleures approches en vue d'une hybridation de mesures syntaxiques, sémantiques et utilisant les Word Embedding dans le système d'évaluation automatique et d'un autre côté, l'élaboration de plusieurs approches que nous reprenons dans ce travail sans redéveloppement des codes associés.

Conclusion

Dans ce chapitre nous avons présenté les approches développées au paravent sur l'évaluation automatique des réponses courtes ainsi la génération des paraphrases nous tentons dans la suite de notre travail d'améliorer nos résultats par apport au approche déjà faite.

chapitre III. Système de génération automatique des réponses modèles dans un système ASAG

Ce chapitre traite le noyau de notre travail qui consiste en l'approche suivie pour évaluer l'impact des synonymes(générés par l'espace sémantique et les word embedding) sur l'évaluation des systèmes ASAG en utilisant les mesures de similarité sémantique avec la combinaison des mesures syntaxiques .

1. Méthodologie

Il y a peu de tentatives de recherche dans le domaine de l'évaluation automatique en langue arabe en raison de sa nature complexe et le manque de ressources, notre approche va évaluer automatiquement les réponses des étudiants et générer un dictionnaire en langue arabe ainsi qu'un corpus de toutes les réponses possibles.

Les réponses courtes, celles de l'étudiant et celles de l'enseignant à évaluer, vont être traités comme des entrées lesquelles vont subir à leur tour un prétraitement et un choix à faire entre l'approche d'espace sémantique et WE par la suite une génération du dictionnaire des synonymes sera faite qui jouera un rôle dans la notion de calcul de similarité entre la réponse modèle(RM) et la réponse d'étudiant(RE), une note sera définie en conséquence Notre travail permettra de :

- générer un dictionnaire des synonymes en appliquant la notion des WE et la notion d'espace sémantique (une explication par la suite sera présentée)
- l'évaluation des réponses des étudiants en utilisant les synonymes : la génération des synonymes va jouer un rôle dans la notion de calcul de similarité entre RM et RE.

Dans ce qui suit, nous allons résumer les grandes lignes (visualisées au niveau de la Figure III-1) par les points suivants :

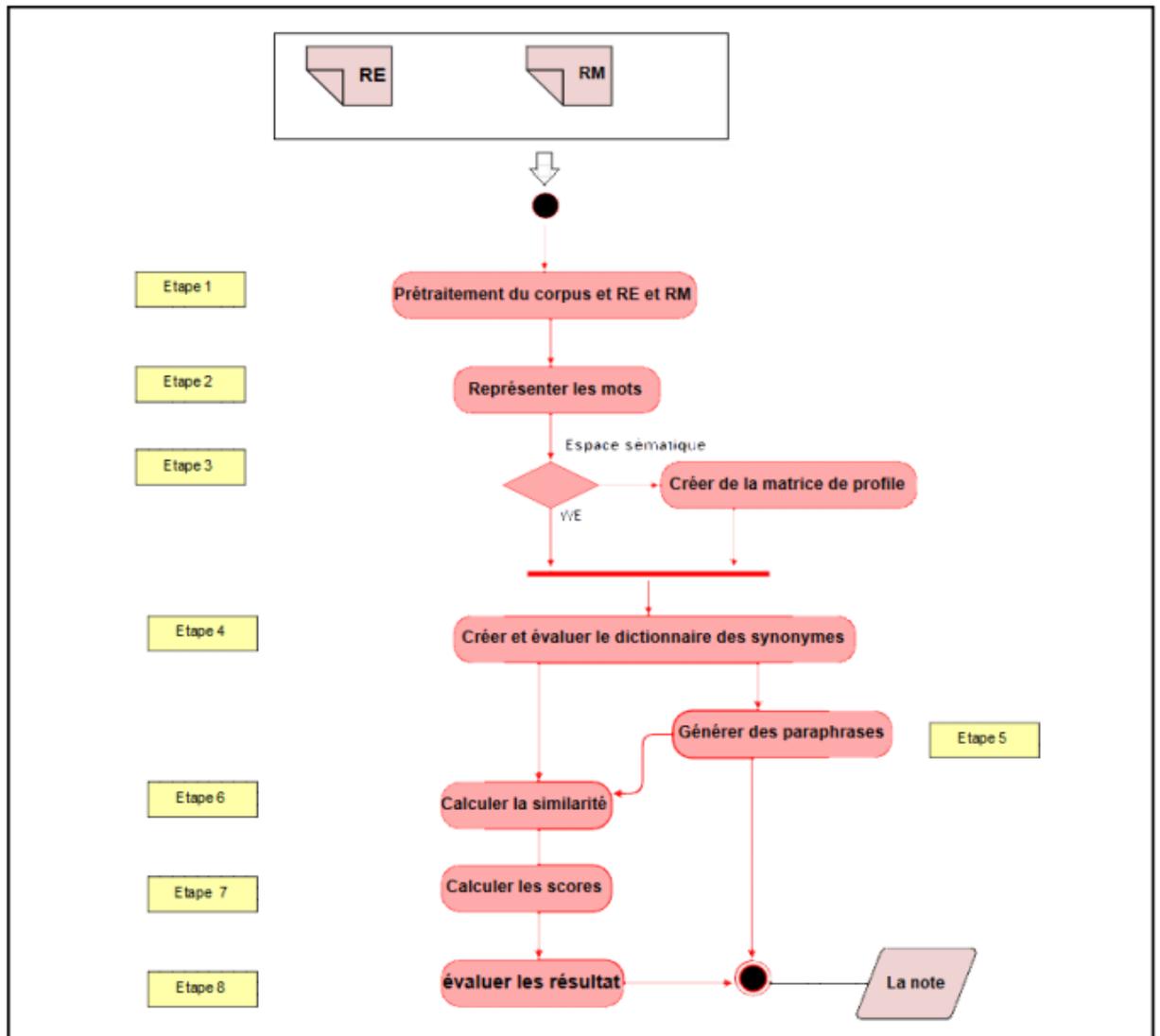


Figure III-2: schéma de représentation des étapes de notre système

Etape 1 : Le prétraitement de la réponse modèle et étudiant

Les deux réponses vont être normalisés et stemmés soit avec tashaphyne stemer, stanford-corenlp pour le stem Legé ou khoja stem pour le stem lourd

Etape 2 : La représentation des mots

La représentation des mots est sous forme des vecteurs sémantique dans un espace sémantique ou sous forme des vecteurs dans un modèle WE

Etape 3 :

Création de la matrice de profil

Après obtention de l'espace sémantique nous construisons la matrice de profils, les mots vont être représentés sous forme des vecteurs de similarité (chaque mots a une similarité par rapport à un autre mot)

L'acquisition des Word Embeddings

Nous avons évaluées l'impact des Word Embeddings sur le processus de génération des réponses modèles pour l'évaluation en langue arabe, nous avons utilisé les deux seuls modèles de WE déjà générés pour les mots de la langue arabe, les WE de Zahran décrit dans : [35] et disponible sur le web [36]

Et les WE d'aravec décrit dans : [37] et disponible sur le web [38]

Etape 4 : Création du dictionnaire

De la matrice de profil obtenu ou le modèle des WE, nous avons fait ressortir des vecteurs pour 10 mots les plus similaires qui doivent être classés et ordonnés, au fur et à mesure le dictionnaire des synonymes va être construit.

L'évaluation des synonymes se fait par un expert (l'enseignant).

Etape 5 : La génération des paraphrases

Nous allons combiner tous les mots qui sont dans notre dictionnaire pour générer des paraphrases

Etape 6 : Calcul de similarité

Nous avons deux modèles pour le calcul de similarité sémantique qui sont :

Le modèle somme-vecteurs avec et sans pondération,

Le modèle calcul-matriciel sans et avec pondération

Etape 7 : Passage au score

Après avoir calculé la similarité, nous ferons une classification des similarités en utilisant le Classifieur non supervisé K-Means pour obtenir les scores

Etape 8 : Evaluation de scores

Calculer le pourcentage de Pearson et l'erreur quadratique pour un jeu de test complet.

Nous présentons les différentes phases de notre approche, tout d'abord notre réponse modèle et celle de l'étudiant vont subir un prétraitement.

1.1. La normalisation

Il s'agit de prétraiter le texte des documents textuels (corpus, réponse model, réponse étudiant) :

- Supprimer toutes les lettres d'une autre langue
- Supprimer les chiffres et la ponctuation
- Elimination des stopwords
- Elimination du tashkeel
- Normalisation des caractères (آ (ا - ا - ا - ا) et (ة) à (ه) et (ي) à (ي)

Exemple :

La phrase avant la normalisation :

P= جرائم النصب والاحتيايل تصنف في المجموعة الأولى التي تقع على الانترنت أي أن الشبكة تكون محلا لها

Après normalisation :

P = جرائم النصب الاحتيايل تصنف المجموعه الاولي تقع الانترنت الشبكة محلا

1.2. Le stemming

Lemmatisation, Stemming qui consiste à réduire le mot à sa racine ou bien l'élimination des suffixes, préfixes..., nous avons utilisé :

Le stem léger (Light Stemming) : Dans ce type le stemming s'applique sur la suppression des préfixes et des suffixes, sans tenter d'identifier la racine du mot.

Le stem lourd (Root-Stemming) : Ce type consiste à supprimer les préfixes et les suffixes bien connus pour extraire la racine de base d'un mot et à identifier le motif en correspondance avec le mot restant.

Dans notre cas nous avons testé 3 stems

- Stemmer StanfordCoreNLP³
- Tahapyne Stemmer⁴
- Khoja Stemmer⁵

Sur les 61 réponses modèles du dataset Gomaa et 48 réponses modèle du data set cybercriminalité . Le tableau en dessous montre les différents stemmers sur la même phrase

Exemple :

Avant stemming :

P جرائم النصب الاحتيال تصنف المجموعه الاولى تقع الانترنت الشبكه محلا =

Après Stemming :

³ <https://stanfordnlp.github.io/CoreNLP/> (Dernier accès le 23/06/2018)

⁴ <https://pypi.org/project/Tashaphyne/>

⁵ <https://github.com/motazsaad/khoja-stemmer-command-line/blob/master/khoja/KhojaStem.java>

les mots	Stanford_ Core_NLP	Tahapyne	Khoja
جرائم	جرائم	جرائم	جرم
النصب	النصب	نصب	نصب
الاحتيال	الاحتيال	حتيال	حول
تصنف	تصنف	صنف	صنف
المجموعة	المجموع	جموع	جمع
الأولى	الاولي	ول	ولي
تقع	تقع	قع	قعي
الانترنت	الانترنت	ترن	رني
الشبكة	الشبكة	شبيك	شبيك
محلا	محلا	حل	حلي

Tableau 1: Exemple de stemmer

On remarque que lemmatisation en utilisant khoja et tashapyne change le sens de quelque mot dû aux les stemmers qui existent pour la langue arabe ne présentent pas de documentation disponible et ne présentent pas une évaluation de la précision des résultats obtenus.

Après le prétraitement de nos réponses nous passerons à la représentation des mots et la génération des synonymes.

2. Construction du dictionnaire des synonymes

Cette partie présente en détail la deuxième ,troisième et la quatrième étape de la «Figure III-1».

Pour la construction de notre dictionnaire il faudra passer par l'approche espace sémantique (figure III-2) ou bien l'approche WE

2.1. Dictionnaire basé sur l'espace sémantique

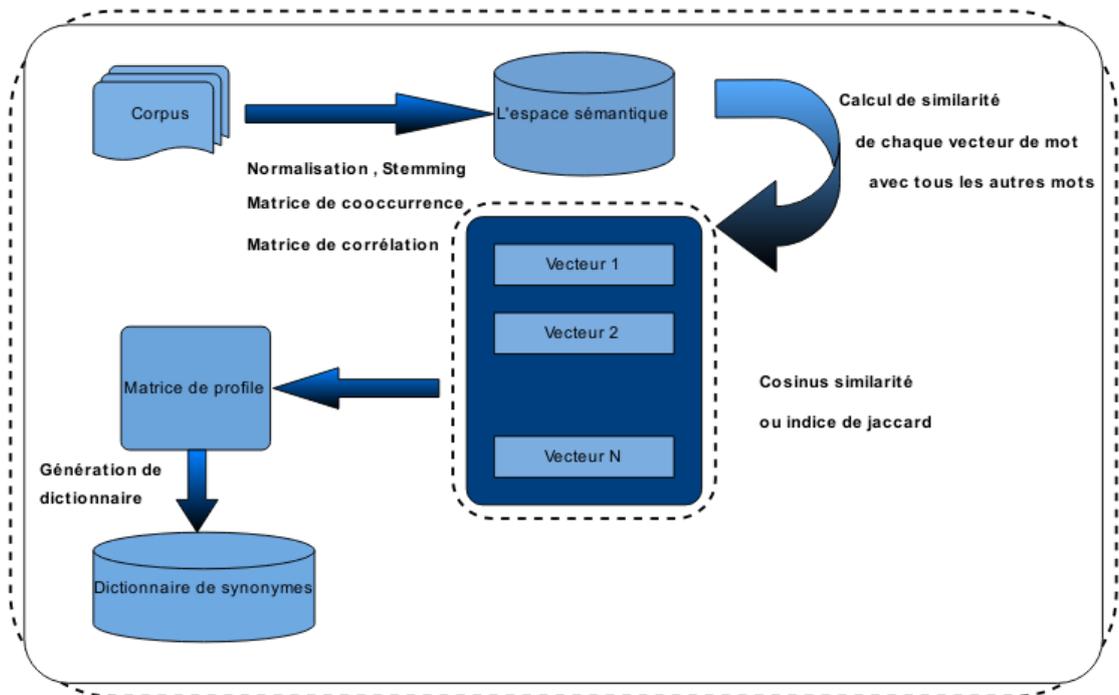


Figure III-3: phase de création du dictionnaire avec l'espace sémantique

2.1.1. Construction de l'espace sémantique

Dans la « Figure III-3» les différentes phases de création de l'espace sémantique sont présentées. Tout d'abord, l'acquisition et l'analyse du corpus afin de choisir celui qui convient. Ensuite, la phase du prétraitement qui consiste à préparer le corpus pour tout usage. La phase du traitement du corpus comprend l'application de l'une des approches de similarité dans le but de construction de l'espace sémantique. Enfin, le post traitement du corpus qui prend en considération l'importance des mots dans le corpus[16].

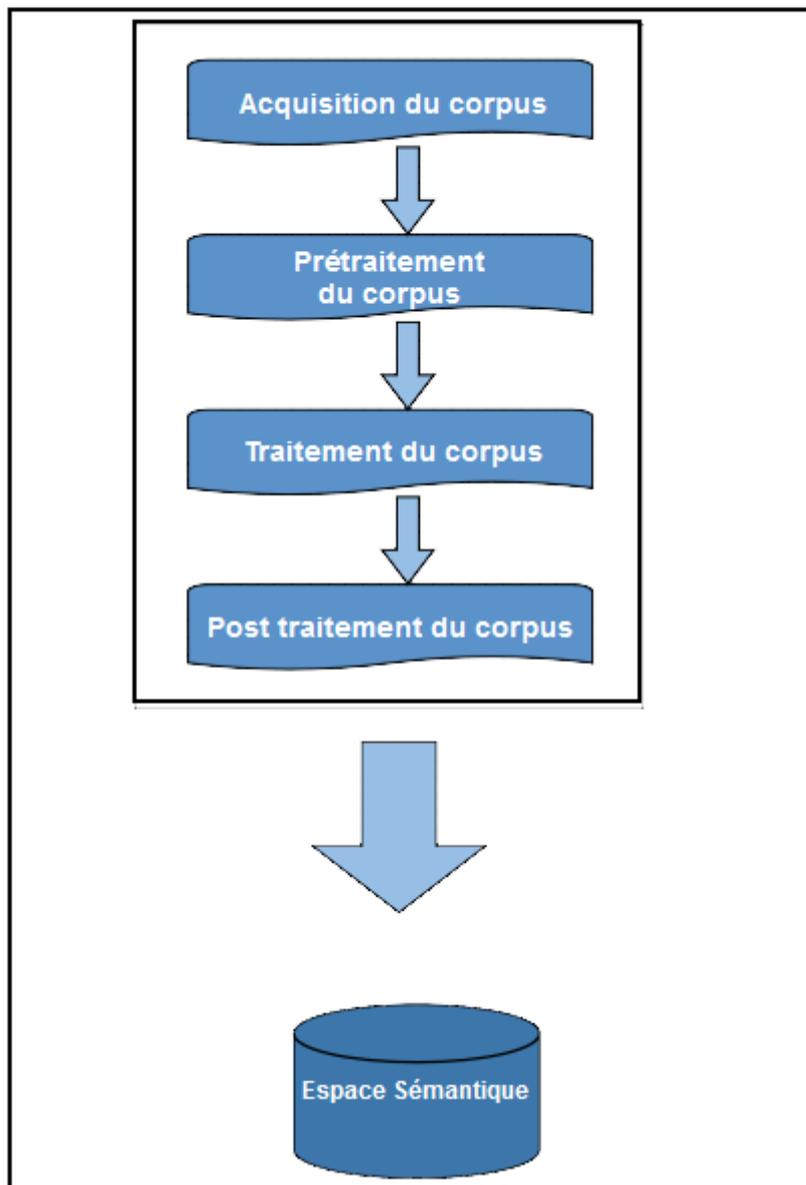


Figure III-4: Phase de création de l'espace sémantique

PHASE 1: Acquisition du corpus

Analyse et traitement des corpus disponibles (CNN⁶, CNN + BBC⁷, Khaleej⁸, Cybercriminalité⁹)

⁶ <https://sourceforge.net/projects/ar-text-mining/files/Arabic-Corpora/>

⁷ <https://sourceforge.net/projects/ar-text-mining/files/Arabic-Corpora/>

⁸ <https://sourceforge.net/projects/arabiccorpus/files/khaleej-2004corpus/>

⁹ https://drive.google.com/drive/folders/13DoT0HqDwrDnwcwTFucJsceM9rfK197C?fbclid=IwAR0wizvduK8PGNy_8mRdFjAJF8Vc7Z9O9aHplhN6uXYbmCfmOSLRPRzyAYk

PHASE 2: Prétraitement du corpus

Pour cette tâche, le travail doit être fait sur les corpus utilisés, le prétraitement du corpus se compose de deux autres sous-tâches : la normalisation et le stemming.

PHASE 3: Traitement du corpus

Dans cette démarche, l'approche de similarité statistique « Corpus-Based » est adoptée à cause du manque de ressources arabe comme les dictionnaires et les lexicons dont l'approche topologique « knowledge-based » impose. Le modèle BOW est adopté (il permet de compter tous les mots dans un texte., il crée une matrice d'occurrences pour une phrase ou un document, sans tenir compte l'ordre des mots) . L'approche se base sur le concept disant que les mots qui sont sémantiquement liés se trouvent dans le même contexte. Pour cela cette approche repose sur la notion du voisinage (mots voisins). Cependant la présence d'un corpus qui sera manipulé à l'aide du modèle BOW est indispensable. En outre, une fenêtre avec une taille prédéfinie est importante pour concrétiser la notion du voisinage. La figure III-5 illustre la représentation du corpus par le modèle Bag of words.

