

**République Algérienne Démocratique et Populaire**  
**Ministère de l'Enseignement Supérieur et de la Recherche Scientifique**  
**UNIVERSITE SAAD DAHLEB BLIDA 1**



**Faculté des Sciences**  
**Département d'Informatique**

Mémoire Présenté  
Pour l'obtention de diplôme de

**Master en Informatique**

**Spécialité : Ingénierie du logiciel**

**Titre :**

**Elaboration d'un corpus de test pour un système  
d'évaluation automatique des réponses courtes**

**Réalisé par :**

- ❖ Oukina Faiza Radia.
- ❖ Amar setti Imene.

**Proposé et encadré par :**

- ❖ Mme Leila Ouahrani

**Composition de jury :**

Président : Mr Cherif Zahar.  
Examineur : Mr Ferfera.

**Soutenu le : 30/09/2019**

# Résumé

Le « E-Learning » est un mode d'apprentissage basé sur l'accès à des formations en ligne. Son objectif principal est d'améliorer l'autonomie des apprenants. L'acquisition du contenu pédagogique se concrétise par des tests sur les connaissances de l'apprenant. Nous parlons donc de l'évaluation automatique. Les Système d'évaluation automatique des réponses courte (ASAGS) évaluent la réponse de l'élève en la comparant aux réponses modèles de l'enseignant. Plus la corrélation entre les réponses de l'élève et de l'enseignant est élevée, plus le système de notation est efficace. Nous nous intéressons dans ce travail aux questions ouvertes à réponses courtes (quelques mots à quelques phrases construites en langage naturel).

Notre travail englobe le développement d'un corpus de test qui va être utilisé comme ressource d'évaluation des systèmes d'évaluation automatiques. Nous développons entre autres un système qui intègre :

- Un outil de manipulation de dataset (le corpus de test) qui est basé sur un enseignement pédagogique sur un contenu de programme diffusé en langue arabe suivi par un test utilisant des questions aux réponses courtes. Pour une variété d'utilisation, le data set est élaboré dans les cinq versions : Texte, Word, XML, Moodle XML et base de données.
- Un outil de création automatique de corpus de domaine, à partir du web, par interrogation automatique le moteur de recherche et nous avons intégré avec cet outil l'outil de création d'espace sémantique.
- Un outil d'évaluation pour évaluer le dataset (le corpus de test) par rapport au corpus de textes qui ont été construit de même domaine pour étudier l'impact du domaine commun sur les résultats obtenus.

**Mots clés :** Corpus de test(Dataset), Corpus de textes, la langue arabe, Evaluation automatique, Stem, traitement du langage naturel, espace sémantique, similarité sémantique.

# Abstract

"E-Learning" is a learning mode based on access to online training. Its main objective is to improve the autonomy of learners. The acquisition of the educational content is concretized by tests on the knowledge of the learner. We are talking about automatic evaluation. Automatic short answer grading systems (ASAGS) evaluate the student's response by comparing it to the teacher's model responses. The higher the correlation between student and teacher responses, the more effective the rating system is. In this work, we are interested in open questions with short answers (a few words to some sentences built in natural language).

Our work includes the development of a test corpus that will be used as an evaluation resource for automatic evaluation systems. We develop among others a system that integrates:

- A dataset manipulation tool (the test corpus) which is based on pedagogical instruction on Arabic-language program content followed by a test using short-answer questions. For a variety of uses, the data set is developed in five versions: Text, word, XML, Moodle XML and database.
- A tool for automatic creation of corpus domain, from the web, by automatic querying the search engine and we integrated with this tool the semantic space creation tool.
- An evaluation tool to evaluate the dataset (the test corpus) against the body of texts that have been constructed in the same domain to study the impact of the common domain on the results obtained.

**Key words:** Test corpus (Dataset), Text corpus, Arabic language, Automatic evaluation, Stem, Natural language processing, Semantic space, Semantic similarity.

# المخلص

" التعليم الإلكتروني" هو وضع التعلم القائم على الوصول إلى التدريب عبر الإنترنت. هدفها الرئيسي هو تحسين استقلالية المتعلمين. يتم اكتساب المحتوى التعليمي عن طريق اختبارات معرفة المتعلم. نحن نتحدث عن النظام التقييم التلقائي للإجابة القصيرة (ASAGS).

تقوم أنظمة التقييم التلقائي بتقييم اجابة الطالب من خلال مقارنتها بإجابات نموذج المعلم. كلما زاد الارتباط بين اجابات الطلاب والمعلمين، زاد فعالية نظام التصنيف. في هذا العمل، نحن مهتمون بالأسئلة المفتوحة ذات الإجابات القصيرة (بضع كلمات لبعض الجمل المبنية بلغة طبيعية).

يشمل عملنا تطوير مجموعة اختبار يتم استخدامها كمورد تقييم لأنظمة التقييم التلقائي. نحن نطور من بين أمور أخرى نظامًا يدمج:

- أداة معالجة مجموعة البيانات (قواعد البيانات) والتي تعتمد على التعليم البيداغوجي حول محتوى برنامج اللغة العربية متبوعًا باختبار يستخدم أسئلة الإجابة القصيرة. لمجموعة متنوعة من الاستخدامات، تم تطوير مجموعة البيانات في خمسة إصدارات: النص، الورد، XML، Moodle XML وقاعدة البيانات.
- أداة لإنشاء مجال مجموعة النصوص تلقائيًا، من الويب، عن طريق الاستعلام التلقائي لمحرك البحث و قمنا بدمج هذه الأداة مع أداة إنشاء الفضاء الدلالي.
- أداة تقييم لتقييم مجموعة البيانات (قواعد البيانات) مقابل مجموعة النصوص التي تم إنشاؤها في نفس المجال لدراسة تأثير المجال المشترك على النتائج التي تم الحصول عليها.

**الكلمات المفتاحية:** اختبار الإحصار (قواعد البيانات)، النص الأساسي، اللغة العربية، التقييم التلقائي، الجذعية، معالجة اللغة الطبيعية، الفضاء الدلالي، التشابه الدلالي.

## ***Remerciements***

*Notre cursus universitaire tend à sa fin, en cette heureuse occasion  
Nous remercions d'abord Allah le tout puissant de nous avoir donné La force, la santé et la  
volonté pour arriver à accomplir ce travail.*

*Nous tenons à exprimer notre profonde gratitude et nos remerciements et considération à  
notre promotrice Mme L. OUAHRANI qui n'a tari d'aucun effort pour nous conseiller ; nous  
orienter et nous montrer la voie de la réussite.*

*Et nous remercions M. Cherif Zahar d'avoir réalisé la 2ème annotation manuelle de  
l'ensemble  
du dataset élaboré.*

*Notre gratitude va également à l'Université de Bouira qui nous a fourni l'accès à un serveur  
distant pour accomplir notre travail.*

*Et puis nos parents qui ont été la cause et le support de notre vie, Que Dieu les protège.  
Notre parfaite estime, notre considération et notre profonde reconnaissance s'adressent à  
tous les enseignants du USDB, qui étaient très passions, sage et coopératives avec nous,  
Durant toutes ces cinq années d'études pour leur précieux conseil.*

*Aussi, les membres du jury pour nous avoir honorés en acceptant de juger notre travail.  
Enfin toute personne qui a participé de près ou de loin à l'achèvement de ce travail, tous nos  
amies, nos collègues, à toute la promo2019.*

*Merci de tout cœur.*

# Liste des tables

Tableau 1 : Exemple sur les questions 1.6 et 2.11 du Dataset CCASAG.....	40
Tableau 2: Statistiques sur notre Dataset CCASAG.....	41
Tableau 3 : Valeurs de corrélation de Pearson.....	74
Tableau 4 : Résultats du modèle SV et CM avec l'espace sémantique Cyber.....	77
Tableau 5 : Résultats du modèle SV et CM avec l'espace sémantique Khaleej.....	77
Tableau 6 : Résultats du modèle SV et CM avec l'espace sémantique CNN.....	77
Tableau 7 : Résultats du modèle SV et CM avec l'espace sémantique BBC+CNN.....	78
Tableau 8 : Résultats de la combinaison des modèles CM et SV.....	78
Tableau 9 : Résultats des similarités syntaxiques.....	79
Tableau 10 : Résultats de la combinaison de la similarité syntaxique.....	80
Tableau 11 : Résultats de la combinaison du modèle CM avec la mesure syntaxique (Dice).....	81
Tableau 12 : Résultats de la combinaison des modèles SV et Dice syntaxique.....	81
Tableau 13 : Résultats des WE "Aavec CBOW Kmeans".....	83
Tableau 14 : Résultats des WE "Aavec SkipGram Kmeans".....	83
Tableau 15 : Résultats des WE "Zahran CBOW Kmeans".....	84
Tableau 16 : Résultats des WE "Zahran SkipGram Kmeans".....	84
Tableau 17 : Résultats des combinaisons WE.....	84
Tableau 18 : Récapitulatif des résultats.....	85
Tableau 19 : Analyse approfondie entre la note automatique et la note manuel pour le DS Cyber.....	86
Tableau 20 : Analyse approfondie entre les deux scores manuels pour le DS Cyber.....	86
Tableau 21 : Des informations sur le corpus Cyber.....	88

# Liste des figures

Figure 1: Les ères et les tendances du classement automatique des réponses courtes.....	13
Figure 2: Les approches d'évaluation automatique des réponses courtes.....	15
Figure 3 : Mesures de similarité syntaxique.....	17
Figure 4 : Mesures de similarité sémantique.....	20
Figure 5 : Les dix langues les plus utilisées sur Internet.....	32
Figure 6 : Architecture de fonctionnement du système.....	38
Figure 7: Diagramme de classe de la base de données (Dataset CCASAG).....	41
Figure 8 : L'interface de création de corpus.....	43
Figure 9 : Les étapes de création de corpus.....	44
Figure 10 : Exemple de la page web.....	45
Figure 11 : Diagramme de séquence sur la création d'un corpus avec les 3 modes.....	48
Figure 12 : Schéma des étapes principales.....	49
Figure 13 : Phases de création de l'espace sémantique.....	50
Figure 14 : Exemple de heavy Stemming.....	53
Figure 15 : Exemple de light Stemming.....	54
Figure 16 : Exemple illustrant le traitement d'un corpus.....	56
Figure 17 : Modèles du calcul de similarité sémantique entre deux réponses courtes.....	61
Figure 18 : Vue globale sur le fonctionnement du modèle SV.....	61
Figure 19 : La phase de combinaison des termes clés.....	66
Figure 20 : La phase de collecte des URLs.....	67
Figure 21 : Outils de normalisation.....	68
Figure 22 : Outils de stemming.....	68
Figure 23: Outil NLP.....	69
Figure 24 : Outil de création de l'espace sémantique.....	70
Figure 25 : L'ajout des questions avec la réponse modèle sur le dataset.....	71
Figure 26 : L'ajout de réponses étudiantes avec leurs scores sur le dataset.....	72
Figure 27 : L'enregistrement de différentes versions des fichiers.....	72
Figure 28 : Outil d'évaluation automatique des réponses courtes.....	73
Figure 29 : Les notes attribuées par l'enseignant, SV, CM : pour la réponse de 10 élèves à la question " عرف مصطلح الجريمة الإلكترونية ".....	82
Figure 30 : Les notes attribuées par l'enseignant, DICE, COSINE : pour la réponse de 10 élèves à la question " عرف مصطلح الجريمة الإلكترونية ".....	82
Figure 31 : Références du serveur à distance.....	87

# Liste des abréviations

<b>ASAGS</b>	<b>Système d'évaluation automatique des réponses courtes (Automatic Short Answer Grading System)</b>
<b>CCASAGD</b>	<b>Cyber Criminality Arabic Short Answer Grading DataSet</b>
<b>DS</b>	<b>Dataset</b>
<b>CUDSG</b>	<b>Cairo University Dataset Gomaa</b>
<b>RM</b>	<b>Réponse Modèle</b>
<b>RE</b>	<b>Réponse Etudiant</b>
<b>TALN</b>	<b>Traitement automatique du langage naturel</b>
<b>CM</b>	<b>Modèle Calcul Matriciel</b>
<b>SV</b>	<b>Modèle Somme-Vecteurs</b>
<b>WE</b>	<b>Word-Embeddings</b>