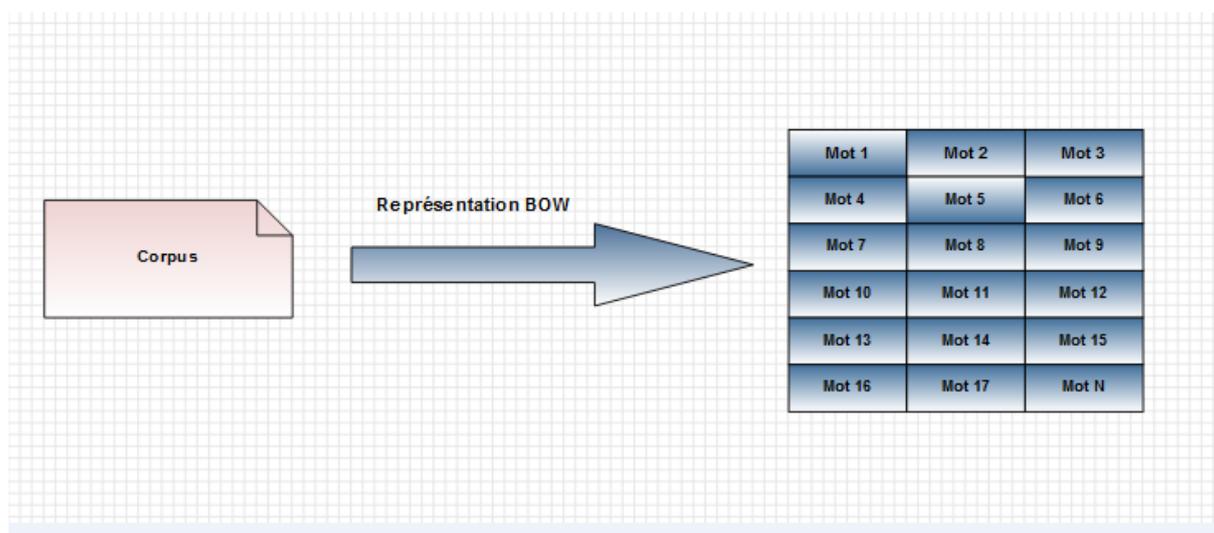


Figure III-5: la représentation du corpus par le modèle Bag of words

a) Les corpus sont acquis et prétraités selon l'étape du « Prétraitement du corpus ».

b) Construction de la matrice des cooccurrences : Ou matrice de fréquences/poids. Cette étape signifie la transformation des données textuelles au

sein du corpus en données numériques sous forme d'une matrice. Chaque case représente la somme des poids de l'apparence du $i^{ème}$ terme (mot) avec le $j^{ème}$ terme. Cette somme est calculée par rapport à l'emplacement des termes du corpus autour du $i^{ème}$ terme (effectuer le même calcul pour chaque occurrence du $i^{ème}$ terme).

c) Construction de la matrice des corrélations : Après avoir construit la matrice des cooccurrences, une stratégie de normalisation est appliquée. Donc une nouvelle matrice est construite en appliquant la formule ci-dessous prise de l'algorithme COALS pour chaque case/élément de la matrice des cooccurrences.

$$\left\{ \begin{array}{l} W'_{a,b} = \frac{Tw_{a,b} - \sum_j w_{a,j} \cdot \sum_i w_{i,b}}{(\sum_j w_{a,j} \cdot (T - \sum_j w_{a,j}) \cdot \sum_i w_{i,b} (T - \sum_i w_{i,b}))^{\frac{1}{2}}} \\ T = \sum_i \sum_j w_{i,j} \end{array} \right.$$

Où :

- a et b sont les deux termes de la matrice de cooccurrences (ligne et colonne).
- $w_{a,b}$ st l'élément de la matrice de cooccurrences du terme a et b.
- i est l'indice des lignes tandis que j est celui des colonnes.
- $\sum_j w_{a,j}$ est la somme des colonnes de la ligne du terme a.
- $\sum_j w_{i,b}$ est la somme des lignes de la colonne du terme b.
- $T = \sum_i \sum_j w_{i,j}$ est la somme de tous les éléments de la matrice de cooccurrences.

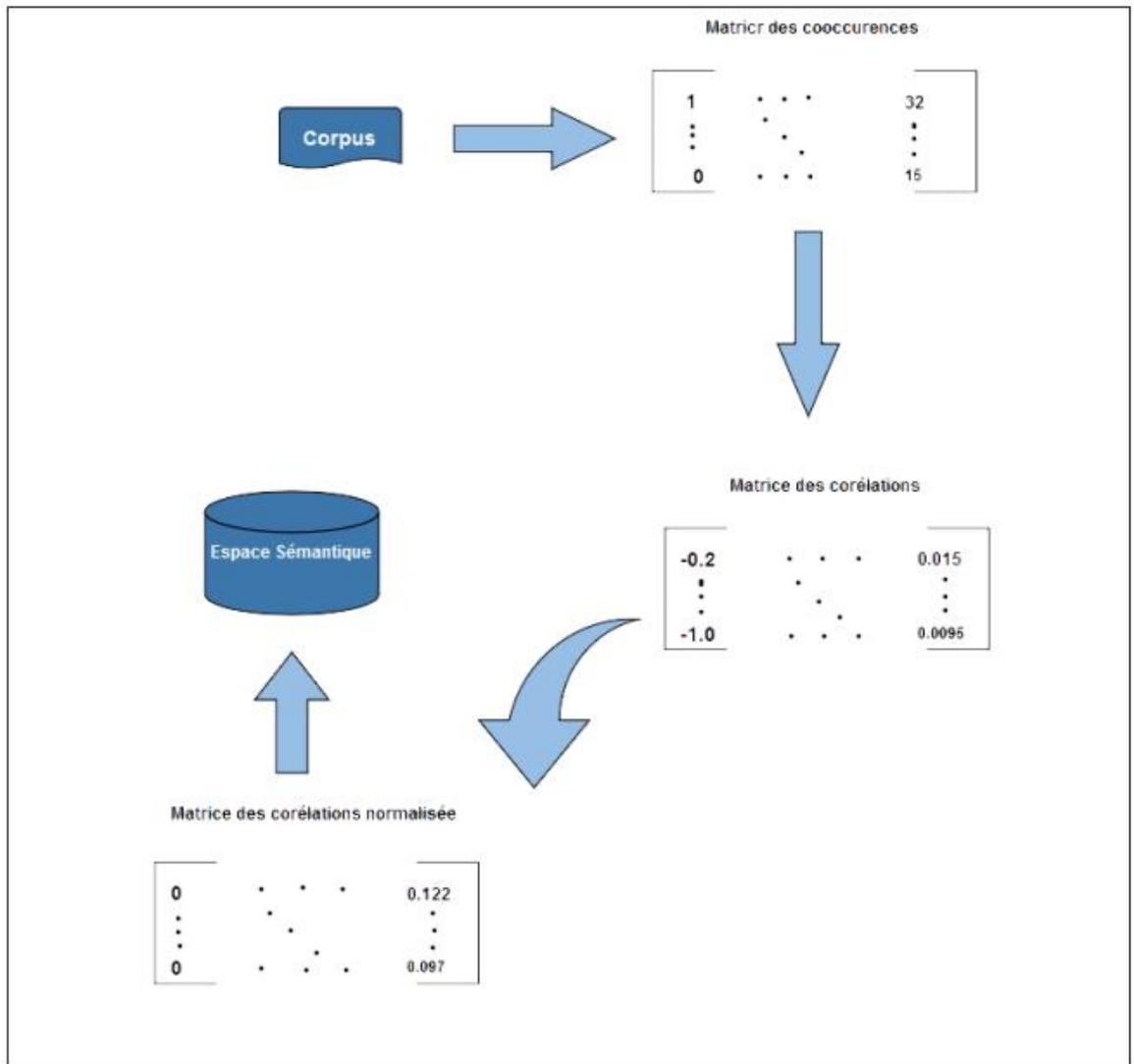


Figure III-6: Exemple de traitement d'un corpus

d) Construction de la matrice de corrélation normalisée :

Les corrélations négatives transportent très peu d'information, alors, nous effectuons encore une autre normalisation. Les valeurs négatives sont normalisées à 0.

Tandis que les valeurs positives prennent leurs racine carré afin d'amplifier l'importance des nombreuses petites valeurs par rapport aux grandes valeurs [17].

PHASE 4 : Post traitement du corpus

En effet, la similarité sémantique est fondée sur l'idée disant que les termes similaires se trouvent fréquemment dans le même contexte. Néanmoins, dans un

contexte, les termes n'ont pas tous la même importance. Par conséquent, il est plus judicieux de pondérer l'impact des termes avant l'application d'une mesure de similarité. La fréquence d'apparition d'un terme dans un document est un bon indicateur de l'importance de ce terme.

Une fonction de pondération attribue à chaque terme de chaque document une valeur. Cette valeur (ou poids) est calculée en tenant compte de deux grands critères [39]:

La force (capacité) locale du terme dans le document (c'est-à-dire mesurer l'importance du terme dans le document dans lequel il paraît).

La force globale (c'est-à-dire mesurer l'importance du terme dans tout le corpus). Plus un terme est présent (fréquent) dans un document, plus sa force locale est importante et plus ce terme est présent dans le corpus, plus sa force globale est élevée.

Une fois l'espace sémantique est généré, illustré en la figure III-2, nous calculerons la similarité pour chaque vecteur de mot dans la colonne avec chaque vecteur de mot dans la ligne en utilisant :

Similarité cosinus : qui a été déjà présentée dans l'état de l'art

Prenons un petit espace sémantique pour mieux comprendre :

	برنامج	المعلومات	المخترق	المستخدم	البرامج	الاجهز	انشاء	مواقع	اغراق	النظام
برنامج	0.000	0.582	0.274	0.000	0.000	0.000	0.000	0.000	0.000	0.000
المعلومات	0.582	0.322	0.423	0.270	0.000	0.000	0.000	0.000	0.000	0.000
المخترق	0.274	0.423	0.000	0.330	0.226	0.000	0.000	0.000	0.000	0.000
المستخدم	0.000	0.270	0.330	0.000	0.316	0.210	0.000	0.000	0.000	0.000
البرامج	0.000	0.000	0.226	0.316	0.000	0.316	0.226	0.084	0.000	0.000
الاجهز	0.000	0.000	0.000	0.210	0.316	0.000	0.330	0.259	0.170	0.000
انشاء	0.000	0.000	0.000	0.000	0.226	0.330	0.000	0.369	0.321	0.274
مواقع	0.000	0.000	0.000	0.000	0.084	0.259	0.369	0.000	0.436	0.412
اغراق	0.000	0.000	0.000	0.000	0.000	0.170	0.321	0.436	0.000	0.540
النظام	0.000	0.000	0.000	0.000	0.000	0.000	0.274	0.412	0.540	0.000

Figure III-7: Un espace sémantique

Les lignes et les colonnes de la figure III-7 (espace sémantique) représentent tous les mots qui se trouve dans le corpus choisis .Chaque mot dans notre espace possède un vecteur sémantique. Le tableau en dessous, représente chaque mot de l'espace sémantique avec son vecteur sémantique .Nous prenons chaque vecteur de mot afin de calculer la similarité avec tous les autres mots.

v1- [0. 0.582 0.274 0. 0. 0. 0. 0. 0. 0. 0.]	برنامج
v2-[0.582 0.322 0.423 0.27 0. 0. 0. 0. 0. 0.]	المعلومات
v3- [0.274 0.423 0. 0.33 0.226 0.0.0.0.0.0.]	المخترق
v4- [0. 0.27 0.33 0. 0.316 0.21 0. 0. 0. 0.]	المستخدم
v5- [0. 0. 0.226 0.316 0. 0.316 0.226 0.084 0. 0.]	البرامج

Prenons le même vecteur 1 (v1) et nous calculerons sa similarité cosinus

v1= [0. 0.582 0.274 0. 0. 0. 0. 0. 0. 0.]

On remplace les 0 par des False et les autres nombres par des True

v1= [[False True True False False False False False False]]

v1= [[False True True False False False False False False]]

V1 intersect v1=2

V1 union v1 =2

sim_jacc=2÷2=1

Exemple 2:

Nous prenons le vecteur v1 du mot (برنامج) avec le vecteur v2 du mot (المعلومات) et nous calculons la similarité

v1= [0. 0.582 0.274 0. 0. 0. 0. 0. 0. 0.]

v2= [0.582 0.322 0.423 0.27 0. 0. 0. 0. 0.]

v1= [[False True True False False False False False False]]

v2= [[True True True True False False False False False]]

V1 intersect v2= $V1 \cap V2$ = [[False True True False False False False False False False]]

V1 union v2= $V1 \cup V2$ = [[True True True True False False False False False False]]

sim_jacc= 2÷4 =0.5

Nous appliquons le même processus sur le reste des mots de ligne et colonne nous obtiendrons la matrice de profile dans la figure III 9:

	برنامج	المعلومات	المخترق	المستخدم	البرامج	الاجهز	الشاء	مواقع	اغراق	النظام
برنامج	1.000000	0.500000	0.200000	0.500000	0.166667	0.000000	0.000000	0.000000	0.000000	0.000000
المعلومات	0.500000	1.000000	0.600000	0.333333	0.285714	0.125000	0.000000	0.000000	0.000000	0.000000
المخترق	0.200000	0.600000	1.000000	0.333333	0.125000	0.285714	0.125000	0.125000	0.000000	0.000000
المستخدم	0.500000	0.333333	0.333333	1.000000	0.285714	0.125000	0.285714	0.285714	0.142857	0.000000
البرامج	0.166667	0.285714	0.125000	0.285714	1.000000	0.428571	0.250000	0.250000	0.500000	0.333333
الاجهز	0.000000	0.125000	0.285714	0.125000	0.428571	1.000000	0.428571	0.428571	0.285714	0.600000
الشاء	0.000000	0.000000	0.125000	0.285714	0.250000	0.428571	1.000000	0.666667	0.500000	0.333333
مواقع	0.000000	0.000000	0.125000	0.285714	0.250000	0.428571	0.666667	1.000000	0.500000	0.333333
اغراق	0.000000	0.000000	0.000000	0.142857	0.500000	0.285714	0.500000	0.500000	1.000000	0.400000
النظام	0.000000	0.000000	0.000000	0.000000	0.333333	0.600000	0.333333	0.333333	0.400000	1.000000

Figure III-9: une matrice de profils

Selon la matrice de profiles, nous ordonnons chaque vecteur de mots en prenant les 10 mots qui ont une grande similarité ce dernier sera sauvegardé dans un fichier « txt ». Chaque ligne de ce fichier représente un mot avec les dix 10 mots les plus similaires. Au moment, où l'enseignant saisit la RM (réponse modèle) chaque mot sera traité dans le dictionnaire qui va faire ressortir les dix (10) mots les plus similaires si le mot existe, sinon le mot sera ignoré. A la fin une évaluation sera faite par l'enseignant.

Exemple :

Le vecteur du mot :

0.059294 0.0516490.0026340.0000000.0628490.0297340.010484] الهجمات
[.....0.004303 0.0109690.0000000.000000...0.0756280.0856810.107280

Après le sort : [1. 0.67643638 0.562849 0.37566716 ... 0. 0. 0.]

On prend les 10 mots qui sont plus similaire

الهجمات , السبيرانه , الالكترونه , هجمات , التهديدات , ردع , الهجوم , لهجمات , شن , المخاطر , هجوم , الجرائم , الردع , سبيران

Après l'évaluation On prend les synonymes du mot الهجمات

الهجمات = [هجوم , الهجوم , المخاطر , شن , الجرائم]

2.2. Dictionnaire basé sur une représentation par Les word Embedding

Une fois la réponse modèle est stemé les 10 premiers mots les plus similaire de chaque mot de la RM seront calculés en appliquant les WE de zahran et les WE d'Aravec avec la représentation CBOW et SkipGram de chacun. Par la suite Le même processus sera répété pour tous les RM, au fur et à mesure le dictionnaire de mot va être construit

La figure III - 10 présente les étapes de la génération du dictionnaire en utilisant les WE

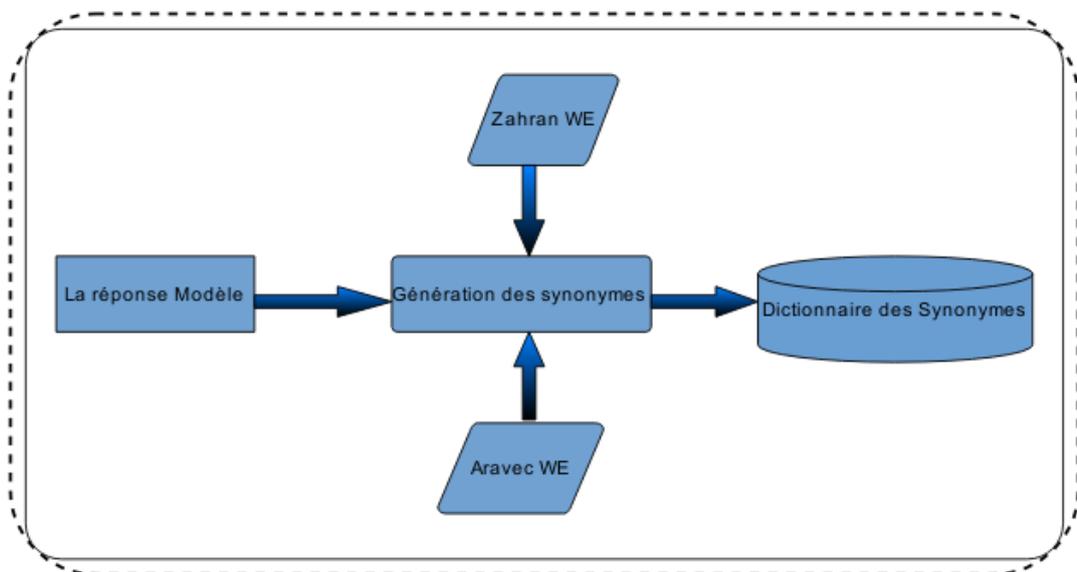


Figure III-10: phase de création du dictionnaire avec les WE

Exemple : dictionnaire pour chaque modèle de WE

D'après les résultats des tableaux 2, 3, 4, 5 nous remarquons une variété de synonymes obtenue de chaque modèles de WE, dans le cas ou le mot n'existe pas dans le modèle on n'obtient pas de synonymes.

Dans notre cas nous avons fixé le nombre de synonymes à 5 mais dans certains cas on peut trouver moins (exemple dans le tableau 5).