# Table des matières

<b>Chapitre 1 : Contexte et problématique .....</b>	<b>9</b>
1. Introduction générale .....	9
2. Problématique .....	10
3. Objectifs du travail .....	10
4. Importance de notre travail .....	10
5. Structure du mémoire .....	11
<b>Chapitre 2 : Etat de l'art.....</b>	<b>12</b>
<b>Introduction.....</b>	<b>12</b>
1. Les systèmes (ASAG) .....	12
2. Les approches d'évaluation automatique des réponses courte.....	15
3. Les mesures s de similarité.....	17
4. Corpus de texte.....	21
5. Outils et ressources.....	23
6. Le Traitement automatique du langage naturel (TALN) et la langue arabe.....	30
7. Travaux connexes à notre travail.....	34
<b>Conclusion.....</b>	<b>36</b>
<b>Chapitre 3 : Système d'évaluation automatique des réponses courtes et l'élaboration d'un corpus de domaine.....</b>	<b>37</b>
<b>Introduction.....</b>	<b>37</b>
<b>Méthodologie .....</b>	<b>38</b>
1. Construction du DataSet.....	38
2. Approche de création de corpus.....	42
3. Approche d'évaluation automatique déployée pour l'évaluation du Dataset (CCASAG).....	48
<b>Conclusion.....</b>	<b>64</b>
<b>Chapitre 4 : Résultats expérimentaux et évaluation.....</b>	<b>65</b>
1. Démarche expérimentale.....	65
2. Le module d'évaluation .....	75
3. Résultats et discussion.....	78
4. Récapitulation des résultats et discussion.....	86
5. Ressources matérielles et logicielles utilisées lors du développement.....	88
<b>Conclusion.....</b>	<b>89</b>
<b>Conclusion et perspectives.....</b>	<b>91</b>
<b>Bibliographie.....</b>	<b>92</b>
<b>Annexe.....</b>	<b>97</b>

# Chapitre 1 : Contexte et problématique

## 1. Introduction générale

L'utilisation d'un ordinateur à des fins d'évaluation en matière d'enseignement a considérablement augmenté au cours des dernières années. Les objectifs de l'utilisation de l'évaluation automatique dans l'enseignement comprennent la réalisation et la consolidation des avantages d'un système présentant les caractéristiques suivantes : *Premièrement*, réduire la charge de travail des enseignants en automatisant une partie de la tâche d'évaluation, *Deuxièmement*, fournir aux apprenants des informations détaillées sur leur période d'apprentissage de manière plus efficace que l'évaluation traditionnelle, et *Enfin*, intégrer la culture d'évaluation au travail quotidien des apprenants dans un environnement d'e-Learning<sup>1</sup>. Nous nous intéressons dans ce travail aux questions ouvertes<sup>2</sup> à réponses courtes (quelques mots à quelques phrases construites en **langage naturel**) visant l'avantage de demander à l'apprenant de construire une réponse plutôt que de choisir parmi un certain nombre d'options prédéterminées (rappel et reproduction des connaissances acquises). L'**automatisation** de l'évaluation de ces réponses n'est pas simple en raison de variations linguistiques (une réponse donnée pourrait être articulée de différentes façons), nature subjective de l'évaluation (multiples réponses possibles), manque de cohérence dans la notation humaine, ... etc.

Le travail est orienté vers la **langue arabe** qui, bien que largement utilisée aujourd'hui, n'a pas encore bénéficié de recherche et de résultats matures dans le domaine de l'évaluation automatique. Plus le Traitement Automatique des Langues Naturelles (TALN) qui est un domaine à la frontière de la linguistique et l'informatique et de l'intelligence artificielle ; elle concerne la conception et le développement de programmes et techniques informatiques capables de traiter de façon automatique des données exprimées dans des divers aspects de la langue, tels que le niveau morphologique<sup>3</sup>, le niveau sémantique<sup>4</sup>, le niveau linguistique<sup>5</sup>, ou le niveau subjective<sup>6</sup>.

---

<sup>1</sup> Le e-learning -ou apprentissage en ligne- est une méthode d'apprentissage qui repose sur la mise en disposition des ressources pédagogiques à travers un support électronique.

<sup>2</sup> Une question ouverte est une question pour laquelle il n'y a pas de réponses préétablies proposées au répondant, celui-ci est donc entièrement libre dans sa réponse. [48]

<sup>3</sup> La morphologie est l'étude de la formation des mots et de leurs variations. Autrement, c'est le regroupement de différents mots à travers leurs parties, comme les suffixes, préfixes, radicaux.

<sup>4</sup> La sémantique lexicale est l'étude du sens des mots -ou plutôt des morphèmes- d'une langue donnée.

<sup>5</sup> Une réponse donnée pourrait être articulée de différentes façons qui ont tous le même sens.

<sup>6</sup> Une question peut avoir de multiples réponses possibles.

## 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. La plupart des recherches dans l'évaluation automatique des réponses courtes traitent de *l'anglais* pour lequel plusieurs approches existent et sont évaluées et validées sur des corpus de test (benchmarks ou data sets) sur des métriques de performance. Quand on doit travailler sur l'évaluation automatique en langue arabe, une limite importante est constatée par le manque considérable de ressources linguistiques (corpus arabes, lexiques et dictionnaires, outils de traitement TALN...) liée au processus de validation et d'évaluation des approches proposées dans la langue arabe et c'est pour cela qu'on construit des corpus de données et de test. Un seul dataset, celui du « Cairo University Dataset » (CUDS) [49] existe dans la littérature. Souvent, les auteurs valident leurs approches sur quelques paires de (réponse étudiant, réponse modèle) ce qui ne donne pas une bonne appréciation de l'apport apporté par de nouvelles approches par rapport à ce qui existe déjà.

## 3. Objectifs du travail

- ✓ Elaborer un dataset (corpus de test) en langue arabe basé sur un enseignement pédagogique sur un contenu de programme diffusé en langue arabe suivi par un test utilisant des questions aux réponses courtes. Pour une variété d'utilisation, le data set doit être élaboré dans les cinq versions : Texte, Word, XML, Moodle XML et base de données.
- ✓ Valider le dataset par rapport à un corpus de textes de même domaine pour étudier l'impact du domaine commun sur les résultats obtenus. La création d'un corpus de textes de domaine doit être facilitée par l'automatisation du processus d'élaboration.
- ✓ Plusieurs approches ont déjà été proposées dans le cadre de ce projet et doivent être validées et évaluées en utilisant le corpus de test élaboré. [61] [62] [65].

## 4. Importance de notre travail

Les objectifs de l'utilisation de l'évaluation automatique dans l'enseignement comprennent la réalisation et la consolidation des avantages d'un système présentant les caractéristiques suivantes :

- Réduire la charge de travail des enseignants en automatisant une partie de la tâche d'évaluation des apprenants. En effet, la correction manuelle des copies est une des
-

tâches les plus coûteuses en temps et en charge cognitive pour un enseignant et nécessite des efforts ainsi que des ressources qui pourraient être économisées.