Aravec CBOW					
جرائم	بجرائم	جريمه	لجرائم	وجرائم	الجرائم
النصب	التذكاري	والنصب	التمثال	الصريح	تذكاري
الاحتيال	الغش	الابتزاز	التزوير	السرقه	السطو
تصنف	وتصنف	توصف	تندرج	صنفت	تتواجد
المجموع	مجموع	المعدل	المقياس	الاجمالي	للعدد
الاولي	الثانيه	الثالثه	الاخيره	الثانيه	الرابعه
تقع	وتقع	يقع	الواقعه	ويقع	والواقعه
الانترنت	الانترنت	الويب	انترنت	الفيديوك	الوب
الشبكه	الشبكات	للشبكة	الويب	الخرادم	شبكتها
محلا	سوقا	متجرا	مطعما	دكانا	مصنعا

Tableau 2: Exemple de génération des synonymes avec le modèle Aravec

CBOW

Aravec SkipGram					
جرائم	الجرائم	وجرائم	لجرائم	القتل	بجرائم
النصب	التذكاري	والنصب	تذكاري	نصب	للنصب
الاحتيال	الابتزاز	والاحتيال	والاختلاس	التزييف	والابتزاز
تصنف	وتصنف	تصنيفها	صنفت	مصنفة	فتصنف
المجموع	مجموع	والمجموع	الاجمالي	فالعدد	كمجموع
الاولي	الثانيه	الثالثه	والثانيه	الثانيه	الاولى
تقع	وتقع	الواقعه	تقع	يقع	فتقع
الانترنت	الانترنت	الشابكه	الويب	انترنت	ويوتيوب
الشبكه	الشبكات	شبكة	للشبكة	بالشبكة	الانترنت
محلا	سوقا	مطعما	متجرا	دكانا	مصنعا

Tableau 3: Exemple de génération des synonymes avec le modèle Aravec

SkipGram

Zahran CBOW					
جرائم	الجرائم	جريمه	لجرائم	بجرائم	وجرائم
النصب	والنصب	نصب	للنصب	بالنصب	الاحتيال
الاحتيال	عمليات الاحتيال	النصب والاحتيال	احتيال	والاحتيال	الخداع
تصنف	وتصنف	صنفت	تصنف_ضمن	تصنفها	يصنف
المجموع	مجموع	الصفحات_المقروءه	المحيط_بالتكليف	زوار_الامس	السجلات_عرض
الاولي	الاولى	الثانيه	الثالثه	الاولي_	الثالثه_والاخيره
تقع	وتقع	يقع	ويقع	الواقعه	فتقع
الانترنت	شبكة_الانترنت	عبر_الانترنت	انترنت	البنغو	الانترنت_كازينو
الشبكه	الشبكات	للشبكة	بالشبكة	شبكة	اكسترنانت
محلا	ومحلا	محلا_تجاريا	متجرا	موضعا	مستودعا

Tableau 4: Exemple génération des synonymes avec le modèle Zahran CBOW

Zahran SkipGram					
جرائم	الجرائم	وجرائم	لجرائم	جريمه	بجرائم
النصب الاحتيال	والنصب والاحتيال	لنصب عمليات الاحتيال	نصب للاحتيال	بالنصب احتيال	فالنصب الغش والاحتيال
تصنف المجموع الاولي	وتصنف والمجموع الاولي	تصنيفها مجموع	تصنف_ضمن بالمجموع	فتصنف للمجموع	صنفت لمجموع
تقع الانترنت	وتقع شبكة_الانترنت	يقع بشبكة الانترنت	ويقع الانترنت	فتقع بالانترنت	
الشبكة محلا	للشبكة ومحلا	بالشبكة محل	الشبكات بمحل	شبكة المحل	والشبكة

Tableau 5: Exemple de génération des synonymes avec le modèle Zahran SkipGram

3. Génération des paraphrases pour la réponse modèle

3.1. La combinaison des phrases

Après obtention du dictionnaire nous combinons tous les mots pour obtenir des phrases

La figure III - 11 montre le processus de génération des paraphrases :

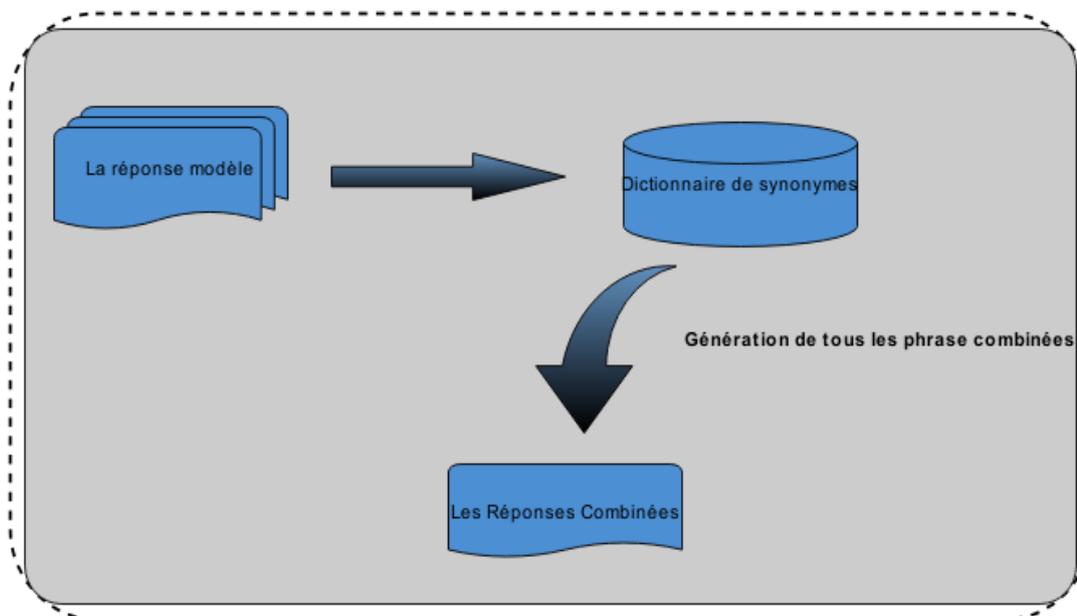


Figure III-11: La génération des paraphrases

Comme entrée on a l'ensemble des mots synonymes par exemple :

جرائم , الجرائم , الجريم , جريم
الانترنت , الشبكه , شبك
الشبكه , الانترنت , الشبكات , النظام , شبك
الاولي
التصنيفات , تصنف
احتيالي , التزوير , النصب , سرقة , الاحتيال
الاحتيال , السرقة , خيان , النصب
تقع
المجموع
محلا

Une combinaison de tous les mots de la liste(1) seront combinés avec tous les mots se trouvant dans chaque liste citée (2.....n)

Pour obtenir à la fin toute les combinaisons possibles.

Exemple :

جرائم الانترنت الشبكه الاولى تصنف الاحتيال النصب تقع المجموع محلا

جرائم الانترنت الشبكه الاولى تصنف الاحتيال الاحتيال تقع المجموع محلا

جرائم الانترنت الشبكه الاولى تصنف الاحتيال السرقة تقع المجموع محلا

جرائم الانترنت الشبكه الاولى تصنف الاحتيال خيان تقع المجموع محلا

جرائم الانترنت الشبكه الاولى تصنف احتيالي النصب تقع المجموع محلا

جرائم الانترنت الشبكه الاولى تصنف احتيالي الاحتيال تقع المجموع محلا

جرائم الانترنت الشبكه الاولى تصنف احتيالي السرقة تقع المجموع محلا

جرائم الانترنت الشبكه الاولى تصنف احتيالي خيان تقع المجموع محلا

جرائم الانترنت الشبكه الاولى تصنف التزوير النصب تقع المجموع محلا

جرائم الانترنت الشبكه الاولى تصنف التزوير الاحتيال تقع المجموع محلا

جرائم الانترنت الشبكه الاولى تصنف التزوير السرقة تقع المجموع محلا

جرائم الانترنت الشبكه الاولى تصنف التزوير خيان تقع المجموع محلا

جرائم الانترنت الشبكه الاولى تصنف النصب النصب تقع المجموع محلا

جرائم الانترنت الشبكة الاولى تصنف النصب الاحتيال تقع المجموع محلا

جرائم الانترنت الشبكة الاولى تصنف النصب السرقة تقع المجموع محلا

جرائم الانترنت الشبكة الاولى تصنف النصب خيان تقع المجموع محلا

جرائم الانترنت الشبكة الاولى تصنف سرقة النصب تقع المجموع محلا

جرائم الانترنت الشبكة الاولى تصنف سرقة الاحتيال تقع المجموع محلا

جرائم الانترنت الشبكة الاولى تصنف سرقة السرقة تقع المجموع محلا

جرائم الانترنت الشبكة الاولى تصنف سرقة خيان تقع المجموع محلا

جرائم الانترنت الشبكة الاولى التصنيفات الاحتيال النصب تقع المجموع محلا

جرائم الانترنت الشبكة الاولى التصنيفات الاحتيال الاحتيال تقع المجموع محلا

جرائم الانترنت الشبكة الاولى التصنيفات الاحتيال السرقة تقع المجموع محلا

جرائم الانترنت الشبكة الاولى التصنيفات الاحتيال خيان تقع المجموع محلا

جرائم الانترنت الشبكة الاولى التصنيفات احتيالي النصب تقع المجموع محلا

جرائم الانترنت الشبكة الاولى التصنيفات احتيالي الاحتيال تقع المجموع محلا

جرائم الانترنت الشبكة الاولى التصنيفات احتيالي السرقة تقع المجموع محلا

جرائم الانترنت الشبكة الاولى التصنيفات احتيالي خيان تقع المجموع محلا

جرائم الانترنت الشبكة الاولى التصنيفات التزوير النصب تقع المجموع محلا

.

.

.

جريم شبك شبك الاولى التصنيفات سرقة خيان تقع المجموع محلا

4. Modèles du calcul de similarité sémantique entre deux réponses courtes

On entamera le calcul de similarité sémantique, après la génération d'espace sémantique ou le modèle des WE et la construction du dictionnaire. Dans notre travail, nous avons utilisé 2 approches de calcul de similarité matriciel et le calcul de similarité somme des vecteurs

4.1. Le modèle calcul-matriciel (CM)

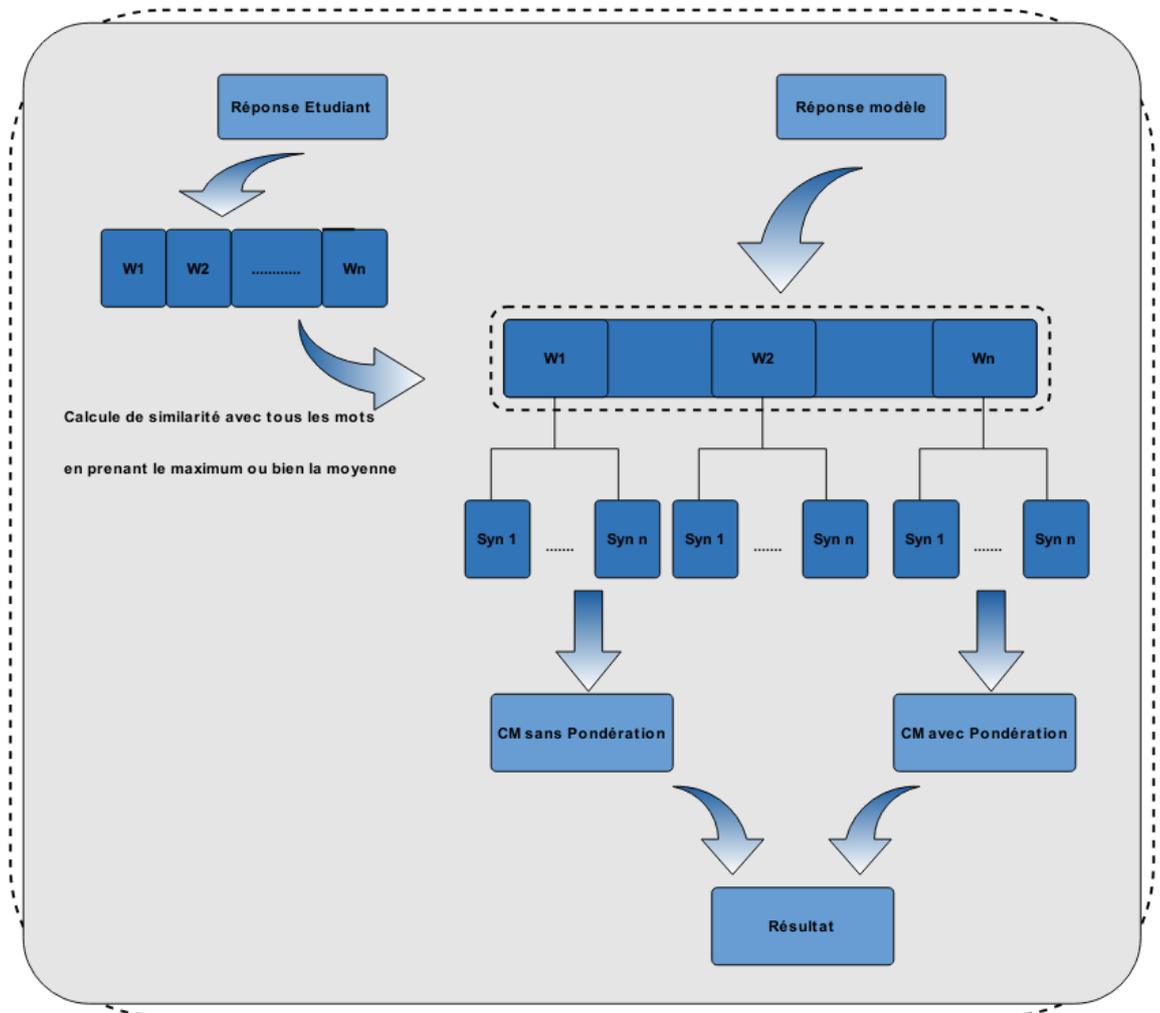


Figure III-12: Vue globale sur le fonctionnement du modèle CM

Le détail des étapes du fonctionnement de la CM représentées au niveau de la Figure III - 12:

a) Au début on fait un prétraitement (normalisation, stemming) pour les deux réponses RM, RE, tandis que chaque mot est représenté par un vecteur.

b) Une matrice est construite qui contient des lignes et des colonnes. Chaque ligne va représenter un mot de la réponse étudiant et chaque colonne représente un mot de notre réponse modèle

c) Le calcul de la similarité entre les mots de RM et RE, cette étape est une procédure itérative qui intervient dans la prise du vecteur d'espace sémantique pour

chaque mot de la réponse étudiant (Ei) et le vecteur de chaque mot de la réponse modèle (Mi) avec tous ses synonymes (Si) pour calculer la similarité entre eux , à la fin la valeur maximale des similarités sera prise:

$$\text{Max (Cosinus_sim (V(Ei), V(Mj), Cosinus_sim (V(Ei),V(Sji)).)}$$

Tels que :

V(Ei) est le vecteur de chaque mot de la RE

V(Mj) est le vecteur de chaque mot de la RM

V(Sji) est le vecteur de chaque synonyme de chaque mot de RM

Pour appliquer une pondération il suffit de multiplier les vecteurs V(Mi) et V(Si) par la pondération

d) Au début on procède à l'extraction des valeurs qui possèdent 1 , le nombre des 1 sera mis dans une variable k et nous retirons tous les lignes et les colonnes qui possèdent cette valeur par la suite on extrait le maximum de la matrice pour l'ajouter à une liste des max (qu'on notera liste_des_max) et retirer la ligne et la colonne de ce maximum. A la fin une multiplication sera faite avec une division sur le nombre des lignes et le nombre des colonnes

$$S = 2 * (k + \sum_i^N nbrM) / (m + n)$$

k : nombre des mots qui une similarité 1

nbrM : la similarité max

m :longueur en nombre de mots de la réponse étudiant

n : longueur en nombre de mots de la réponse modèle

Exemple explicatif :

Dataset cybercrime réponse modèle 9 réponse étudiant 4

1- Phase du cleaning :

	Réponse modèle	Réponse étudiant
Avant le cleaning	جرائم النصب والاحتيال تصنف في المجموعة الأولى التي تقع على الانترنت أي أن الشبكة تكون محلا لها	جرائم النصب الاحتيال، التصنيف: جرائم تقع على الانترنت
Après le cleaning	جرائم النصب الاحتيال تصنف المجموع الاولي تقع الانترنت الشبكه محلا	جرائم النصب الاحتيال التصنيف جرائم تقع الانترنت

2- La phase d'extraction des synonymes :

Mot	Synonymes
جرائم	[الجرائم , الجريم , جريم]
النصب	[الاحتيال , السرقة , خيان]
الاحتيال	[احتيالي , التزوير , النصب , سرق]
تصنف	[التصنيفات]
المجموع	[]
الاولي	[]
تقع	[]
الانترنت	[الشبكة , شبك]
الشبكه	[الانترنت , الشبكات , النظام , شبك]

محلا	[]
------	----

3- Phase de calcul de similarité :

Nous calculons la similarité de chaque mot de la réponse étudiant avec chaque mot de la réponse modèle ainsi que ses synonymes et on prend la similarité la plus grande (maximum).

En suivant l'algorithme qui se présente dans la figure III - 13 :

debut

Entrée la réponse modele et la réponse etudiant

creer une matrice $T[n][m]$ de taille $n=\text{len}(\text{RM})$ et $m =\text{len}(\text{RE})$

for i to n

for j to m

$T[i][j]$ = similarité max entre chaque mot de RE et de RM avec ses synonymes

eleminer les $T[i][j]$ qui on une valeur 1 et les mettres dans une variable K

eleminer les $T[i][j]$ qui on une valeur maximal et les mettre dans une liste L

calculer la similarité ajouter $(k+\text{valeur de la liste L}) * 2 / \text{la taille de RM} + \text{La taille de RE}$

fin

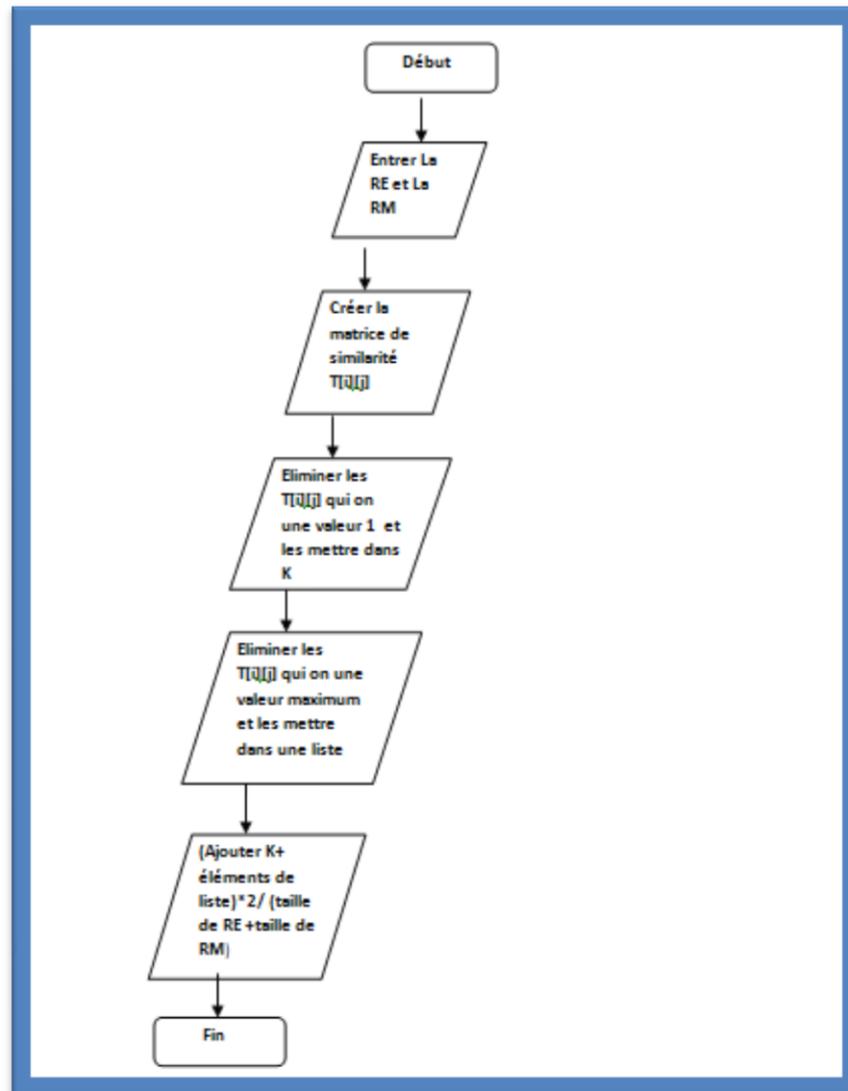


Figure III-13: Algorithme de calcul de similarité

Voici la matrice générée avec le modèle CM dans le cas d'une matrice sémantique.

	جرائم	الانترنت	الشبكة	الاولي	تصنف	الاحتيال	النصب	تقع	المجموع	محلا
التصنيف	0.02046472	0.02211852	0.03024137	0.02207419	0.14854973	0.02940406	0.02391638	0.01058252	0.01648083	0.00856668
الانترنت	0.20910721	1.0	1.0	0.05145364	0.0516847	0.08217033	0.06865717	0.05303181	0.06128818	0.03836153
الاحتيال	0.19427229	0.06865717	0.07724245	0.03494288	0.0407328	1.0	1.0	0.05573644	0.01920122	0.10392065
تقع	0.16017144	0.05303181	0.07038358	0.05457011	0.02216877	0.05573644	0.05675568	1.0	0.01536784	0.04770903
النصب	0.31502404	0.05844568	0.063912	0.0303258	0.02598409	1.0	1.0	0.05493472	0.01346871	0.11664466
جرائم	1.0	0.20910721	0.20910721	0.06302011	0.04575652	0.17762487	0.1559738	0.06342536	0.02492503	0.10110207

Tableau 6 : matrice générée avec le modèle CM

	جرائم	الانترنت	الشبكة	الاولي	تصنف	الاحتيال	النصب	تقع	المجموع	محلا
التصنيف	0.02046472	0.02211852	0.03024137	0.02207419	0.14854973	0.02940406	0.02391638	0.01058252	0.01648083	0.00856668
الانترنت	0.20910721	1.0	1.0	0.05145364	0.0516847	0.08217033	0.06865717	0.05303181	0.06128818	0.03836153
الاحتيال	0.19427229	0.06865717	0.07724245	0.03494288	0.0407328	1.0	1.0	0.05573644	0.01920122	0.10392065
تقع	0.16017144	0.05303181	0.07038358	0.05457011	0.02216877	0.05573644	0.05675568	1.0	0.01536784	0.04770903
النصب	0.31502404	0.05844568	0.063912	0.0303258	0.02598409	1.0	1.0	0.05493472	0.01346871	0.11664466
جرائم	1.0	0.20910721	0.20910721	0.06302011	0.04575652	0.17762487	0.1559738	0.06342536	0.02492503	0.10110207

Tableau 7: sélection des nombre dans la matrice générée avec le modèle CM
 Par la suite, nous appliquons une itération pour récupérer le nombre des 1 dans la matrice et supprimer la ligne et la colonne de cet élément :

	الاولي	تصنيف	المجموع	محلا
التصنيف	0.0220 7419	0.14854 973	0.01648 083	0.0085 6668

Tableau 8 : la matrice après élimination des mots en commun

Par la suite, nous appliquons une itération pour récupérer le maximum de la matrice et supprimer la ligne et la colonne de cet élément

	الاولي	المجموع	محلا

Tableau 9: la matrice après élimination du maximum

Calcule de similarité :

Nombre de k=5

liste_des_max : [0.14854973]

$$s=2*(0.14854973+5)/16=0.6435$$

La note calculée = $0.2761*5=$ **3.21**

Rappelons que la note accordée par l'enseignant correcteur est de **3.87**

4.2. Le modèle somme-vecteurs (SV)

4.2.1. Somme des vecteurs en utilisant la moyenne

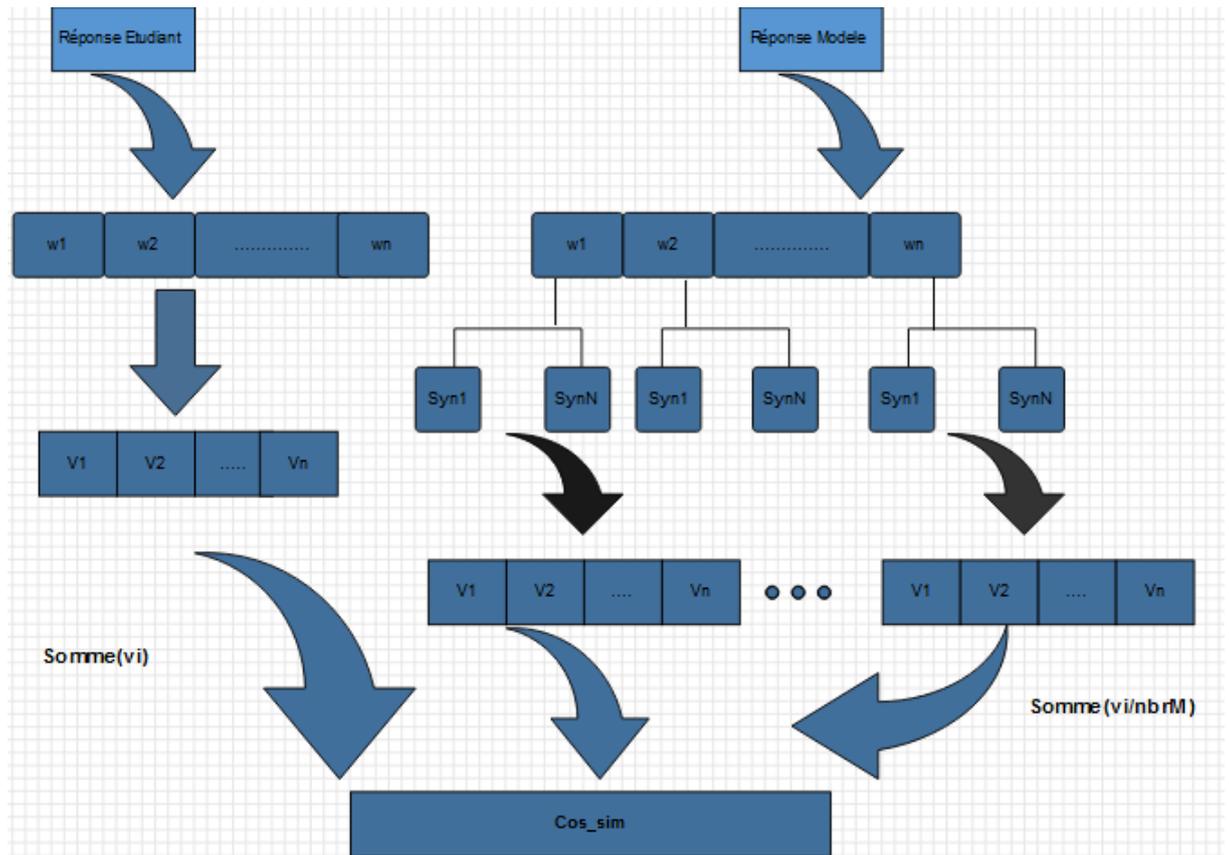


Figure III-14: Vue globale sur le fonctionnement du modèle SV en utilisant la moyenne

La somme des vecteurs est la somme des vecteurs représentatifs de chaque mot d'une phrase. La figure III - 14 montre le fonctionnement du modèle SV en utilisant la moyenne

Soit :

Une phrase RE composée des mots $E_1, E_2 \dots E_N$

Une phrase RM composée des mots $M_1, M_2 \dots M$

Chaque Mot de RM a ses synonymes $\{S_1, S_2, \dots, S_n\}$

La somme des vecteurs pour la réponse étudiant (RE), consiste à sommer les vecteurs représentatifs de chaque mot de la phrase.

La somme des vecteurs pour la réponse modèle(RM), est la somme des vecteurs représentatifs pour chaque mot avec ses synonymes et on prend la moyenne des vecteurs de synonymes de chaque mot

A la fin on obtient deux vecteurs que leur similarité va être calculé par le cos similarité ou la combinaison entre les deux. La formule en dessous

$$V1 = \sum_i^N V(Ei)$$

$$V2 = \sum_i^N Moy(\sum_i^s V(Si))$$

SYM_Cos(VM ,VE) ou bien SYM_Jaccard(VM ,VE) ou bien la combinaison entre les deux

Dans la somme des vecteurs simple (sans pondération) , le coefficient β est égal à 1, tandis que la somme des vecteurs avec pondération consiste à modifier la pondération des vecteurs de mot en affectant à β une valeur représentative de l'importance de chaque terme (IDF, TFminmax, TFlog, TF-IDFminmax, POS,une combinaison entre TF-IDFminmax et le POS).

Exemple :

Réponde modèle	Réponse étudiant
جرائم النصب الاحتيال تصنف المجموع الاولي تقع الانترنت الشبكه محلا	جرائم النصب الاحتيال التصنيف جرائم تقع الانترنت

$$VE = v \times \beta + (جرائم) \times \beta + v \times \beta + (النصب) \times \beta + v \times \beta + (الاحتيال) \times \beta + v \times \beta + (التصنيف) \times \beta + v \times \beta + (تقع) \times \beta + v \times \beta + (الانترنت) \times \beta$$

$$VM = (سرق) \times \beta + v \times \beta + (النصب) \times \beta + v \times \beta + (التزوير) \times \beta + v \times \beta + (احتمالي) \times \beta + v \times \beta + (الاحتيال) \times \beta + v \times \beta + (محل) \times \beta + v \times \beta + (الشبكه) \times \beta + v \times \beta + (الانترنت) \times \beta + v \times \beta + (المجموع) \times \beta + (تقع) \times \beta + (شبكة) \times \beta + (النظام) \times \beta + v \times \beta + (شبكة) \times \beta + v \times \beta + (الانترنت) \times \beta + v \times \beta + (الشبكه) \times \beta + v \times \beta + (محل) \times \beta$$

$$\times \beta) \div nbr_{syn} +$$

$$(v \times \beta + (التصنيفات) \times \beta) \div nbr_{syn} +$$

$$(v \times \beta + (المجموع) \times \beta) + (v \times \beta + (الاولي) \times \beta) + (v \times \beta + (تقع) \times \beta) +$$

$$(v \times \beta + (الانترنت) \times \beta) + (v \times \beta + (الشبكه) \times \beta) + (v \times \beta + (شبكة) \times \beta) \div nbr_{syn} +$$

$$(v \times \beta + (الشبكه) \times \beta) + (v \times \beta + (النظام) \times \beta) + v \times \beta + (شبكة) \times \beta + v \times \beta + (الانترنت) \times \beta + v \times \beta + (الشبكه) \times \beta + v \times \beta + (محل) \times \beta$$

$$(v \times \beta + (محل) \times \beta)$$

tel que :

V : le vecteur des mot

VM : la somme des vecteur de la RM

VE : la somme des vecteurs de la RE

nbr_{syn} : Nombre des synonymes de chaque mot de la réponse étudiant

Nous calculerons la similarité entre VM et VE

$$SYM_{Cos} = (VM, VE) = 0.71$$

La note calculée est = 3.55

Rappelons que la note accordée par l'enseignant correcteur est de 3.87

4.2.2. Somme des vecteurs en utilisant le maximum

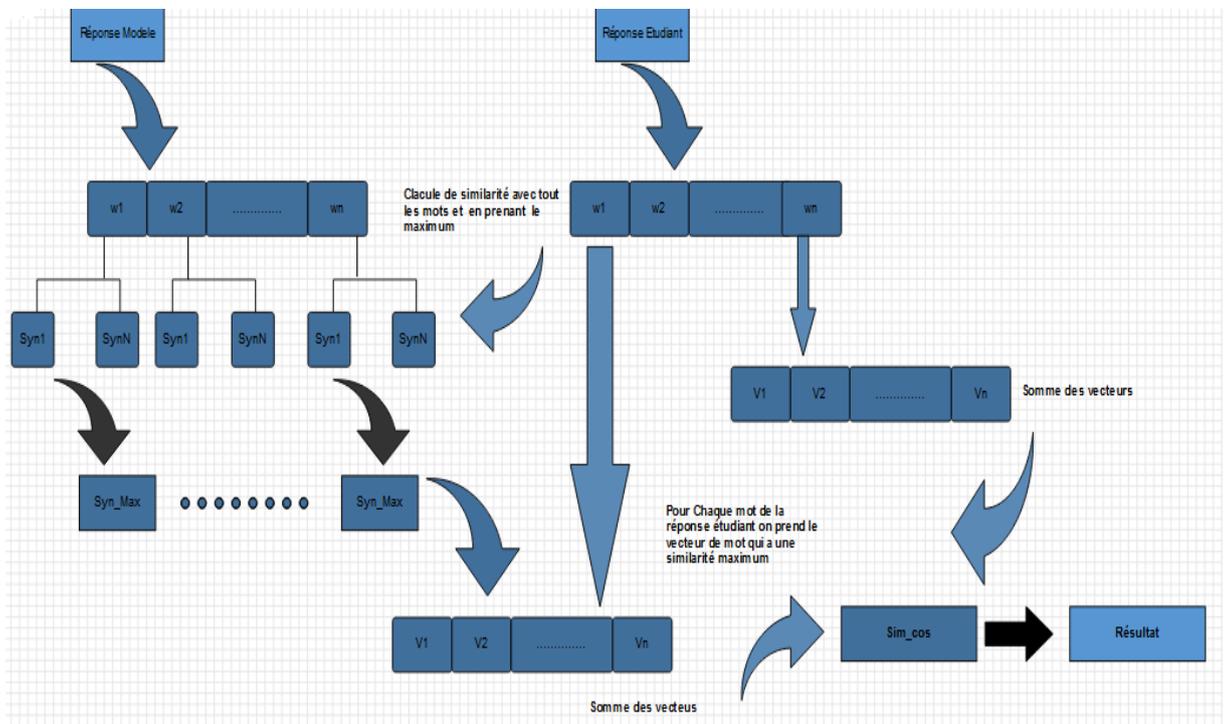


Figure III-15: Vue globale sur le fonctionnement du modèle SV en utilisant le max

La somme des vecteurs pour la réponse modèle(RM), consiste à calculer la similarité de chaque mot de la réponse étudiant avec chaque mot de la réponse modèle ainsi que ses synonymes et on prend la similarité la plus élevée (figure III-

15). Ensuite on somme les vecteurs représentatifs de chaque mot de la réponse modèle qui a une similarité élevée par rapport à chaque mot de la réponse étudiant.

La formule :

$$V1 = \sum_{k=0}^n V(E_k)$$
$$V2 = \sum_{k=0}^n \text{similarité_max} (F_{\text{modèle}}, \text{mot}_{\text{étudiant}}^k)$$

Tel que :

v1 : la somme des vecteurs de la réponse étudiant

v2 : la somme des vecteurs de la réponse modèle

F_modèle : c'est tous les mots de la réponse modèle

mot_étudiant : chaque mot de la réponse étudiant

similarité_max : la similarité maximum entre F_modèle et mot_étudiant

Exemple :

Légende de la Figure III - 16:

Colonne : Réponse étudiant (RE)

ligne : Réponse Modèle (RM)

v : le vecteur du mot

vm: résultat de la somme des vecteurs de la réponse modèle

ve : résultat de la somme des vecteurs de la réponse étudiant

	جرائم	الانترنت	الشبكة	الاولي	تصنف	الاحتيال	النصب	تقع	المجموع	محلا
التصنيف	0.020465	0.022119	0.030241	0.022074	0.148550	0.029404	0.023916	0.010583	0.016481	0.008567
الانترنت	0.209107	1.000000	1.000000	0.051454	0.051685	0.082170	0.068657	0.053032	0.061288	0.038362
الاحتيال	0.194272	0.068657	0.077242	0.034943	0.040733	1.000000	1.000000	0.055736	0.019201	0.103921
تقع	0.160171	0.053032	0.070384	0.054570	0.022169	0.055736	0.056756	1.000000	0.015368	0.047709
النصب	0.315024	0.058446	0.063912	0.030326	0.025984	1.000000	1.000000	0.054935	0.013469	0.116645
جرائم	1.000000	0.209107	0.209107	0.063020	0.045757	0.177625	0.155974	0.063425	0.024925	0.101102

Figure III-16: Matrice des similarités max

Nous prenons le vecteur du mot qui à la similarité maximum entre chaque mots de la réponse étudiant et avec tous les mots de la réponse modèle avec ces synonymes:

$$V_m = v(\text{mot_similarité_max}(\text{التصنيف}, \text{réponse_modèle}) = \text{التصنيفات}) +$$

$$v(\text{mot_similarité_max}(\text{الانترنت}, \text{réponse_modèle}) = \text{الانترنت}) +$$

$$v(\text{mot_similarité_max}(\text{الاحتيال}, \text{réponse_modèle}) = \text{الاحتيال}) +$$

$$v(\text{mot_similarité_max}(\text{تقع}, \text{réponse_modèle}) = \text{تقع}) +$$

$$v(\text{mot_similarité_max}(\text{النصب}, \text{réponse_modèle}) = \text{النصب}) +$$

$$v(\text{mot_similarité_max}(\text{جرائم}, \text{réponse_modèle}) = \text{جرائم}) +$$

$$\text{SYMCos} = (V_M, V_E) = 0.96$$

La note = **4.8**

Rappelons que la note accordée par l'enseignant correcteur est de **3.87**.

4.2.3. Somme des vecteurs en utilisant les combinaisons de paraphrases

Elle consiste à sommer les vecteurs représentatifs de chaque mot pour chaque phrase qui est dans notre fichier (combinaison de toute les réponses modèle possible) après on prend la similarité max entre la somme des vecteurs de la réponse étudiant et la somme des vecteurs de toute les réponses modèle possible. La figure III - 17 montre le fonctionnement du modèle SV en utilisant les paraphrases

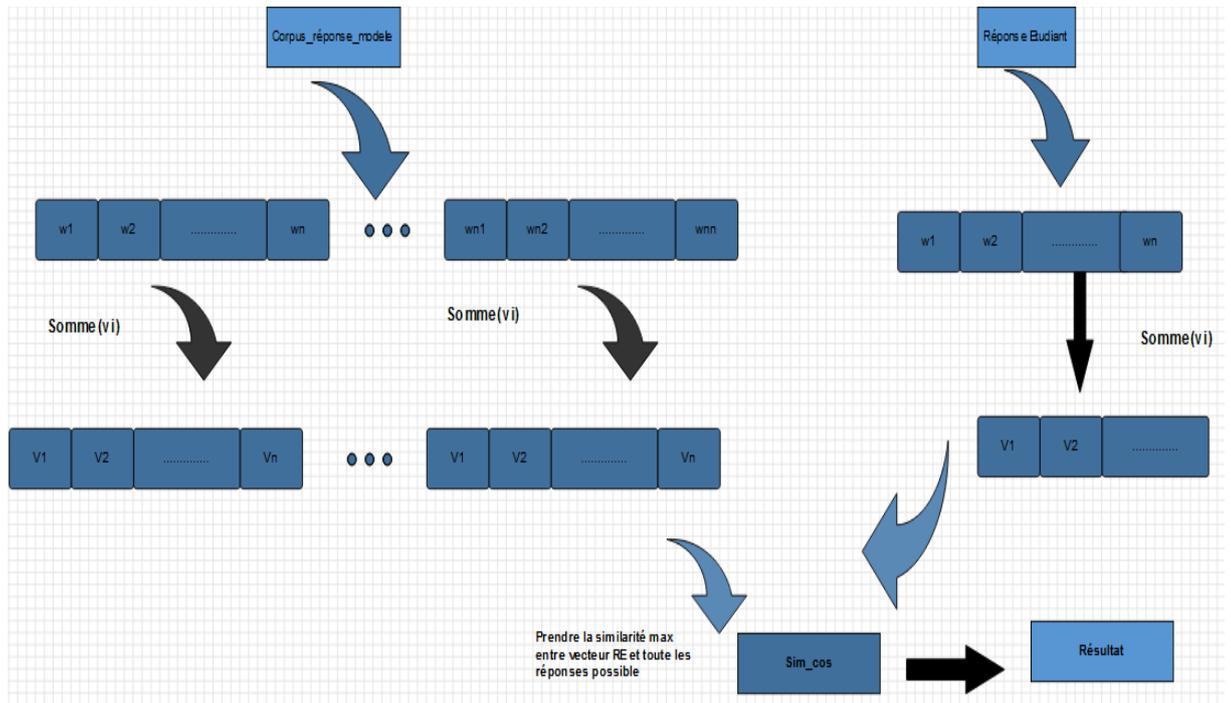


Figure III-17: Vue globale sur le fonctionnement du modèle SV en utilisant les paraphrases

5. Pondération des mots basée sur la fréquence de mots dans le corpus

Les méthodes de pondération sont utilisées pour distinguer l'importance des termes les uns et les autres ou du vocabulaire des deux documents pour lesquels nous mesurons leur similitude. Dans La thématique d'espace sémantique, le couple de réponses (modèle, étudiant) représente les deux documents pour l'approche WE dans le les documents sont substitués par les réponses à comparer. Ce vocabulaire est très petit pour le prendre comme document de source pour le calcul statistique des pondérations (vocabulaire non significatif). Par conséquent, le concept disant qu'un corpus de la langue en question est un bon représentant des termes de la langue en générale.