- Fournir aux étudiants des informations détaillées sur leur période d'apprentissage de manière plus efficace que l'évaluation traditionnelle.
- Fournir aux apprenants des résultats immédiats ainsi que des commentaires détaillés à propos de leurs erreurs ou une redirection vers la partie du cours à revoir.
- Libérer l'enseignant de la correction manuelle subjective qui peut être influencée par son humeur (stress, joie...etc.)
- Intégrer la culture d'évaluation au travail quotidien des apprenants dans un environnement d'e-Learning.
- Un système d'évaluation non supervisé et totalement automatisé réduirait le risque d'erreurs humaines et les soupçons de triche ou de subjectivité.
- Permet à un chercheur avantage d'exportation, pour construire une banque de question pour l'enseignement.

## 1.7. Structure du mémoire

Notre mémoire est organisé de la manière suivante :

- **Chapitre 1** : Introduction générale qui englobe le contexte et problématique du travail.
- **Chapitre 2** : Etat de l'art traitant des outils d'évaluation automatique des réponses courtes et des mesures de similarité.
- **Chapitre 3** : La construction des ressources (un dataset et un corpus de domaine) qu'on test sur le système d'évaluation automatique des réponses courtes déjà construit et développement de l'approche méthodologique et des modèles de calcul de similarité.
- **Chapitre 4** : présente le cadre applicatif de nos travaux, la présentation de l'implémentation réalisée et discussions des résultats obtenus.
- Nous terminons par une conclusion mettant l'accent sur le travail réalisé et des perspectives projetées.

# Chapitre 2 : Etat de l'art

## Introduction :

L'évaluation joue un rôle important dans l'enseignement et à des effets importants sur l'apprentissage. Néanmoins, cette tâche présente une charge contraignante pour les enseignants en termes de temps, de concentration et de précision. Pour cela les chercheurs se sont focalisés sur le domaine de l'évaluation afin d'automatiser cette tâche et par la suite améliorer la performance de l'apprentissage des étudiants, ainsi que la réduction de la charge des enseignants et enfin intégrer la culture d'évaluation au travail quotidien des apprenants dans un environnement du e-Learning.

Nous nous intéressons aux questions ouvertes à réponses courtes (quelques mots à quelques phrases construites en langage naturel). Le concept principal de cette dernière consiste à comparer la réponse de l'apprenant(RA) à la réponse de référence de l'enseignant appelée Réponse Modèle(RM) et à mesurer la similitude (ou similarité) entre les deux réponses puis à convertir cette similarité en note appelée aussi score. La première étape qui est le calcul de la similarité constitue une difficulté essentielle dans l'automatisation du processus de correction. Chaque apprenant a son propre niveau de maîtrise linguistique et un vocabulaire différent, ce qui rend difficile la mise en œuvre d'un système capable de reconnaître toutes les réponses similaires à la réponse type et d'évaluer la consistance d'une réponse donnée.

Les corpus jouent un rôle très important dans l'évolution de la linguistique arabe. Cette ressource a beaucoup retenu l'attention en TALN; il aide les chercheurs à éviter les généralisations linguistiques basées sur sa perception cognitive intériorisée du langage.

## 1. Les systèmes (ASAGS)

### 1.1. Vue historique des ASAGS

Il existe plusieurs approches qui traitent le sujet d'évaluation automatique des réponses courtes dont la recherche s'est développée depuis 1996. [1] dans leur article ont identifié 35 systèmes dans 4 méthodes différentes (représentés dans la Figure 1) :

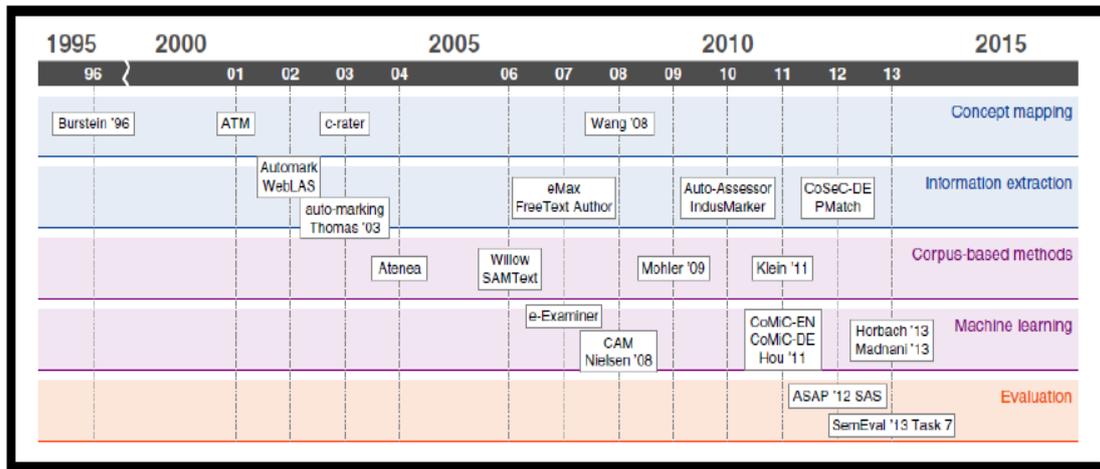


Figure 1 : Les ères et les tendances du classement automatique des réponses courtes [1]

➤ **Les méthodes basées sur le mappage de concepts (Concept Mapping) :**

L'idée de la cartographie des concepts est de considérer les réponses des élèves comme composées de plusieurs concepts, et détecter la présence ou l'absence de chaque concept lors de la notation. Des questions appropriées doivent donc faciliter cette idée, comme une question qui demande une solution à un problème plus une justification, ou une question qui demande plusieurs explications au même problème.

**ATM** : Le marqueur de texte automatique ATM (Automatic text Marker) [2] ce système décompose les réponses des enseignants et des étudiants dans des listes de concepts minimaux comprenant pas plus de quelques mots chacune, Une fois que la réponse est décomposée, le système compte le nombre de concepts communs pour noter la réponse. Chaque concept dans lequel la réponse est cassée est essentiellement l'unité la plus petite possible dans une réponse. Une pondération est attribuée à ces concepts aux fins de la notation. Les pondérations obtenues par les réponses des élèves sont résumées pour donner le score global de la réponse des élèves. [2]

**C-rater** : L'évaluateur conceptuel (Concept rater ou C-rater) [3] C-Rater est basé sur une approche de reconnaissance paraphrase. Il emploie modèles de réponses correctes comprenant des points essentiels formulé en langage naturel. C-Rater vise à scoring automatique et se concentre sur le sens, don tolérer les erreurs de forme. En particulier, les réponses données par l'enseignant sont entrées phrase par phrase pour chaque concept. Cela simplifie l'évaluation de la réponse des élèves, puisqu'un seul concept est considéré à la fois lors de la notation. Cette technique évite le recours à une solution indirecte, telle que la division de la question en

plusieurs parties [4] et il est avancé que cela peut conduire à une plus grande précision [5]. En outre, le format de saisie en langage naturel est avantageux par rapport aux autres systèmes nécessitant une expertise et l'utilisation du langage de balisage [6] [1].