Nous avons utilisés les différentes pondérations (TF, IDF, TF IDF) des mots de chacun corpus prétraités (CNN , CNN+BBC, Khaleej ,Cybercriminalité)

5.1. Calcul des poids

Méthode utilisée avec le Tf-Idf, Elle permet d'évaluer l'importance d'un terme contenu dans un document, relativement à une collection. Le poids augmente

proportionnellement avec le nombre d'occurrences du mot dans le document. Il varie également en fonction de la fréquence du mot dans la collection.

Ainsi, la fréquence inverse du document (idf) est une mesure de l'importance du terme dans l'ensemble des documents. Dans le cas du tf-idf, elle vise à donner un poids plus important aux termes les moins fréquents, considérés comme les plus discriminants. Tout d'abord on va calculer le TF (calculée par le logarithme du nombre de fois où le mot apparaît dans le corpus divisé par le nombre de mot total dans le corpus diviser par le max des tf-logs obtenus).

5.1.1. TF_MIN_MAX

Le TF sert au calcul de la fréquence ou densité d'un mot clé au sein d'un document: nombre de fois que le mot clé apparaît / nombre de mots. [35]

$$TF = TFlog(w) = -\log\left(\frac{nbr_M}{N}\right)$$

$$TFminmax(w) = TFlog(W)/Max(TFlogs)$$

nbr_M : nombre d'apparition du mot dans un document

N : taille du document

Max(TFlogs): nombre maximal dans le TFlogs.

5.1.2. IDF

L'IDF introduit la notion d'importance qualitative du mot clé: nombre de contenus / nombre de contenus contenant le terme. Il évalue la rareté du mot clé dans un corpus d'autres contenus ayant un même champ lexical.

$$IDF = \log\left(\frac{N}{Df}\right)$$

Où :

Df: Le nombre de documents contenant le terme.

N : Le nombre total de documents de la base documentaire.

5.1.3. TF-IDF

Génère une estimation globale de l'importance du mot par rapport à la langue entière.

$$TFIDF(w) = IDF(w) \times TF(w)$$

5.2. Pondération des mots basée sur l'étiquetage morphosyntaxique

5.2.1. POS (Part Of Speech)

Segmentation, analyse, Marquage POS (Part Of Speech) qui est une méthode qui prend en considération le côté grammatical et le contexte (Nom, verbe, adjectif...), subdivision de phrases, modèles syntaxiques, tokenization, segmentation de mots. elle consiste à identifier pour chaque mot sa classe morphosyntaxique à partir de son contexte et de connaissances lexicales .L'étiquetage morphosyntaxique a la réputation d'être la tâche d'analyse linguistique automatique la plus facile à réaliser, puisque le niveau de performance atteint par les systèmes est comparable à celui des humains en terme de qualité d'étiquetage, avec l'avantage que les systèmes offrent une vitesse d'annotation sans commune mesure avec celle d'un humain, qui ne peut annoter au mieux que quelques milliers de mots à l'heure[40].

Pour cette raison que nous nous sommes orientés vers cette approche.

Dans notre cas, nous avons utilisé l'analyseur « Stanford coreNLP »[41] pour générer l'étiquetage des réponses, ensuite, nous avons réalisé une série de tests qui a abouti à la considération de 3 catégories : les noms "N", les verbes "V" et la 3eme catégories "A" pour le reste. Voici quelques exemples :

CC : conjonction, pondéré par 0.1

CD : cardinal ou numérique pondéré par 0.1

IN : préposition pondéré par 0.1

JJ : Adjectif, pondéré par 0.1

NN : nom commun, pondéré par 0.4

RB : adverbe, pondéré par 0.1

VB : verbe, pondéré par 0.5

5.3. Combinaison entre le POS et le TF_MIN_MAX

$$\text{Com (W)} = \text{POS (W)} * \text{TFminmax (W)}$$

La pondération des mots permet d'améliorer la similarité. L'impact de la pondération des mots est illustré dans la partie expérimentation.

6. Passage au score

Après l'obtention des similarités entre les RE et les RM, un passage au score va être effectué et cela par conversion des similarités au score pour cela nous avons utilisé un algorithme non supervisé de clustering non hiérarchique K-means

6.1. Classifieur k-means [42] [43]

K-means (ou K-moyennes) : C'est l'un des algorithmes de clustering les plus répandus. Il permet d'analyser un jeu de données caractérisées par un ensemble de descripteurs, afin de regrouper les données "similaires" en groupes (ou clusters). La similarité entre deux données peut être inférée grâce à la "distance" séparant leurs descripteurs ; ainsi deux données très similaires sont deux données dont les descripteurs sont très proches¹⁰. Cette définition permet de formuler le problème de partitionnement des données comme la recherche de K "centroïdes", autour desquelles peuvent être regroupées les autres données. Pour traiter les datasets, où les notes varient de 1 à 5, nous avons choisi 11 classes, k=11 pour permettre un pas de 0.5 dans cet intervalle.

La figure III- 18 présente un exemple de classification des notes :

¹⁰<https://dataanalyticspost.com/Lexique/k-means-ou-k-moyennes/>

Cluster Pour les notes des étudiants

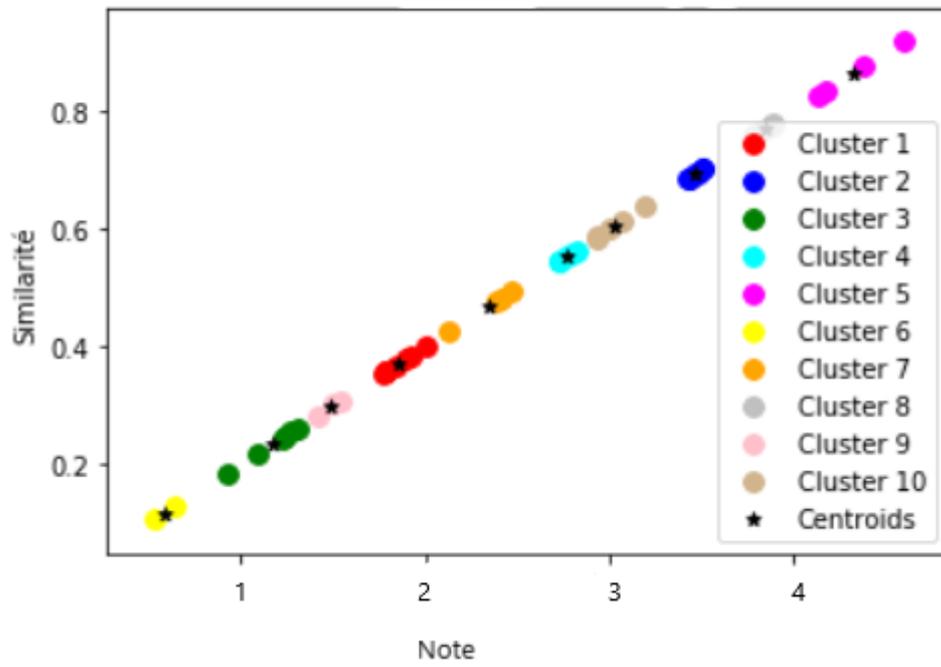


Figure III-18: Classification selon le K-means

chapitre IV. Résultats expérimentaux et évaluation

Dans ce chapitre, nous montrons les expériences et les résultats des tests réalisés afin de voir l'amélioration des résultats par rapport au travail déjà fait [16][34], ainsi nous discutons les résultats obtenus pour montrer l'impact de la construction du corpus de réponses modèles.

1. Démarche expérimentale

Après la définition des grandes lignes de notre approche dans le chapitre 3, nous allons évaluer :

L'impact de l'élaboration de corpus des réponses modèle dans l'évaluation automatique des réponses par composition à base de synonymes. Entre autre évaluer l'apport :

- L'impact du stemmer
- La qualité du dictionnaire dans le calcul des similarités entre RM et RE

1.1. L'impact de l'élaboration de corpus des réponses modèle dans l'évaluation automatique des réponses par composition à base de synonymes

Nous voulons analyser l'impact des synonymes sur l'évaluation des réponses des étudiants.

1.1.1. Calcul de similarité

Nous avons développé des méthodes de calcul de similarité : Somme des vecteurs, Calcul matriciel

Calcul matriciel (CM)

Ce module regroupe toutes nos méthodes dérivées du Calcul matriciel. Qui sont :

- CM_sans_Ponderation : Calcul matriciel unitaire où $\beta = 1$
- CM_TFMINMAX, CM_IDF, CM_TF_IDF, CM_POS,
- CM_TFMINMAX_POS .où β varie selon les mesures générées par les méthodes dont le nom est porté par la méthode.

Somme des Vecteur(SV) :

Somme des vecteur unitaire où $\beta = 1$

1.1.2. Calcul des notes et évaluation

Les modèles de calcul de similarité renvoient des valeurs entre 0 et 1 , ces valeurs doivent être convertie en note afin d'évaluer correctement les modèles, la classification est faite par le Kmeans.

1.2. L'impact du stemmer

Nous projetons d'analyser l'impact du stem sur l'évaluation automatique des questions à réponses courtes en langue arabe en utilisant les synonymes. Les meilleurs résultats étaient avec le lourd-stem (khouja) et Tashapyne pour l'approche espace sémantique tandis que pour l'approche WE le light-stem (Stanford-coreNLP) à donner des bons résultats

1.3. La qualité du dictionnaire (évaluation du dictionnaire)

On déduit La meilleure approche qui génère un dictionnaire ou ses synonymes sont corrects prenant un exemple :

1.3.1. Évaluation des synonymes par rapport au mot de sketch engine

Dans notre cas le choix des synonymes se fait par un expert qui est l'enseignant. Pour une meilleure évaluation nous avons évalués notre dictionnaire par rapport aux synonymes de sketch engine¹¹.

Pour l'évaluation nous avons calculé la concordance des synonymes, le calcul se fait comme suit :

Nous prenons 32 mots en suite pour chaque mot nous prenons les 10 premiers synonymes avec le dictionnaire de sketch engine et notre dictionnaire, par la suite on applique la formule :

$$\frac{\left(\sum \frac{N}{N_s}\right)}{len(mot)}$$

¹¹ Un outil ultime pour explorer le fonctionnement de la langue.il contient 500 corpus prêts à l'emploi dans plus de 90 langues, chacun ayant une taille allant jusqu'à 30 milliards de mots disponible. Il a un outil de génération de synonymes disponible sur : <https://embeddings.sketchengine.co.uk>

Ou :

N : le nombre de synonymes en commun entre L1 et L2

L1 : la liste de synonymes de dictionnaire de sketch engine

L2 : la liste de synonymes de notre dictionnaire

Ns : le nombre de synonymes de chaque mot (dans notre cas 10)

Len (mot) : le nombre de mots (dans notre cas 32)

Nous prenons pour exemple :

S = [, تتناول, مكونات, حيه, الاطار, الانسان, يحصل, مقومات, يشترك, سائر, تطبيق المعلومات, , المجالات, المعرفيه, الفيزيائيه, الكيمائيه, يتخذ, كافيه, السليبه, النهر, الصحراء, محدود, حيز, تبادلات, ,] يتعلق, يصف, نظام, التوازن, مستقره, تصل, نمو, معدلات, تغير

Aravec CBOW	Aravec SkipGram	Zahran CBOW	Zahran SkipGram	Espace Sémantique Khaleej
0.24	0.35	0.36	0.30	0.03

Tableau 10: évaluation des synonymes selon les mots de sketch engine

D'après les résultats du tableau 10 nous avons constaté qu'avec le modèle ZAHRAN CBOW, le résultat était meilleur par rapport à celui de l'espace sémantique de KHALEEJ qui était mauvais (d'une part les mots sont sans stem et d'autre part ils n'existent pas dans cet espace).

Nos résultats seront comparés à ceux obtenus par les travaux précédents dans le même projet afin d'en apprécier l'impact de l'introduction de la notion de synonymes. Pour cela nous utilisons deux métriques d'évaluation et deux data sets.

2. Métriques d'évaluation

Le succès de tout projet ne peut être prononcé que suite à l'évaluation d'une approche proposée en comparant un résultat produit avec des résultats corrects attendus. Pour évaluer les scores obtenus par rapport aux scores manuels ; nous avons utilisés deux métriques : La corrélation de Pearson [44] c'est la métrique la

plus fréquemment utilisée par les recherches dans ce domaine nous avons évalué les scores obtenus par rapport aux scores manuels fournis, L'erreur quadratique moyenne (Root Mean Squared Error (RMSE) [45]) pour quantifier la différence entre le résultat (score) obtenu par le système et celui obtenu par l'expert humain.

2.1. Coefficient de Pearson

En statistiques, étudier la corrélation entre deux ou plusieurs variables statistiques numériques, c'est étudier l'intensité de la liaison ("proportionnalité") qui peut exister entre ces variables. La mesure de la corrélation linéaire entre les deux se fait alors par le calcul du coefficient de corrélation linéaire, noté r . Ce coefficient est égal au rapport de leur covariance et du produit non nul de leurs écarts types. Le coefficient de corrélation est compris entre -1 et 1

Corrélation	Négative	Positive
Faible	de -0,5 à 0,0	de 0,0 à 0,5
Forte	de -1,0 à -0,5	de 0,5 à 1,0

Tableau 11: Signification des valeurs la corrélation de pearson

Plus le coefficient est proche des valeurs extrêmes -1 et 1, plus la corrélation linéaire entre les variables est forte ; on emploie simplement l'expression « fortement corrélées » pour qualifier les deux variables. Une corrélation égale à 0 signifie que les variables ne sont pas corrélées linéairement. Le coefficient de corrélation est multiplié par 100 pour exprimer un pourcentage de corrélation. Dans notre cas les variables statistiques à considérer sont celles définies dans deux vecteurs l'un contenant les valeurs de scores entre les couples de réponses du dataset (réponse de l'étudiant, réponse modèle de l'enseignant) calculés automatiquement, le deuxième vecteur contient les scores, pour les mêmes couples de réponses, calculées par l'expert humain. L'objectif dans notre travail revient à maximiser ce coefficient.

2.2. Erreur quadratique RMSE (Root Mean Squared Error (RMSE))

L'erreur quadratique moyenne permet de quantifier une mesure synthétique de l'erreur globale commise. Pour calculer l'erreur quadratique moyenne RMSE, les erreurs individuelles sont tout d'abord élevées au carré, puis additionnées les unes

aux autres. On divise ensuite le résultat obtenu par le nombre total d'erreurs individuelles, puis on en prend la racine carrée. L'erreur quadratique est probablement le critère quantitatif le plus utilisé pour comparer des valeurs calculées (ici les scores ou notes automatiques) et valeurs observées (scores manuels attribués par l'expert humain. C'est cette fonction que nous tentons de minimiser dans le cadre de ce travail.

3. Jeux de données (Datasets)

Nous avons effectué des tests sur le data set :

Gomaa dans le but de comparer nos résultats avec les résultats de [16] [34]

Cybercriminalité dans le but de comparer les résultats avec les approches de calcul de similarité sans synonymes

3.1. Gomaa Dataset (Cairo university dataset)[46][47]

Les questions présentées dans le data set couvrent un chapitre du programme d'études égyptien. L'ensemble de données contient 61 questions, 10 réponses pour chacune, avec un nombre total de 610 réponses.

3.2. Cyber criminality Arabic Short Answer Grading Data set (CCASAG)¹² :

Les questions présentées dans le data set couvrent des cours sur la cybercriminalité enseigné à l'université de Saad Dahleb blida1 durant l'année 2018/2019. L'ensemble des réponses modèles contient 48 questions, avec un nombre total de 2133 réponses. L'ensemble de données contient une collection de réponses et notes des élèves, notées par deux enseignants qui ont donné des notes entre 0 et 5 et obtenu un coefficient de corrélation de Pearson (r) et une erreur quadratique moyenne (RMSE) de **0.83**, respectivement entre les deux annotateurs [48]

¹² <https://drive.google.com/open?id=13DoT0HqDwrDnwcwTFucJsceM9rfK197C>

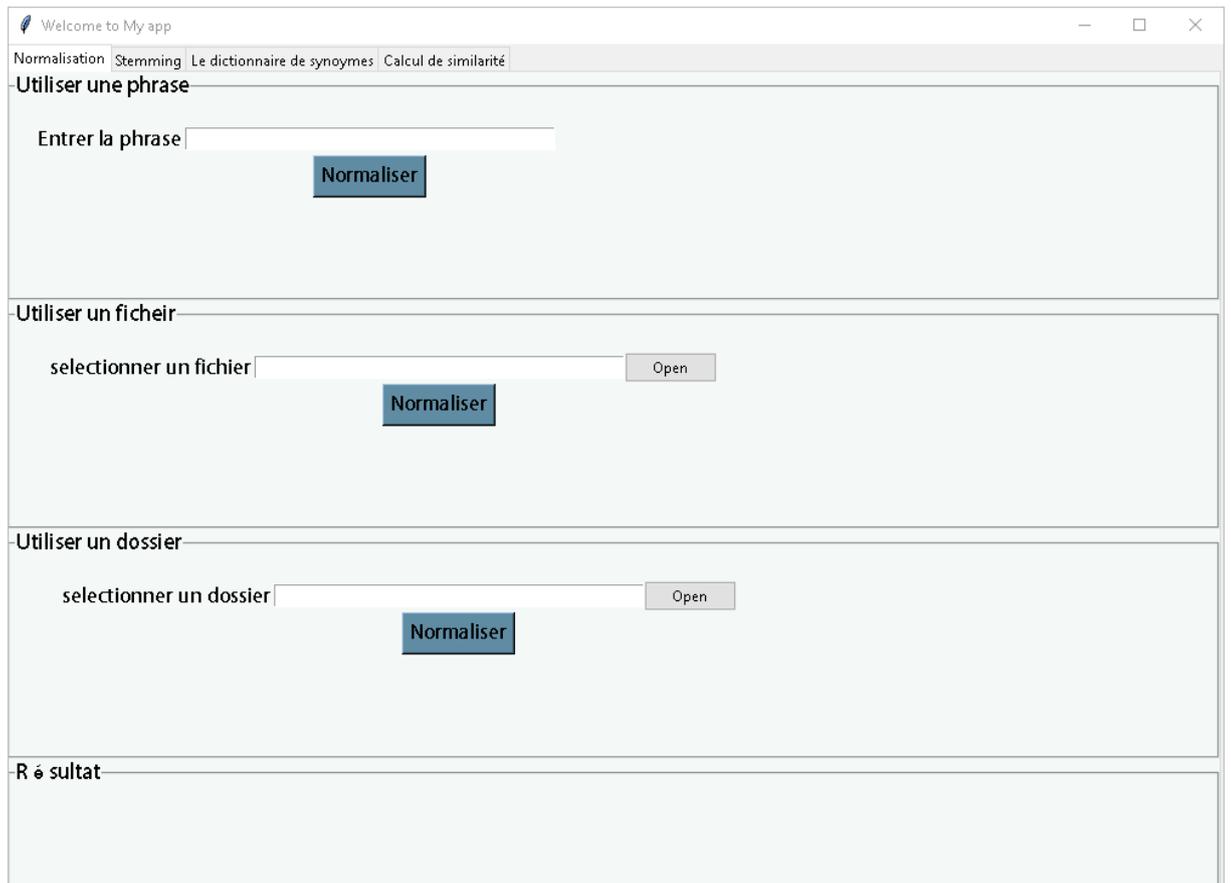


Figure IV-1: Outils de normalisation

4.1.2. Stemming :

Cet outil accorde un choix entre les 3 stemmer que nous avons utilisé(Khoja , Tashapyne et stanfordNLP,). En fin du traitement, notre texte ou bien corpus va être stemer « voir Figure IV - 2»