➤ **Les méthodes basées-corpus (Corpus Based Methods) :**

Le principe est d'utiliser des propriétés statistiques des corpus, qui sont des ensembles de textes. Ces méthodes peuvent être utiles lors de l'interprétation des synonymes dans les réponses courtes. Alors, afin de limiter les réponses correctes qui peuvent être identifiées, utiliser seulement le vocabulaire des réponses modèles. Ensuite, pour renforcer ce vocabulaire, prendre en considération ses synonymes et effectuer des traductions en autre langue afin d'éviter les difficultés de la langue source.

➤ **Les méthodes basées sur l'apprentissage automatique (Machine learning) :**

Les systèmes d'apprentissage par machine utilisent généralement un certain nombre de mesures extraites du langage naturel (Techniques de traitement et similarité), qui sont ensuite combinés en une seule note ou un score en utilisant Un modèle (fonction d'estimation) de classification ou de régression. Cela peut être pris en charge par une boîte à outils d'apprentissage machine telle que Weka [9].

## **1.2. Définition des ASAGS**

Les systèmes de notation automatique à réponse courte (ASAG) sont conçus pour évaluer automatiquement les réponses courtes en langage naturel comportant une longueur de quelques mots à quelques phrases. Dans le monde réel, de nombreux facteurs, tels que le niveau de difficulté et la diversité des réponses des élèves, varient considérablement d'une question à l'autre, ce qui permet à différentes techniques ASAG de devenir supérieures pour différentes mesures d'évaluation.

Les réponses textuelles peuvent être des réponses courtes ou des essais, et chacun doit être évalué différemment. Normalement, les systèmes de notation automatique des réponses courtes évaluent et marquent la réponse d'un apprenant en le comparant à une ou plusieurs réponses correctes. [28]

Les ASAGS se basent sur le contenu plutôt que sur le style. Une mauvaise qualité d'écriture (de formulation) est, jusqu'à un certain point, tolérée facilitant ainsi la production d'une idée, une méthode, ou un produit original qui est la meilleure façon de vérifier l'acquisition des informations et des connaissances selon la taxonomie de Bloom [10] il existe plusieurs taxonomies représentant ces niveaux d'acquisition de connaissances. Dans ce domaine, la taxonomie la plus fréquemment citée est celle de Bloom.

## 2. Les approches d'évaluation automatique des réponses courtes :

On peut classer l'évaluation automatique des réponses courtes en trois types d'approches différentes et importantes qu'on peut les représenter comme suit « voir la figure 2 » :

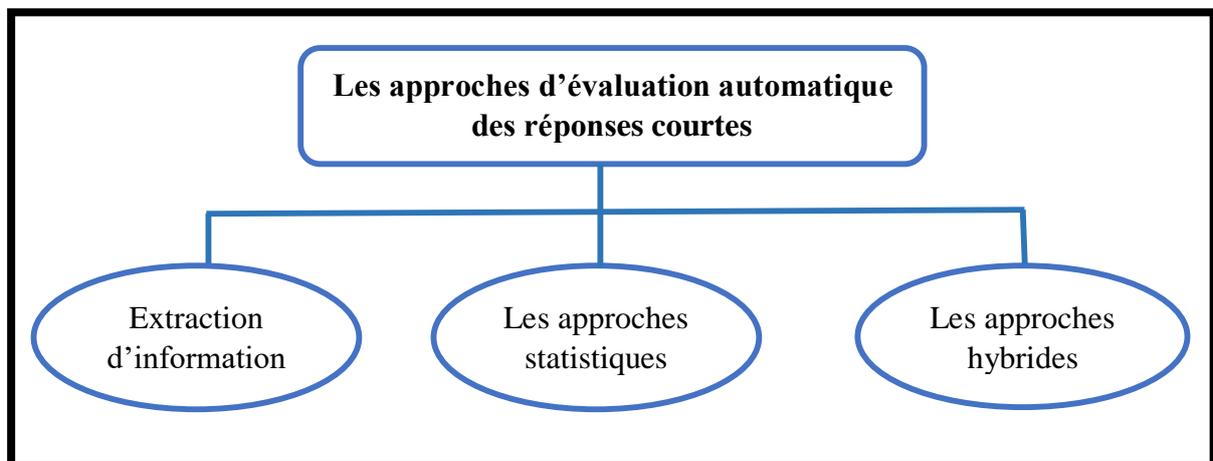


Figure 2 : Les approches d'évaluation automatique des réponses courtes

### 2.1. Extraction d'information (Information Extraction) :

Les techniques d'extraction d'informations (IE) extraient des informations pertinentes à partir de réponses textuelles analysées syntaxiquement en appliquant un ensemble de modèles. Les motifs sont définis soit sur du texte en surface (mots, phrases), soit sur des éléments structurels tels que des balises de parties de discours (PoS). Dans le cas de réponses courtes en texte libre, elles sont généralement créées par des experts en la matière afin d'indiquer des concepts importants qui devraient figurer dans les réponses.

**Marquage automatique :** Le marquage automatique [7] utilise des modèles fabriqués à la main qui sont adaptés à un ensemble de formation pour la construction de modèles réduits.

Deux motifs sont formés pour chaque question, chaque question valant deux points. L'évaluation empirique montre que l'approche est plus efficace que la ligne de base du voisin le plus proche  $k$  avec des fonctions de sac de mots pondérées par  $tf.idf$ . [8]. Cependant, la quantité de données n'est pas rapportée, ce qui rend difficile la comparaison de cette approche avec l'approche artisanale.

## **2.2. Les approches statistiques**

Basée sur des techniques de l'apprentissage automatique, et une classification au niveau des textes entiers. Le fonctionnement de cette approche est basé sur l'extraction des phrases à l'aide de la méthode symbolique après l'entraînement des modèles des corpus sur les extraits des textes, et enfin la classification des textes.