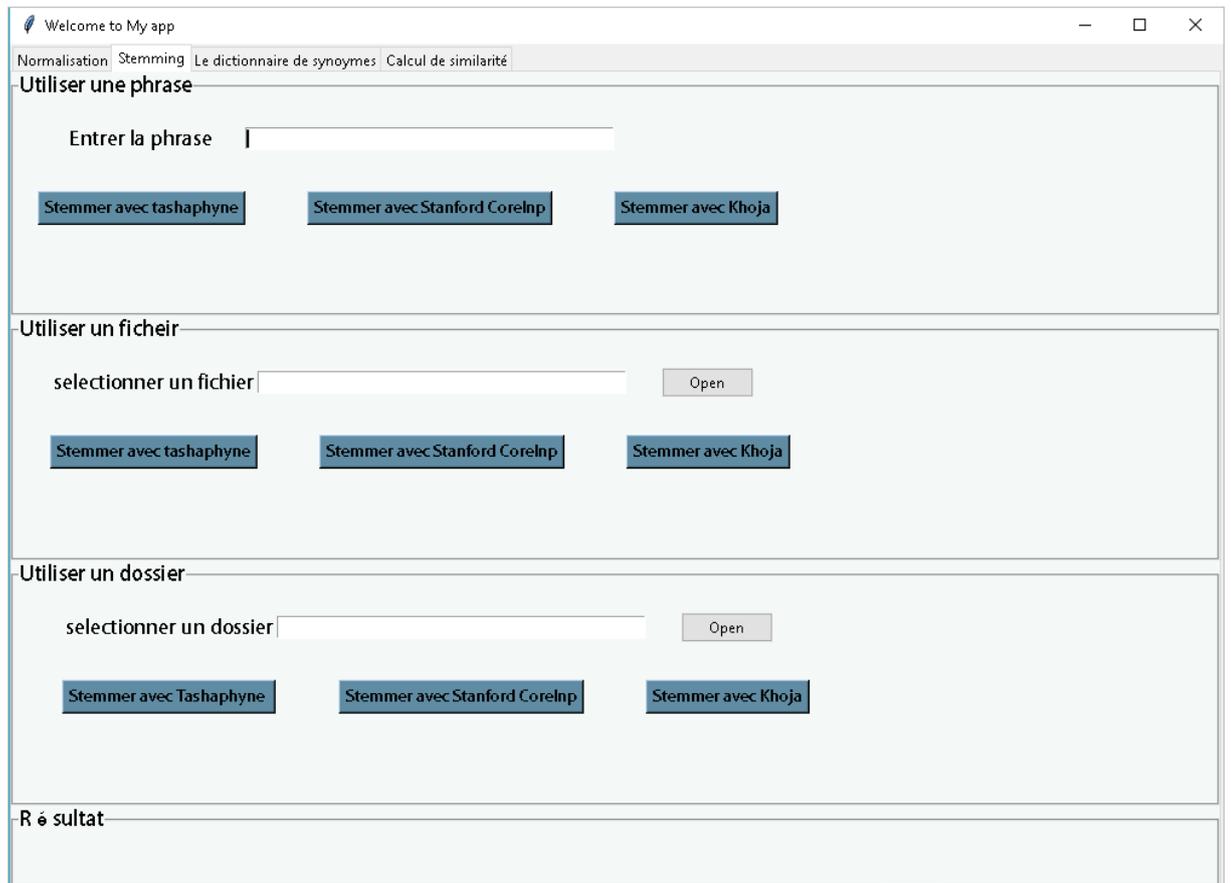


Figure IV-2: Outils de stemming

4.2. Outils de la génération du dictionnaire des synonymes pour les réponses modèle

L'utilisateur doit spécifier la réponse modèle, l'espace sémantique ou bien le modèle de WE en cliquant sur le bouton générer ; le dictionnaire se produira et s'affichera « Figure IV- 3»

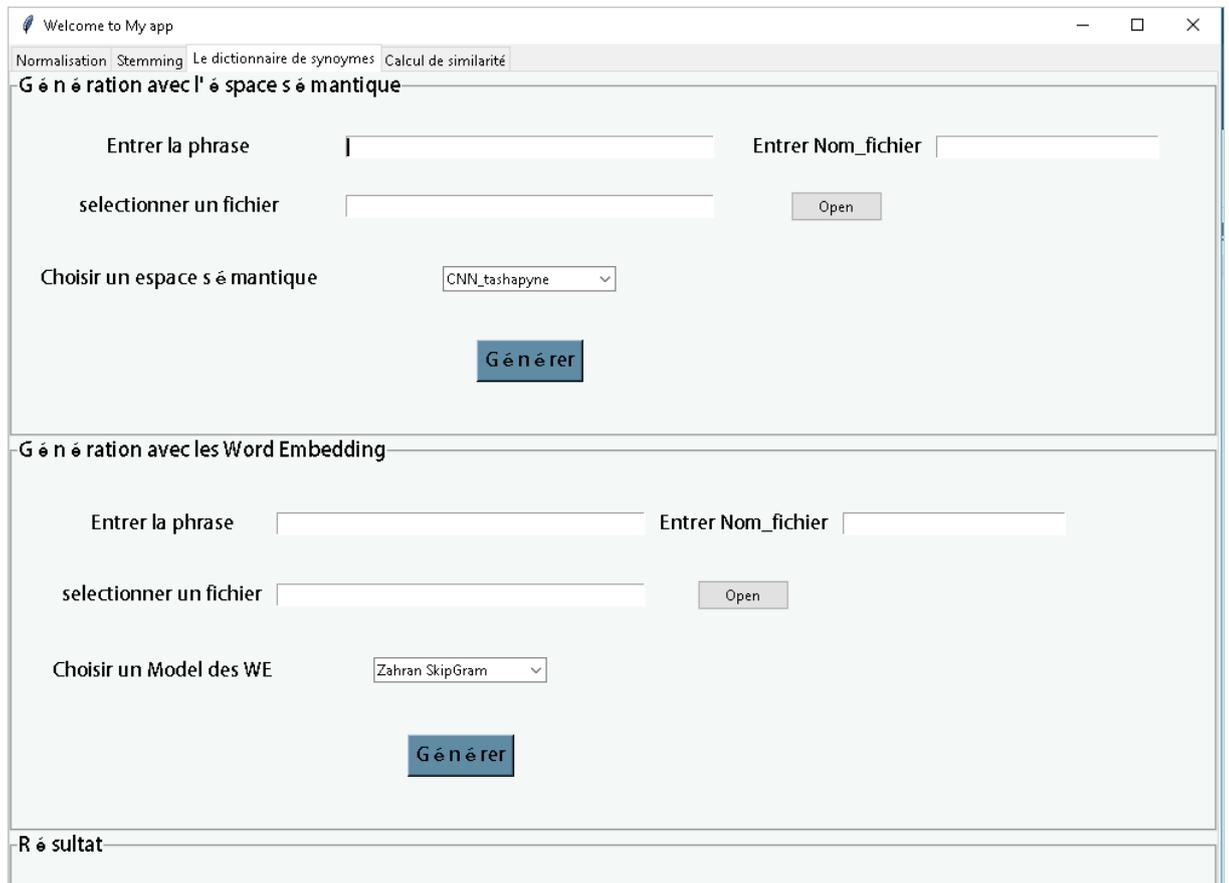


Figure IV-3: Outils de création du dictionnaire

4.3. Outils d'évaluation automatique des réponses courte

Cet outil est le noyau de notre travail « La figure IV - 4 », il nous permet de calculer la similarité entre la réponse modèle et la réponse étudiant en utilisant l'approche synonyme. Ses réponses (RE et RM) peuvent être saisie manuellement ou bien depuis des fichiers qui vont être choisi par l'utilisateur. Une fois les entrées saisies, l'outil apporte la possibilité du choix des options suivantes :

- 1- Le choix des modèles (l'approche) :

Modèle espace sémantique, matrice de profils : L'utilisateur peut choisir le type d'espace sémantique (CNN, Khaleej, BBC+CNN, Cybercrime)

Modèle de Word Embedding : L'application fournit l'utilisateur de quatre modèles au choix entre Zahran (CBOW, SkipGram) et Aravec (CBOW, SkipGram)

- 2- Sélection le dictionnaire des synonymes

3- Calcul de similarité : nous avons donné le choix à l'utilisateur entre

Le calcul matriciel CM sans Pondération

Le calcul matriciel CM avec Pondération

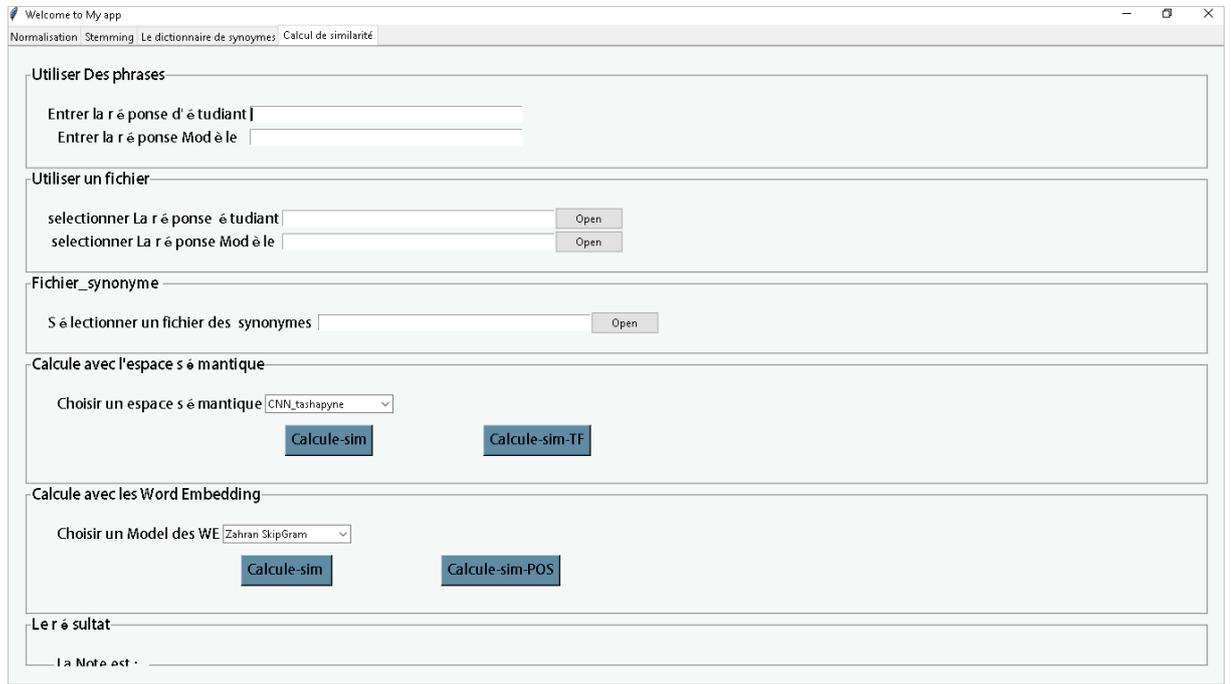


Figure IV-4: Outils d'évaluation automatique des réponses courtes

5. Ressources matérielles et logicielles utilisées lors du développement

Nous avons rencontré un problème matériel dans la phase principale de notre approche « la génération de l'espace sémantique et la création de la matrice de profils ». Ce dernier n'a pas abouti sa fin à cause de l'insuffisance de la mémoire de nos PC portables (4 Go de RAM). Pour cela, nous avons eu la chance de travailler sur un serveur à distance fournit par l'université de Bouira de 16 go de Ram .Nous travaillons avec le langage de programmation *Python 3.6* sur nos machines personnelles et sur le serveur à distance

Pour les interfaces, nous avons travaillé avec Tkinter, l'environnement de développement était avec spyder

6. Résultats et discussion

Dans la suite, nous présentons les résultats obtenus dans différentes perspectives:

Dimensionnalité de la matrice de profil et impact du stemming,

Les résultats des modèles de similarité proposés et de leurs variantes (en termes de pondération)

Hybridation avec les mesures syntaxiques

Evaluation par rapport aux résultats obtenus l'année passée

Les résultats des deux modèles de similarité proposés (espace sémantique et WE)

6.1. Résultat en utilisant les espaces sémantique Pour le modèle CM

Gomaa Dataset :

Le tableau 13 présente les résultats du modèle CM sans et avec synonymes pour data set Gomaa avec l'espaces sémantiques de CNN avec Tashapyne stemmer

		Corpus cnn			
Gomaa Dataset	Pondération	sans	TF	IDF	TF-IDF
résultat sans synonymes	RMSE	1.29	1.20	1.22	1.25
	CP	69.63	73.11	74.04	72.65
résultat avec synonymes	RMSE	1.17	1.36	1.14	1.37
	CP	80.03	77.99	79.81	78.25

Tableau 13: Résultats du modèle CM avec l'espace sémantique CNN Tashapyne stem

On remarque une amélioration de **10.4%** pour la méthode de base et une amélioration de **5.77%** pour le meilleur résultat (Cm avec pondération (IDF))

Le tableau 14 représente les résultats du modèle CM avec et sans pondération obtenus par l'espace sémantique CNN avec Tashapyne stemmer

		Corpus khaleej			
Gomaa Dataset	Pondération	sans	TF	IDF	TF-IDF
résultat sans synonymes	RMSE	1.28	1.14	1.19	1.20
	CP	70.58	73.94	75.17	73.64
résultat avec synonymes	RMSE	1.15	1.15	1.15	1.15
	CP	79.86	80.04	79.86	79.86

Tableau 14: Résultats du modèle CM avec l'espace sémantique Khaleej avec Tashapyne stem de cette année et celle de [16]

On remarque une amélioration de **9.28%** pour la méthode de base et une amélioration de **4.69%** pour le meilleur résultat (Cm avec pondération (IDF))

Le tableau 15 représente les résultats du modèle CM avec et sans pondération obtenus par l'espace sémantique CNN avec Khoja stemmer

		Corpus cnn			
	Pondération	sans	TF	IDF	TF-IDF
Gomaa	RMSE	1.14	1.13	1.22	1.19
	CP	80.28	80.64	78.79	79.68

Tableau 15: Résultats du modèle CM avec l'espace sémantique CNN avec khoja stem

Le meilleur résultat obtenu [16] est de : **76.5 %** il ya une amélioration de **4.14%**

		corpus bbc+cnn			
	Pondération	sans	TF	IDF	TF-IDF
Gomaa Dataset	RMSE	1.14	1.13	1.12	1.12
	CP	79.7	80.64	80.26	80.36

Tableau 16: Résultats du modèle CM avec l'espace sémantique BBC+CNN avec Khoja stem

Le meilleur résultat obtenu [16] est de : **75.92%** il ya une amélioration de **4.72%**

Le tableau 17 représente les résultats du modèle CM avec et sans pondération obtenus par l'espace sémantique Khaleej avec Khoja stemmer

		corpus Khaleej			
	Pondérat	sans	TF	IDF	TF-IDF
Gomaa Dataset	RMSE	1.14	1.15	1.14	1.14
	CP	79.2	79.22	79.62	79.45

Tableau 17: Résultats du modèle CM avec l'espace sémantique Khaleej avec Khoja stem

Le meilleur résultat obtenu [16] est de : **76.17%** il ya une amélioration de **3.45%**

Combinaison avec les mesures syntaxiques :

Une combinaison des deux matrices CM sémantique et CM syntaxique a été réalisée avec la mesure syntaxique DICE. Les résultats sont reportés dans le « Tableau 18 ».

		Corpus khaleej				
		Pondération	sans	TF	IDF	TF-IDF
Gomaa Dataset	RMSE	1.11	1.13	1.17	1.32	
	dice_max	80.036	80.02	80.87	79.09	
		RMSE	1.11	1.19	1.11	1.11
		dice_moyen	80.25	79.32	80.25	80.25

Tableau 18: Résultats du modèle CM avec l'espace sémantique Khaleej avec tashapyne stem combiné avec dice

CCASAG data set :

Le tableau 19 représente les résultats du modèle CM avec synonyme et sans synonymes obtenus par l'espace sémantique Cybercriminalité avec Tashaphyne stemmer

		Corpus Cyber Criminality			
Cyber Criminality dataset	Pondération	sans	TF	IDF	TF-IDF
résultat sans synonymes	RMSE	1.67	1.71		
	CP	55.38	53.91		
résultat avec synonymes	RMSE	1.35	1.32	1.38	1.36
	CP	66.34	66.59	65.78	66.73

Tableau 19: Résultats du modèle CM avec l'espace sémantique cybercriminalité avec Tashaphyne stem

Le meilleur résultat est obtenu avec tfidf, une amélioration de **11.35 %** par rapport au résultat de [48] qui est : **55.38%**

Le tableau 20 représente les résultats du modèle CM avec synonymes et sans synonymes obtenus par l'espace sémantique Cybercriminalité avec Khoja stemmer

	Corpus Cyber Criminality				
Cyber Criminality dataset	Pondération	sans	TF	IDF	TF-IDF
résultat sans synonymes	RMSE	1.4	1.46		
	CP	58.46	56.59		
résultat avec synonymes	RMSE	1.34	1.34	1.34	1.34
	CP	66.1	66.1	66.15	66.1

Tableau 20: Résultats du modèle CM avec l'espace sémantique cybercriminalité avec Khoja stem

Le meilleur résultat est obtenu avec tfidf, une amélioration de **7.69%** par rapport au résultat de [48] qui est **58.46%**

Le tableau 21 représente les résultats du modèle CM avec et sans pondération obtenus par l'espace sémantique Cybercriminalité avec Stanford_core_nlp stemmer.

		Corpus Cyber			
	Pondération	sans	TF	IDF	TF-IDF
cyber	RMSE	1.61	1.61	1.6	1.59
	CP	60.67	59.91	60.25	60.66

Tableau 21: Résultats du modèle CM avec l'espace sémantique cybercriminalité avec Stanford_core_nlp stem

On remarque avec le light stem stanford_core_nlp une diminution des résultats.

Combinaison avec les mesures syntaxiques :

Une combinaison des deux matrices CM sémantique et CM syntaxique a été réalisée avec la mesure syntaxique DICE. Les résultats sont reportés dans le « Tableau 22 ».

		Corpus cyber			
Pondération		sans	TF	IDF	TF-IDF
cyber	RMSE	1.35	1.32	1.37	1.38
	dice_max	66.18	66.75	66.58	65.76
RMSE		1.33	1.33	1.33	1.33
dice_moyen		67.87	67.87	67.87	67.87

Tableau 22: Résultats du modèle CM avec l'espace sémantique cybercriminalité avec tashapyne stem combiné avec dice

Le meilleur résultat est obtenu par la combinaison de Dice **67.87%**.

On déduit que l'approche de génération des synonymes avec l'espace sémantique a donné des résultats escompté, et que l'impact de stem a joué son rôle (les meilleurs résultats sont avec KHOja et Tashapyne Stem)

6.2. Résultat en utilisant les WE

Le data set de goma :

Le tableau 23 présente les résultats obtenus en utilisant les WE en appliquant la méthode de CM avec un stem léger

		CM	tfidfMinMax	POS	POS+TfMinMax
Zahran	CBow	CP: 78.30 RMSE:1.24	CP :78.51 RMSE :1.21	CP :78.61 RMSE :1.17	CP :78.61 RMSE :1.17
	SkipGram	CP: 75.59 RMSE: 1.13	CP :74.59 RMSE :1.20	CP:74.30 RMSE :1.15	CP:75.01 RMSE :1.33
Aravec	CBow	CP: 77.55 RMSE: 1.38	CP:78.14 RMSE:1.28	CP:77.87 RMSE:1.24	CP:77.87 RMSE:1.24
	SkipGram	CP: 77.47 RMSE: 1.29	CP:77.28 RMSE:1.24	CP:77.29 RMSE:1.25	CP:77.29 RMSE:1.25

Tableau 23: Résultat du modèle CM en utilisant les WE en avec stanford stem

Le WE de zahran CBOW avec le part of speech a donné un léger accroissement des résultats par rapport à [34]

Combinaison avec les mesures syntaxiques :

Le tableau 24 présente les résultats des combinaisons avec les mesures syntaxiques
DICE

		Dice	Dice + TfIdfMinMax	Dice + Pos+TfIdfminmax	Dice Moy + POS
Zahran	CBow	CP :79.90 RMSE :1.31	CP :78.61 RMSE :1.17	CP :79.78 RMSE :1.17	CP: 79.78 RMSE: 1.17
	SkipGram	CP:78.33 RMSE :1.16	CP:78.10 RMSE :1.15	CP :77.38 RMSE :1.21	CP:77.63 RMSE :1.20
Aravec	CBow	CP:78.85 RMSE :1.38	CP:79.66 RMSE :1.13	CP :79.003 RMSE :1.30	CP: 79.18 RMSE :1.30
	SkipGram	CP:79.20 RMSE :1.15	CP :79.94 RMSE :1.13	CP: 79.03 RMSE :1.23	CP: 79.03 RMSE :1.23

Tableau 24: Résultat du modèle CM en utilisant les WE Combiner avec Dice

Le meilleur résultat est avec WE d'Aravec SkipGram dont une augmentations de 2% par rapport [34].

On constate que l'approche de génération des synonymes avec les WE n'a pas donné une grande amélioration par rapport à [34]

CCASAG data set:

Le tableau 25 présente les résultats obtenus en utilisant les WE en appliquant la méthode de CM avec un stem léger.

		CM	CM *tfIdfMinMax	CM*POS	CM*POS*TfMinMax
Zahran	CBow	CP: 66.59 RMES :1.33	CP :66.04 EQ :1.44	CP :62.80 EQ :1.40	CP :62.80 EQ :1.40
	SkipGram	CP: 65.36 EQ:1.14	CP :64.88 EQ :1.41 CP:63.60 EQ:1.30	CP:61.13 EQ :1,56	CP: 61.13 EQ :1.30
Aravec	CBow	CP: 63.60 EQ: 1.23	CP:63.92 EQ:1.31	CP:60.55 EQ:1.33	CP: EQ:
	SkipGram	CP: 65.30 EQ: 1.23	CP:65.24 EQ:1.42	CP:61.85 EQ:1.20	CP:61.85 EQ: 1.20

Tableau 25: Résultat du modèle CM en utilisant les WE en avec stanford stem

Le meilleur résultat était avec le modèle de ZahranCBOW sans pondération

Combinaison avec les mesures syntaxiques :

Le tableau 26 présente les résultats des combinaisons avec les mesures syntaxiques
DICE

		CM*Dice Moy	CM*Dice Moy * TfMinMax	CM*Dice Moy *POS	CM*Dice Moy *POS
Zahran	CBow	CP :65.85 EQ: 1,47	CP :67.18 EQ :1.38	CP:65.03 EQ:1.35	CP:65.03 EQ:1.35
	SkipGram	CP:64.72 EQ:1.16	CP:67.92 EQ:1.23	CP:64.07 EQ:1.47	CP:64.07 EQ:1.47
Aravec	CBow	CP:63.49 EQ:1.28	CP:65.49 EQ:1.26	CP: 63.53 EQ: 1.51	CP: 63.53 EQ: 1.51
	SkipGram	CP:65.16 EQ: 1.23	CP:65.81 EQ:1.25	CP: 64.34 EQ: 1.39	CP: 64.34 EQ: 1.39

Tableau 26: Résultat du modèle CM en utilisant les WE Combiner avec Dice

Le meilleur résultat était avec le modèle de ZahranCBOW et Zahran SkipGram combiner avec la mesure syntaxique Dice avec le TFMINMAX

6.3. Résultat en utilisant l'espace sémantique Pour le modèle SV

Le tableau 27 présente les résultats de [16] et de cette année pour le modèle SV sur l'espace sémantiques Khaleej avec Stem Tashapyne, une comparaison sera faite .le meilleurs résultats obtenus c'était avec l'espace Khaleej StemTasahpyne, en utilisant les paraphrases.

corpus kaleej		
Cyber Criminality dataset	Pondération	sans
résultat sans synonymes	RMSE	1.28
	CP	70.58
résultat avec synonymes	RMSE	1.11
	CP	77.18

Tableau 27: Résultats du modèle SV avec l'espace sémantique Khaleej Tashapyne stem

On remarque une amélioration de **6.6%** par rapport au résultat de [16] qui était **70.58%**

Le tableau 28 représente les résultats du modèle Sven utilisant la moyenne pour le data set Goma obtenus par l'espace sémantique Khaleej avec Tashaphyne stemmer.

corpus kaleej		
Cyber Criminality dataset	Pondération	sans
résultat sans synonymes	RMSE	1.28
	CP	70.58
résultat avec synonymes	RMSE	1.45
	CP	66.38

Tableau 28: Résultats du modèle SV en utilisant la moyenne avec l'espace sémantique Khaleej Tashapyne stem

On remarque une diminution de **3.8%**, par rapport à [16]

Le tableau 29 représente les résultats du modèle SV en utilisant le max pour le data set Goma obtenus par l'espace sémantique Khaleej avec Tashaphyne stemmer.

corpus kaleej		
Cyber Criminality dataset	Pondération	sans
résultat sans synonymes	RMSE	1.28
	CP	70.58
résultat avec synonymes	RMSE	1.38
	CP	61.14

Tableau 29: Résultats du modèle SV en utilisant le max avec l'espace sémantique Khaleej Tashapyne stem

Nous avons déduit une diminution de **9.44%**, par rapport à [16]. Nous avons constaté que la méthode SV en utilisant la moyenne et le max n'a pas

donné des bon résultats, par contre le modèle SV en utilisant les paraphrases à améliorer les résultats de [16] sauf l'inconvénient c'est le temps énorme consommé dans le traitement.

En conclusion : l'approche WE n'a pas beaucoup amélioré les travaux de [34], par contre la génération du dictionnaire à présenter une bonne performance des synonymes. Après les tests et les travaux de l'approche espace sémantique les résultats obtenus sont bons.

chapitre V. Conclusion et Perspectives

L'évaluation automatique des questions à réponses courtes pour la langue arabe est un problème difficile à cause du manque considérable d'outils de TAL et de ressource, pour cela nous avons axé notre recherche sur la création d'un corpus des réponses modèle dans l'évaluation automatique des réponses par composition à base de synonymes qui va être générée en utilisant l'approche espace sémantique (création de la matrice de profil) ou bien l'approche word embedding. Par cette élaboration l'enseignant serait déchargé de la création d'un corpus de réponses modèles possibles.

Notre travail se situe à la conjoncture de plusieurs domaines, ce qui nous a permis de présenter, dans ce mémoire, un état de l'art sur les systèmes d'évaluation automatique des réponses courtes, sur les approches de similarité des textes courts.

Nous avons ensuite proposée une approche basée sur les synonymes permettant la génération automatique d'un corpus de plusieurs réponses modèles.

Nous avons ensuite proposé et comparé notre approche par rapport aux approches existantes. En résultats un système d'évaluation automatique a été développé basé sur l'impact des synonymes qui se compose de plusieurs modules de calcul de similarité en utilisant les Word Embeddings (Zahran et AraVec) ou bien en utilisant l'espace sémantique. Le système a été combiné à d'autres approches syntaxiques et sémantiques. Il retourne une nette amélioration des résultats déjà obtenus en considérant une seule réponse modèle.

Nous avons évalué notre système avec deux data sets, et nous avons obtenu de meilleurs résultats par rapport aux systèmes des travaux déjà réalisés sur le Dataset CairoUniversity (Gomaa) et le Dataset Cyber Criminality Arabic Short Answer Grading. Sachant que les stemmers et les corpus sont les seules ressources dépendantes de la langue utilisées dans notre approche. En perspectives :

- ✓ Intégrer notre système d'évaluation dans une plateforme de e-learning pour tester son comportement à une échelle réelle
- ✓ Explorer les algorithmes de deep learning pour la génération de corpus de réponses modèles pour tenter d'améliorer encore les résultats.

Références Bibliographiques

- [1] W. Bakari, P. Bellot, and M. Neji, "Towards an Automatic Text Comprehension for the Arabic Question- Answering : Semantic and Logical Representation of Texts," no. December, pp. 1–3, 2018.
- [2] J. B. Lovins, « Development of a stemming algorithm », *Mech. Transl. Comput. Linguist.*, vol. 11, no June, p. 22-31, 1968.
- [3] S. Burrows, I. Gurevych, and B. Stein, *The Eras and Trends of Automatic Short Answer Grading*, vol. 25. 2015.
- [4] Singh, P. (1997). Review Essay: Basil Bernstein (1996). *Pedagogy, symbolic control and identity*. London: Taylor & Francis. In *British Journal of Sociology of Education*,18(1), pp.119-124.
- [5] Leacock, C. and Chodorow, M. (2003). C-rater: Automated Scoring of Short-Answer Questions. *Computers and the Humanities*,37(4), 389–405.
- [6]Dezso Sima, Balazs Schmuck, Sandor Szollosi, & Arpad Miklos. (2007). Intelligent short text assessment in eMax. *AFRICON 2007*.
- [7]Jordan,S.andMitchell,T.(2009).e-AssessmentforLearning? ThePotentialofShort-AnswerFree-TextQuestions with Tailored Feedback. *British Journal of Educational Technology*,40(2), 371–385.
- [8] Bukai, O., Pokorny, R., and Haynes, J. (2006). An Automated Short-Free-Text Scoring System: Development and Assessment. In *Proceedings of the Twentieth Interservice/Industry Training, Simulation, and Education Conference*, pages 1–11. National Training and Simulation Association.
- [9] Mohler, Michael Mihalcea, Rada (2004) *Text-to-text Semantic Similarity for Automatic Short Answer Grading*.
- [10] Hall, M., Frank, E.,Holmes ,G., Pfahringer, B.,Reutemann,P.,andWitten, I.H.(2009). *TheWEKADataMining Software: An Update*. *SIGKDD Explorations*,11(1), 10–18.
- [11] Bailey, S. and Meurers, D. (2008). Diagnosing Meaning Errors in Short Answers to Reading Comprehension Questions. In J. Tetreault, J. Burstein, and R. De Felice, editors, *Proceedings of the Third ACL Workshop on InnovativeUseofNLPfor BuildingEducationalApp*.
- [12] Madnani, N., Burstein, J., Sabatini, J., and Reilly, T. O. (2013). Automated Scoring of a Summary Writing Task Designed to Measure Reading Comprehension. In J. Tetreault, J. Burstein, and C. Leacock, editors, *Proceedings ofthe Eighth Workshop onInnovative*.
- [13] E. Negre, "Comparaison de textes: quelques approches...", 2013. hal-00874280f

- [14] [Baeza-Yates and Ribeiro-Neto, 1999] Baeza-Yates, R. A. and Ribeiro-Neto, B. (1999). *Modern Information Retrieval*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.
- [15] [Jaccard, 1901] Jaccard, P. (1901). Étude comparative de la distribution orale dans une portion des alpes et des jura. *Bulletin del la Société Vaudoise des Sciences Naturelles*, 37 :547579.
- [16] ATOUB Yasmine, BENAYAD Asma « Mesures de similarité sémantique pour un système d'évaluation automatique des réponses courtes : Application à la langue arabe». Mémoire master USDB 1. 2018/2019.
- [17] D. L. T. Rohde, L. M. Gonnerman, et D. C. Plaut, « An Improved Method for Deriving Word Meaning from Lexical », *Cogn. Psychol.*, vol. 7, p. 573-605, 2004.
- [18] W. H. Gomaa et A. A. Fahmy, « A Survey of Text Similarity Approaches », *Int. J. Comput. Appl.*, vol. 68, no 13, p. 13-18, 2013.
- [19] R. Mihalea, C. Cole, ad C. “tappaaa, Copus-based and Knowledge-based Measures of Tet “eati “iilait, pp. 775–780, 2005.
- [20] M. Naili, A. H. Chaibi, dan H. H. Ben Ghezala, “Comparative Study of Word Embedding Methods in Topic Segmentation,” *Procedia Computer Science*, Vol. 112, hal. 340–349, 2017.
- [21] D. S. El Moatez Billah Nagoudi, Jérémy Ferrero, « LIM-LIG at SemEval-2017 Task1: Enhancing the Semantic Similarity for Arabic Sentences with Vectors Weighting », in *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval2017)*, no June, p. 134-138, 2017
- [22] B. Furlan, V. Batanović, et B. Nikolić, « Semantic similarity of short text in languages with a deficient natural language processing support », *Decis. Support Syst.*, vol. 55, no 3, p. 710-719, 2013.
- [23] R. Mihalcea, C. Corley, et C. Strapparava, « Corpus-based and knowledge-based measures of text semantic similarity », *Proc. 21st Natl. Conf. Artif. Intell.*, vol. 1, p. 775-780, 2006.
- [24] A. Islam et D. Inkpen, « Semantic text similarity using corpus-based word similarity and string similarity », *ACM Trans. Knowl. Discov. Data*, vol. 2, no 2, p. 1-25, 2008.
- [25] M. Zahran, A. Magooda, A. Mahgoub, H. Raafat, M. Rashwan, et A. Atyia, « Word Representations in Vector Space and their Applications for Arabic ». 2015.
- [26] W. H. Gomaa et A. A. Fahmy, « Arabic Short Answer Scoring with Effective Feedback for Students », *Int. J. Comput. Appl.*, vol. 86, no 2, p. 35-41, 2014.

- [27] P. Kolb, « Disco: A multilingual database of distributionally similar words », Proc. KONVENS-2008, Berlin, no 2003, p. 37-44, 2008.
- [28] A. Magooda, M. A. Zahran, M. Rashwan, H. Raafat, et M. B. Fayek, « Vector Based Techniques for Short Answer Grading », Proc. Twenty-Ninth Int. Florida Artif. Intell. Res. Soc. Conf. Val. Spain, p. 238-243, 2016.
- [29] E. M. Billah Nagoudi et D. Schwab, « Semantic Similarity of Arabic Sentences with Word Embeddings », Proc. Third Arab. Nat. Lang. Process. Work., p. 18-24, 2017.
- [30] Madnani, Nitin ,Dorr, Bonnie J Generating Phrasal and Sentential Paraphrases : A Survey of Data-Driven Methods , November 2009 .
- [31] Koehn, Philipp, Franz Josef Och, and Daniel Marcu. 2003. Statistical phrase-based translation. In Proceedings of HLT-NAACL, pages 48–54, Edmonton.
- [32] Brown, Peter F., John Cocke, Stephen Della Pietra, Vincent J. Della Pietra, Frederick Jelinek, John D. Lafferty, Robert L. Mercer, and Paul S. Roossin. 1990. A Statistical Approach to Machine Translation. Computational Linguistics, 16(2):79–85.
- [33] J. Berant and P. Liang, “Semantic Parsing via Paraphrasing,” no. Figure 1, 2013.
- [34] ABDALLAH Amina, GAROUDJA Khadidja Mesures de similarité syntaxique pour un système d'évaluation automatique des réponses courtes : Application à la langue arabe. Mémoire master USDB 1. 2018/2019
- [35] HANNOUFI Mohammed , Hamza HENNICHE Adel Nassim «Les Word-Embedding pour l'évalutaion automatique des réponses courtes en apprentissage en ligne : Application à la langue arabe ». Mémoire master USDB 1. 2018/2019
- [36] Zahran, Mohamed & Magooda, Ahmed & Mahgoub, Ashraf & Raafat, Hazem & Rashwan, Mohsen & Atyia, Amir. Word Representations in Vector Space and their Applications for Arabic, (2015).
- [37] « Zahran Word Embedding » disponible sur : <https://drive.google.com/drive/folders/1lpzvYO8K8CEu-e69nVKWRWVmzT8yZE0A> (Dernier accès le 27/07/2019)
- [38] Mohammad, Abu Bakr & Eissa, Kareem & El-Beltagy, Samhaa. (2017). AraVec: A set of Arabic Word Embedding Models for use in Arabic NLP. Procedia Computer Science, 117. 256-265.
- [39] <https://github.com/bakrianoo/aravec/tree/master/AraVec%201.0> (Dernier accès le 27/07/2019)
- [40] S. R. Bowman, G. Angeli, C. Potts, et C. D. Manning, « A large annotated corpus for learning natural language inference », 2015

[41] Patrick Paroubek, “Etiquetage morphosyntaxique”,2006 [Enligne] disponible sur : http://www.technolangue.net/imprimer.php3?id_article=296

[42]”Stanford CorNLP – Natural language software” [Enligne] disponible sur: <https://stanfordnlp.github.io/CoreNLP/>

[43] J. Macqueen, « Some methods for classification and analysis of multivariate observations », Proc. Fifth Berkeley Symp. Math. Stat. Probab., vol. 1, no 233, p. 281,297, 1967.

[44] « ckmeans » [Enligne]. <https://github.com/llimllib/ckmeans> (Dernier accès le 20/09/2019)

[45] William H Greene, Econométrie, Paris, Pearson Education, 5e éd. (ISBN 978-2-7440-7097-6) , 2005.

[46] Jacob Cohen, Statistical power analysis for the behavioral sciences (2nd ed.). (1988).

[47] Gomaa, W. H., & Fahmy, Automatic scoring for answers to Arabic test questions. Computer Speech & Language. Elsevier Ltd 2013.

[48] Gomaa, W. H., &Fahmy, Arabic Short Answer Scoring with Effective Feedback for Students. International Journal of Computer Applications (0975 – 8887) Volume 86 – No 2, January 2014.

[49] Oukina faiza Radia Amar setti imene « Elaboration d’un corpus de test pour un système d’évaluation automatique des réponses courtes ». Mémoire master USDB 1. Soutenue le 30/09/2019.