Les méthodes statistiques présentent certains avantages par rapport aux méthodes linguistiques. Elles ne nécessitent pas de connaissances spécifiques de la langue de corpus, et n'utilisent pas des ressources linguistiques externes (dictionnaire, ...). Elles ont l'avantage d'être rapides, simples à mettre en œuvre, et s'appuient sur des calculs des fréquences. Cette approche est performante sur des corpus de taille suffisamment grande. En revanche, l'utilisation des approches linguistiques nécessite la maîtrise des langues des corpus étudiés, ainsi qu'une connaissance parfaite des règles syntaxiques de la langue de corpus. Les approches linguistiques et statistiques présentent certains inconvénients lorsqu'ils sont utilisés seules. D'une part, l'approche statistique est incapable de faire face à des basses fréquences des termes composés. D'autre part, l'approche linguistique dépend de la langue traitée et n'est pas assez flexible pour faire face à des structures complexes de termes composés.

Pour surmonter les faiblesses des deux approches, une approche hybride est proposée en combinant les calculs statistiques et les filtres linguistiques [11]. L'ordre dans lequel cette hybridation est effectuée varie d'un système à un autre. En effet, dans certains systèmes les résultats obtenus par l'analyse statistique sont validés par une analyse linguistique, tandis que dans d'autres systèmes les résultats de l'analyse linguistique sont validés et filtrés par une analyse statistique.

## **2.3. Les approches hybrides**

Les approches hybrides sont une combinaison entre les règles linguistiques avec les informations statistiques afin de résoudre l'ambiguïté morphologique.

Le principe de cette approche est que les termes qui partagent les mêmes contextes sont des termes sémantiquement proches [12].

L'approche hybride est considérée comme la plus efficace et cohérente en termes d'analyse, car elle combine les deux approches et tire profit de leurs avantages.

### 3. Les mesures de similarité

Les mesures de similarité de texte jouent un rôle de plus en plus important dans la recherche liée au texte et dans des applications telles que la recherche d'informations, la classification de textes, la classification de documents, la détection de sujets, le suivi de sujets, la génération de questions, la réponse à des questions, la rédaction de dissertations, la traduction automatique, la traduction de texte résumé et autres.

La similarité sémantique est introduite à travers des algorithmes basés sur le corpus et sur la connaissance. Les mesures basées sur les chaînes fonctionnent sur les séquences de chaînes et la composition des caractères. Une métrique de chaîne est une métrique qui mesure la similarité ou la dissimilarité (distance) entre deux chaînes de texte pour une correspondance ou une comparaison de chaîne approximative. La similarité basée sur le corpus est une mesure de similarité sémantique qui détermine la similarité entre les mots en fonction d'informations obtenues à partir de grands corpus. La similarité basée sur la connaissance est une mesure de similarité sémantique qui détermine le degré de similarité entre des mots à l'aide d'informations dérivées de réseaux sémantiques. [31]

#### 3.1. Similarité syntaxique :

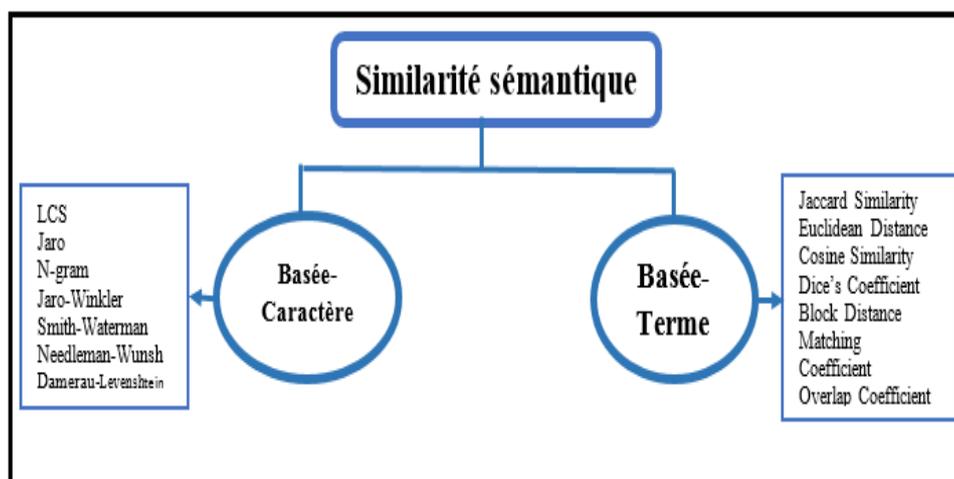


Figure 3 : Mesures de similarité syntaxique

Les mesures de similarité syntaxique peuvent être regroupées en deux catégories, la première basée sur les caractères et la seconde sur les termes, contenant sept mesures pour chacun, mais nous citons les mesures les plus utilisées :

➤ **La similarité cosinus (ou mesure cosinus) :**

Est une mesure de similarité entre deux vecteurs d'un espace de produit interne qui mesure le cosinus de l'angle entre eux, elle permet de calculer la similarité entre deux vecteurs à  $n$  dimensions en déterminant le cosinus de l'angle entre eux. Cette métrique est fréquemment utilisée en fouille de textes<sup>1</sup>.

Soit deux vecteurs  $A$  et  $B$ , l'angle  $\theta$  s'obtient par le produit scalaire et la norme des vecteurs :

$$\text{Sim}_{\text{cosinus}} = \cos \theta = \frac{\vec{A} \cdot \vec{B}}{\|\vec{A}\| \|\vec{B}\|}$$

Comme la valeur  $\cos \theta$  est comprise dans l'intervalle  $[-1,1]$ , la valeur  $-1$  indiquera des vecteurs résolument opposés,  $0$  des vecteurs indépendants (orthogonaux) et  $1$  des vecteurs similaires (colinéaires de coefficient positif). Les valeurs intermédiaires permettent d'évaluer le degré de similarité.

➤ **La distance euclidienne :**

La distance euclidienne calcule la similarité entre deux documents  $d1$  et  $d2$  comme la distance entre leurs représentations vectorielles ramenées à un seul point [32]. C'est la racine carrée de la somme des différences au carré entre les éléments correspondants des deux vecteurs.

$$\text{Sim}_{\text{euclidienne}} = \sqrt{\sum_{i=1}^n (d1_i - d2_i)^2}$$

Où  $n$  est le nombre total de termes représentés, i.e. la taille des vecteurs.

➤ **Le coefficient de Dice :**

Ce coefficient est utilisé en statistique pour déterminer la similarité entre deux échantillons il définit comme étant le double du nombre de termes communs dans les chaînes comparées divisé par le nombre total de termes dans les deux chaînes [32].