Annexes

Les synonymes utilisés pour L'évaluation :

	Sketch engine	Aravec CBOW	Aravec SkipGram	Zahran CBOW	Zahran SkipGram	ES Khaleej
تتناول	تناولت يتناول تتطرق تناولتها تناقش تعالج تناول تتمحور تطرق تركز					تشملها تتناول الإصلاحات ويعيق ضعفها ويوصي ومثل ترقب كسبته تترايد
مكونات	مكونة المكونات للمكونات تركيبية كمكون شراخ المكون المتجانس اطياف تكويناة	عناصر المكونات تركيبات خصائص مكوناتها بروتينات مكوناته مميزات مشتقات جزئيات	مكوناته المكونات ومكونات كمكونات عناصر مكوناتها بمكونات ومكوناته لمكونات مواد	المكونات ومكونات لمكونات مكوناتها مكوناته بمكونات المكونات_الاساسيه خصائص تركيبه عناصر_ومكونات	ومكونات لمكونات المكونات مكوناته والمكونات بمكونات مكوناتها المكونات_الاساسيه ومكوناته للمكونات	
حيه	حيها العدنان اوحاه مريبه اصطفاه ه كليم يوحى النبوه ينطق	بدائيه غريبه نقيه بشريه ناضجه ميته مشوهه انثويه حقيقه	كائنات الحيه لكائنات تدبيه بويضات ميكروبيه ومخلوقات ادميه بكائنات طقوسيه	ميته الحيه الثوره_حقتضل وحي_زوله مربيه ويندوس_هوتميل تلد_الحيه فحصت_خارجا وهادفه_ومباشره تدفن_ابنتها	تصنيف_مستحثات ميته الحيه اناس_وحيوانات مخاليق_حيه الالमित طقس_معلومه امفيبين اناشيط دومستيزيرت	
الاطار	للاطار اطارها السياق المنطلق المنحي الصدد المسار اطار منظورها	السياق النموذج النهج المسار التصور المفهوم المنظور المنحي المعيار	والاطار السياق للاطار النموذج اطار اطاره اطارها المفاهيمي	السياق والاطار اطار اطارها المسار الإطار للاطار المجال اطاره	والاطار للاطار اطار بالاطار كالاطار اطارها السياق ارساءه لاطارها	لشركته الاطار بالمستقبل الهادف بمبادره كلوور تيري المستهلكون سبببت

	اختص اجناسها قاطبة	بجميع غيره بمختلف	وباقي مختلف والامصار	جميع كبقية ببقية	بهذه_المزیه والخصیصه وانحائه	سهمی وغلفا القطاعیه
تطبيق	تطبيقها تطبيقه تطبق التطبيق تنفيذ اعتماد يطبق تطبيقا تقنين تفعيل	تطبيقها تعميم تطبيقه تبسيط تمرير تنفيذ اعتماد تضمين استخدام لتطبيق	لتطبيق بتطبيق وتطبيق التطبيق تطبيق تنفيذ تطبيقه والتوازنات استخدام	تنفيذ تطبيقها لتطبيق وتطبيق التطبيق بتطبيق بتطبيق تطبيق اعتماد استخدام	وتطبيق لتطبيق تطبيقها بتطبيقه فنتطبيق التطبيق وتطبيقها الالتزام_بتطبيق	تطبيق الماسسي والاجتهاد والانضباط الفقه وكفايه يستغرق بالمشاركة عافيته فاذا
المعلومات	للمعلومات البيانات المعلومة معلومات المعلومات الوثائق معلوماتية الحاسوبية الحوسبة	البيانات للمعلومات المعطيات معلومات للبيانات الخبرات المعلومه المعرفه الملفات المستندات	البيانات معلومات للمعلومات والبيانات والمعلومات الببليوغرافيه الاتصالات المعطيات واسترجاعها الببليوجرافيه	البيانات المعلومات المعلومات_والبيانات المعلومات_التفصيليه المعلومات للمعلومات	والمعلومات المعلومات للمعلومات معلومات المعلومات_والبيانات المعلومات المعلومات_البيانات_والمعلومات المعلومات_التفصيليه	المعلومات وتقنيه فتوفير بارترز وادوات ندره والموارد وتكنولوجيا والتعرف بتبادل
المجالات	مجالات مجالة الاصعدة مجال الخبرات المناحي الصعد المجالين القطاعات الشركات	التخصصات الاختصاصات القطاعات مجالات المجالات الميادين مجالاته النواحي السياقات الجوانب	مجالات والمجالات التخصصات بالمجالات القطاعات مجالاتها والتخصصات مجالاتهم للمتخصصين مجال	مجالات والمجالات مختلف_المجالات كاهه_المجالات الميادين القطاعات شتى_المجالات بالمجالات مختلف_الميادين	مجالات مختلف_المجالات والمجالات كاهه_المجالات شتى_المجالات بالمجالات مختلف_الميادين كالمجالات	المجالات الصغيرتين واشدها المدينتين تنافسا التسال وسيحطم يبقى مختلف الملاحه
المعرفيه	الفكریه العلمیه النظریه المعرفیه المعرفیه المعرفیه الثقافیه الروحیه الاخلاقيه المعرف	الادراكيه السلوكيه الذهنيه الدلالیه الوظيفیه المؤسسيه الحياتيه الحسيه اللغويه الموضوعيه	الادراكيه والادراكيه والمعرفيه والسلوكيه والحسيه والانفعاليه والادراك والذهنيه النمائيه والمنطقيه	المعرفيه الفكریه اللغويه الفكریه_والمعرفيه المعرفيه_والعلميه الفكریه_والفلسفیه العلميه_والمعرفيه المعرفيه_والثقافيه الابستمولوجيه الثقافيه_والمعرفيه القيميه	والمعرفيه والفكریه_والمعرفيه المعرفيه_والمنهجيه المعرفيه_والقيميّه المعرفيه_والثقافيه معرفيه اللغويه_والمعرفيه والعمليات_الذهنيه المنهجيه_والنظريه	
الفيزيائيه	الفيزيائيه النظرية الطبع الاشع	الجزئيه التרכيبيه الكمويه التحليليه	والفيزيائيه والكيميائيه الفيزيائيه والجزئيه	الفيزيائيه الفيزياويه الكمويه الرياضيه_والفيزيائيه	الفيزيائيه الفيزيائيه الفيزيائيه والفيزيائيه الذريه_والجزئيه	

	فيزيائية نظريه الجزئيات الكهر ومغناطيسية الجسيمات الكهر بائيه	البيولوجيه الميكانيكيه الفراغيه والكيميائيه الشكليه التشريحيه	النانومتريه الفيزيولوجيه والكهر ومغناطيسيه الكيميائيه اللاخطيه الكسيرييه	الفيزيقيه الكهر ومغناطيسيه الفيزيائيه_والكيميائيه الفيزيائيه الجسيميه الكهرطيسيه	والديناميكا_الحراريه الميكانيكا_الاحصائيه الفيزيائيه_والكيميائيه الديناميكيه_الحراريه الكهر بائيه_والمغناطيه سيه	
الكيميائيه	الكيميائيه الادويه الحيويه الكيميائيه الكيميائيه الكيميائيه الكهر بائيه النوويه الطبيه الغذائيه	الكيميائيه الكيميائيه الكيميائيه المؤكسده الهيدروكربونيه الحفازه الانشطاريه الكيميائيه النسجيه المتطابره	الكيميائيه الكيميائيه الكوريه الكيميائيه الحفازه المخصبات المكلوره والكيميائيه والمذيبات الخلاط	الكيميائيه الكيميائيه الكيميائيه والكيميائيه الكيميائيه والكيميائيه والكيميائيه اللاعضويه الكيميائيه_والفيزيائيه	الكيميائيه الكيميائيه الكيميائيه والكيميائيه المركبات_الكيميائيه والكيميائيه المواد_الكيميائيه والكيميائيه	
يتخذ	يتخذ يتخذها اتخذ اتخاذ يتخذوا تتخذ يتخذون اتخذوا تتخذ متخذوا	ياخذ يتخذ اتخذ تتخذ اتخذ يتخذ يسلك	يتخذ اتخذ يتخذ تتخذ ويتخذ اتخذ يتخذها واتخذ ياخذ سيتخذ	تتخذ اتخذ ويتخذ يتخذ سيتخذ تتخذ يتخذ يتخذها وسيتخذ اتخذ	ويتخذ اتخذ يتخذ تتخذ يتخذها اتخذ يتخذ فيتخذ ليتخذ اتخذ	يتخذ الفولاذيه القضبان مترى هيكل وسلمت بخصوص المكتتب بنسبته اقلها
كافيه	الكافيه قهو ماركت مجانيه ممتاز مخصصه شوب الاضاف صحيه كافي	معقوله ضروريه تكفي فعاله مناسبه مطلوبه محدوده الكافيه ملائمه	تكفي الكافيه فائضه معقوله اللازمه وكافيه يكفي كافي ضروريه مطلوبه	تكفي الكافيه ضروريه كاف محدوده_للغايه ملائمه واقيه يكفي معقوله اضافيه	وكافيه الكافيه تكفي كافيه_ومناسبه مناسبه_وكافيه المناسبه_والكافيه واقيه كاف ملائمه يكفي	كافيه قيامنا بجهود ولفتره ووقعا اتفاقات رسوخ بتخزين لتضمن ومراجعه
السلبيه	الايجابيه سلبيه ايجابيه النفسيه الانانيه السلبيه العصبيه الاخلاقيه الحقيقيه الفكريه	الايجابيه المحتمله الضاره الخاطئه الجسديه الضائره السلوكيه المترتبه السبئه الجائيه	الايجابيه والسلبيه الضائره والعواقب المحتمله والايجابيه الانسحابيه والضاره سلبيه المؤذيه	السالبه السيئه والسلبيه المحبطه الايجابيه_والسلبيه سلبيه الاجابيه الغير_مرغوبه الخاطئه	السلبيه والسلبيه سلبيه والمشاعر_السلبيه والسلوكيات_السلبيه الاجابيه الجوانب_السلبيه العادات_السلبيه المحبطه للجوانب_السلبيه	السلبيه والتقارير توترات والتوترات تطليح التاثيرات لاحتواء وجني

						تحوم وتعيش	
النهر	نهر للنهر البحر ضفتيه جيجون شط الضفاف بحيرة ضفاف ضفة	نهر نهري مجراه الوادي الانهار السييل المضيق الدانوب مياهه	مصب مياهه	نهر نهري جيجون ونهر سيجون اورينوكو دجله وروافده براهاموترا بنهر	نهر للنهر مجرى_النهر مياه_النهر نهر_الفرات الانهار_نهر_دجله البحيره ضفه_النهر مصب_النهر	للنهر نهر بالنهر مجرى_النهر مصب_النهر الانهار قنهر والنهر لنهر ضفه_النهر	
الصحراء	صحرائ الصحراوية الصحاري صحرائنا صحراء السهوب صحراوية القاحلة الواحات الصحاري	بالصحراء والصحراء الواحات للصحراء السهول صحراء الصحاري الباديه السهوب	الجبال	والصحراء صحراء للصحراء بالصحراء الواحات الصحراويه كالاهاري السهول وصحراء القاحله	صحراء للصحراء الصحاري الصحراء_الغريبيه الصحاري بالصحراء والصحراء رمال_الصحراء الصحراء_الكبرى الصحراويه	بالصحراء الصحراويه للصحراء والصحراء الصحاري صحراويه الصحاري الصحراء_القاحله رمال_الصحراء	الصحراء لتختفي رمال واقعيه ستعود سارعت وولاء ميسان كبرت صاحبها
محدود	ضئيل محدودة المحدود كبير محدودا محصور اسع متنام مقتصر محدد	ضئيل قليل كافي كاف معقول محدد لانهايتي متساو منته معتبر	قليل ضئيل محدوده محدد ومحدود محدودا كافي معنوم كبير لانهايتي	محدوده ضئيل المحدود محدود_للغايه محدد كبير محدودا ضئيل_جدا قليل هائل	محدودا محدود المحدود محدود_للغايه ضئيل لانهايتي محدد ومقتصر ضئيل_جدا	محدود يتقدم وصعود اسبوعها تنهي طفيف بقطر جاوزت والامارات اخلاقه	
حيز	الحيز نطاق حيزها فضاء النفاذ نطاقها اطارها المتعلق اطار موفى	النفاذ التنفيذ نطاق فراغ اطار الفعلي فعليا افق لنطاق معيار	التنفيذ النفاذ ستدخل شنغن نطاق وتضعها بنفاذ كقانون التطبيق اطار	لحيز وحيز الحيز حيزا نطاق فضاء مساحه بحيز نطاق الحيز_الضيق	الحيز وحيز لحيز بحيز والحيز حيزا الحيز_الضيق للحيز حيز_مكاني حيز_ضيق	حيز التنفيذ ودخلت ودخل وحدته مصادقه سيستكمل اختتمت قانونا لمبادرات	
تبادلات	مبادلات التبادلات متبادلة علاقات تبادلية تبادل	تأثرات ارتباطات تعاملات لعلاقات تمثيلات وعميقة روابط	فروق	لصراعات متضافره وصفقات انتلافات والموردين تبدلات	وتبادلات اجراء_تبادلات تبادل التبادلات تبادلات_متكرره تبادلات	وتبادلات بتبادلات التبادلات اجراء_تبادلات القيام_بتبادلات تبادلات_متكرره اجريا_تبادلات	

	المبادلات التبادل اتفاقات	خلفيات	فالعلاقات روابطها متصارعه وبوسائل	مبادلات القيام_بتبادلات واجراء_تبادلات تبادلات_تجاريه	للتبادلات المتكرر_للزيارات للزيارات_رفيعه	
يتعلق	يخص تتعلق المتعلقة المتعلق مايتعلق خصوص متعلقة يختص يرتبط المرتبطه	يخص يختص يتصل تتعلق المتعلق يرتبط مرتبط المتعلقه	يختص المتعلقه يخص تتعلق والمتعلقه المتعلق ويتعلق مايتعلق متعلق فيما	فيما_يتعلق تتعلق متعلق يختص المتعلق وسواء_تعلق يرتبط مرتبط فيما_يختص المتعلقه	يختص مايتعلق تتعلق يرتبط المتعلقه فيما_يتعلق متعلق المتعلق ويتعلق يتعلق	يتعلق فيما بالقانون بموضوع مناقشتها وفيما بتعطيل الحوكمه بالقضايا بتوفيق
يصف	يفصف وصف يصفها صف يمتدح ينعت يوصف يسميه يرى يصور	ويصف يشرح يلخص وصف يوضح ووصف يسرد عالج يحلل يتناول	ويصف وصف تصف يشرح يلخص ويوضح يشير يوضح يصفه ووصف	ويصف وصف يصفها تصف يلخص يصفه يفصف ووصف بصفون فوصف	ويصف وصف يصفها يفصف تصف يصفه يلخص ووصف ليصف ويصفه	يصف صرخه الكاتب الشانك المقال طبيي مات مصاص يدعي واتهم
نظام	نظامي انظمة نظام النظام ديكتاتوري دكتاتوري استبدادي نظامها قمعي الانظمة	لنظام بنظام ونظام النظام كنظام انظمه معيار جهاز ميدا نموذج	لنظام ونظام النظام انظمه بنظام كنظام مولنكس وانظمه بيونكس نظامها	ونظام لنظام انظمه بنظام كنظام النظام نظام فنظام نظام_محواسب نظم	ونظام لنظام بنظام كنظام فنظام النظام انظمه نظام نظام ولنظام	نظام متشدد وتشغيل القطبيه الاحادي يستند تنجح تركيب عالمي باعتماد
التوازن	توازن للتوازن التناغم المتوازن الاتزان الاختلال الانسجام التماسك توازنها	توازن الاتزان الترابط التماسك التناسق التكامل التناظر التناغم التجانس	اتوازن الاتزان للتوازن الهيدروستاتيكي الاستتباب اتزان والتوازن الترموديناميكي لاختلال وبخاره	توازن التوازن للتوازن الاتزان تحقيق_التوازن التناغم بالتوازن الانسجام توازن	توازن التوازن والتوازن تحقيق_التوازن للتوازن فالتوازن بالتوازن اختلال_التوازن وتحقيق_التوازن لتحقيق_التوازن	التوازن المنظم لاختلال الضبط بالتخطيط وتغير يظل لاستعاده وقدر يتطلب
مستقره	مستقر مستقرة المستقر جنه حياه	متوازنه نشطه متجانسه محسوسه اعتياديه متماسكه	المستقره مستقر ومستقره متاينه متبلوره	مستقر مستتبه ومستقره هادئه ومستقره مستقره_نسبيا مقلبه	مستقر ومستقره مستقره_نسبيا مستقره المستقره جيده_ومستقره	مستقره وتنخر هشه انظمتها والنزاعات

	مطمئن صحيه مطمئنة يستقر حرجه	المستقره مر غوبه معتاده مجديه	مثاره هشه استقرارا خامله	مطمئنه مضطربه سليمه جيده_ومستقره	مستقره_وجيده متدهوره متقلبه مستقره_وثابته	انظمه فجميعها وغير عديد وابقت
تصل	تبلغ يصل وصلت تتجاوز تتعدي تتعدي ترتفع تقترب تتخطي تجاوزت	يصل وصلت ستصل تتخفض تتجرف تهبط تتجاوز ترتفع	لتصل تتراوح يصل لتصل وصلت تتجاوز فتصل تتراوح تبلغ وتصل تتخفض تتعدي	وتصل يصل لتصل ستصل وصلت تبلغ تتجاوز وتصل لا تصل	وتصل يصل لتصل فتصل ستصل تبلغ تتجاوز تتراوح تصالي تتزايد_تدرجيا	تصل دقتها الليزر اشعه الليله جناح الاقامه زائد مليار ناقص
نمو	النمو نموها للمو انكماش سينمو نمو انتعاش انخفاض ينمو معدلات	النمو نموها تكاثر نموه تطور ونمو ازدهار انتعاش نضج انكماش	النمو ونمو نموها بنمو نموه تطور لنمو نمو للمو والنمو	ونمو النمو لنمو معدل نمو نمو_وتطور بنمو نموها تطور_ونمو انتعاش نموه	ونمو لنمو النمو بنمو للمو نموها معدل نمو نمو_وتطور نموه نمو	نمو وتيره النواتج حاسم تباطا انكماش النمو واعداد المحلي الاقتصاد
معدلات	المعدلات معدلاة معدل انخفاض ارتفاع النمو الانخفاض مستويات التضخم تنخفض	معدل ومعدلات المعدلات بمعدلات وتيره مستويات اسعارها معدلاتها لمعدلات تركيزات	معدل ومعدلات المعدلات معدلاته لمعدلات بمعدلات معدلاتها ومعدل وبمعدلات المراضه	معدل ومعدلات نسبه لمعدلات معدلات بمعدلات المعدلات انخفاض_معدلات معدلاته فمعدلات	ومعدلات بمعدلات المعدلات معدل لمعدلات انخفاض_معدلات ارتفاع_معدلات وانخفاض_معدلات معدلات_مرتفعه لمعدل	رقما معدلات سناثر خافت لتزايد اثارا سنتسبب يطرق سيرتفع حاجزا
تغير	يتغير تبدل تغيرها تتغير تغيرت تغيير فتغير التغير تغيرا غيرت	تغيير وتغير يتغير تبدل تتغير تغيرت غيرت تعديل يغير تحسن	تغيير وتغير يتغير بتغير تتغير تغيرت تغيرات بتغيير التغير تغيرا	تغيير وتغير تبدل يتغير غيرت تغيرت تتغير لتغير يغير بتغير	وتغير يتغير تغيير بتغير لتغير تتغير تغيرت بغير غيرت التغير	تغير تراجمعت يحدث ويثبات باقي وعلاقه انخفضت اذبعت سيعزز واستقرت