---

<sup>1</sup> La **fouille de textes** ou « l'extraction de connaissances » dans les textes est une spécialisation de la fouille de données et fait partie du domaine de l'intelligence artificielle. Cette technique est souvent désignée sous l'anglicisme text mining.

$$\text{Sim}_{\text{dice}} = \frac{2Nc}{N1+N2}$$

➤ **La similarité Jaccard :**

L'indice de Jaccard (ou coefficient de Jaccard) est le rapport entre le cardinal (la taille) de l'intersection des ensembles considérés et le cardinal de l'union des ensembles [33]. C'est un indice statistique utilisé pour comparer la similarité et la diversité des ensembles échantillons. Soit deux ensembles  $d_1$  et  $d_2$ , l'indice est :

$$\text{Sim}_{\text{jaccard}}(d_1, d_2) = \frac{||d_1 \cap d_2||}{||d_1 \cup d_2||}$$

La similarité obtenue  $\text{Sim}_{\text{jaccard}}(d_1, d_2) \in [0, 1]$

➤ **Coefficient de corrélation de Pearson :**

Le coefficient de corrélation de Pearson calcule la similarité entre deux documents  $d_1$  et  $d_2$  comme le cosinus de l'angle entre leurs représentations vectorielles centrées-réduites. La similarité obtenue  $\text{Sim}_{\text{pears}}(d_1; d_2) \in [-1; 1]$  [32].

$$\text{sim}_{\text{pearson}}(d_1, d_2) = \text{sim}_{\text{cosinus}}(d_1 - \overline{d_1}, d_2 - \overline{d_2})$$

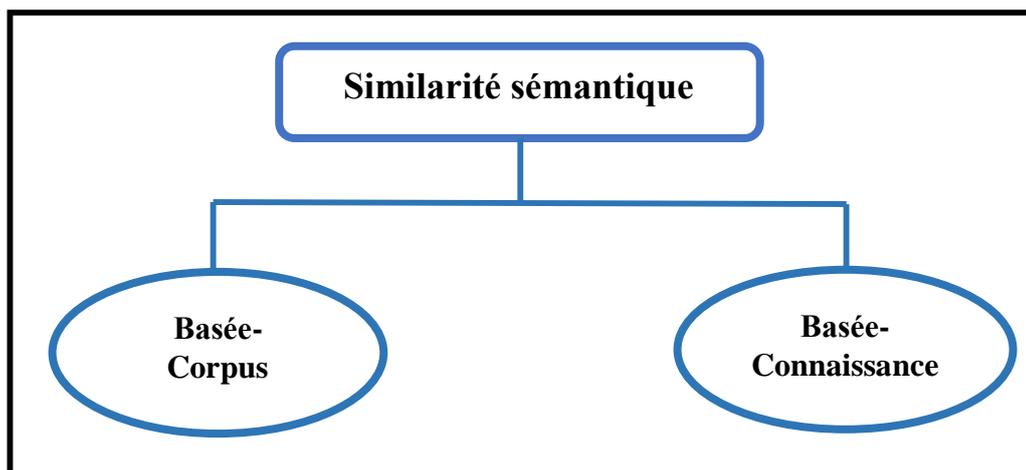
D'où  $\overline{d_1}$  (resp.  $\overline{d_2}$ ) représente la moyenne de  $d_1$  (resp.  $d_2$ ).

### 3.2. Similarité sémantique :

Une mesure de 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. Deux concepts sont considérés comme sémantiquement similaires s'il y a une synonymie, hyponymie, antonymie, ou troponymie<sup>1</sup> entre eux. Deux sens de mots sont considérés comme sémantiquement liés s'il existe au moins une relation lexico sémantique entre eux - classique ou non classique. [32]

---

<sup>1</sup> Relation sémantique entre deux verbes, l'un décrivant de manière plus précise l'action de l'autre. Le premier verbe est dit troponyme du second.



**Figure 4 : Mesures de similarité sémantique**

### ✚ Similarité fondée sur le corpus

La similarité basée sur le corpus est une mesure de similarité sémantique qui détermine la similarité entre les mots en fonction d'informations obtenues à partir de grands corpus. En linguistique, un corpus (corpus pluriel) ou un corpus de texte est un ensemble de textes volumineux et structuré (de nos jours généralement stockés et traités électroniquement). Ils sont utilisés pour effectuer des analyses statistiques et des tests d'hypothèses, vérifier des occurrences ou valider des règles linguistiques sur un univers spécifique.

Toutefois, la similarité basée sur le corpus est une mesure de similarité sémantique 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 [35]. Les mots sont représentés par des vecteurs sémantique d'où la notion d'espace vectoriel VSM (Vector Space Model).

Le modèle VSM vise à convertir l'ensemble des textes du corpus du format textuel en format numérique. Ceci se fait en construisant une matrice des fréquences des termes exhaustifs du corpus (vocabulaire du BOW).

Dans l'approche sémantique, chaque ligne de la matrice représente le vecteur de contexte du terme par rapport à son apparition autour du reste des termes. Cependant, pour calculer la similarité entre deux mots (terme), il faut récupérer leurs vecteurs de contexte et puis appliquer une des mesures syntaxiques citées précédemment sur ces deux vecteurs.

### **La similarité basée sur la connaissance**

La similarité basée sur la connaissance est l'une des mesures de similarité sémantique qui repose sur l'identification du degré de similarité entre les mots à l'aide d'informations dérivées de réseaux sémantiques [1]. WordNet [36] est le réseau sémantique le plus populaire dans le domaine de la mesure de la similarité entre les mots basée sur la connaissance ; WordNet est une grande base de données lexicale en anglais. Les noms, verbes, adjectifs et adverbes sont regroupés en ensembles de synonymes cognitifs (synsets), chacun exprimant un concept distinct. Les synsets sont liés entre eux au moyen de relations conceptuelles, sémantiques et lexicales.

## **4. Corpus de texte :**

### **4.1. Définition de corpus :**

Un corpus est une collection de textes lisibles par machine qui ont été produits dans un cadre de communication naturel. C'est un ensemble de données linguistiques, soit des textes écrits, soit une transcription de paroles enregistrées, qui peuvent être utilisées comme point de départ de la description linguistique ou comme moyen de vérification. Ils représentent une ressource très cruciale pour la construction et l'évaluation des parseurs statistiques. Les annotations des corpus riches ont été aussi utilisées dans une variété d'applications telles que la segmentation, la discrétisation, étiquetage morphosyntaxique, désambiguïsation morphologique, segmentation des syntagmes et étiquetage des rôles thématiques (sémantiques). Sous la contrainte temps, la création des corpus s'affronte au compromis entre la richesse linguistique et la taille du corpus. Ceci est particulièrement le cas pour les langues morpho-syntaxiquement complexes telles que l'arabe.

## 4.2. Variété des contenus d'un corpus

**Texte :** Un corpus écrit est un corpus de textes, s'il contient essentiellement des textes. Il peut être constitué de documents différents (tableau, extraits de textes, etc.). Il peut être présenté comme un livre, une page Web, ou un document (en WORD, ODS, LATEX, INTERLEAF, etc.).

**Parole :** Un corpus de parole est un corpus constitué d'enregistrements de données orales.

**Visuel :** Un corpus visuel est un corpus contenant des vidéos ou images, et ce type de corpus peut contenir les textes pour compléter, décrire, ou présenter les objets.

## 4.3. Variété des formats d'un corpus

À cause de la variété des usages, il existe plusieurs formats de corpus. Le format le plus utilisé est le texte brut (*.txt*). Pour un corpus monolingue, les textes sont sauvegardés dans un ou plusieurs fichiers *.txt*

L'autre type de format le plus utilisé est « le format XML », ou plus exactement des instances de XML, comme TMX<sup>1</sup>, TEI<sup>2</sup>, DOCBOOK<sup>3</sup>, NITF<sup>4</sup>, etc., correspondant à des DTD ou à des schémas XML définis par les professionnels du traitement des documents, de la traduction, et de la localisation.

## 4.4. Construction de corpus

L'utilisation de corpus n'est bien entendue pas récente dans presque toutes les applications du traitement automatique du langage, ainsi que l'utilisation du Web comme base pour la constitution de ressources textuelles est très récente.

Ces dernières années sont le témoin sur les travaux tentant d'exploiter ce type de données. Dans une perspective de traduction automatique dans [81] étudie la possibilité d'utiliser les sites Internet proposant les informations en plusieurs langues pour constituer des corpus parallèles bilingues.

---

<sup>1</sup> [https://en.wikipedia.org/wiki/Translation\\_Memory\\_eXchange](https://en.wikipedia.org/wiki/Translation_Memory_eXchange)

<sup>2</sup> <http://www.tei-c.org/index.xml>

<sup>3</sup> <http://www.docbook.org>

<sup>4</sup> <http://www.gwg.nga.mil/ntb/index.html>

Dans une toute autre approche [74] mettent au point un logiciel pour la constitution d'un corpus de phrases dans le but d'étudier le comportement des noms prédictifs marquant la localisation et le déplacement, afin de mesurer si l'introduction des prépositions dans les requêtes en recherche d'information permet d'améliorer la précision.

Dhaou Ghoul dans [86] a présenté une méthode de construction d'un corpus pour l'Arabe à partir du Web. Qui permet d'identifier les différentes formes du discours de la langue arabe et spécifiquement les mots outils ou tokens<sup>1</sup>. Pour ce faire, tout d'abord il a implémenté un script Perl qui permet d'extraire à partir d'une adresse URL des fichiers html. Ensuite, Il a nettoyé ces fichiers pour obtenir des textes bruts. La détection et l'identification des tokens se fait grâce à un outil appelé *kawâkib*.

## 5. Outils et ressources

Les outils d'évaluation de traitement automatique de la langue arabe sont l'ensemble des développements visant à modéliser et reproduire, à l'aide de machines, la capacité humaine à produire et à comprendre des énoncés linguistiques dans des buts de communication.

### 5.1. Outils du traitement automatiques de la langue arabe

Il y a divers outils qui sont utilisés pour le traitement automatique de la langue. Ceux-ci incluent des outils d'analyse morphologique et des ressources lexicales :

#### 5.1.1. Analyseurs morphologiques en arabe

L'analyse morphologique en arabe s'intéresse, comme les autres langues, aux formats du mot. Mais étant donnée la richesse du mot graphique, l'opération englobe rapidement des aspects formels de la langue, reléguant les traits sémantiques en arrière-plan.

Dans cette partie nous essayerons de présenter les difficultés posées lors d'une analyse morphologique ou morphosyntaxique en détaillant les composants du mot graphique en arabe.

---

<sup>1</sup> Les mots qui n'appartiennent pas au lexique arabe et n'obéissent pas à la dérivation morphologique de l'arabe.

### ➤ **Difficulté de l'analyse morphologique de l'arabe**

En analyse morphologique, le principal problème à résoudre est l'ambiguïté. Il existe différents types d'ambiguïtés.

#### ➤ **Ambiguïtés d'agglutination**

Contrairement aux langues latines, en arabe, les articles, les prépositions, les pronoms, ect. Collent aux adjectifs, nom, verbes, et particules auxquels ils se rapportent. Comparé au français un mot arabe peut parfois correspondre à une phrase française [13] (exemple : le mot arabe « انتذكروننا » correspond en français à la phrase « Est-ce que vous vous souvenez de nous »). Cette caractéristique engendre une ambiguïté morphologique au cours de l'analyse. En effet, il n'est pas toujours facile de distinguer un proclitique ou enclitique d'un caractère original du mot. Par exemple, le caractère « و » dans le mot « وصل » (il est arrivé) est un caractère original alors que dans le mot « وفتح » (et il a ouvert), il s'agit plutôt d'une proclitique.

#### ➤ **Ambiguïté dérivationnelle et flexionnelle**

La flexion est la variation de la forme des mots en fonction de facteurs grammaticaux telle que la conjugaison pour les verbes (exemple : le mot يتأثرون (ils s'influencent) est le résultat de la concaténation du préfixe « ي » indiquant le présent et du suffixe « ون » indiquant le masculin pluriel du verbe « تأثر »

Le problème en analyse morphologique de l'arabe se rapporte surtout au niveau de la dérivation qui est un phénomène plus complexe que la flexion. En effet, la dérivation est la formation de nouveaux mots à partir de mots existants. Dans le cas de la langue arabe, la plupart des mots sont dérivés à partir de racines trilitères ou quadrilitères. Le mot arabe n'est pas le résultat d'une simple concaténation de morphèmes comme c'est le cas en anglais (exemple : unfailingly = un+fail+ing+ly), mais c'est à partir d'une racine, d'une combinaison de voyelles, de préfixes, d'infixes, de suffixes et d'un schème morphologique qu'on obtient un mot (exemple : à partir de la racine « أثر » on peut dériver plusieurs verbes tel que « تأثر » (s'influencer) et plusieurs noms tel que « متأثر » « ému ») [14].

### **5.1.2. Ressources lexicales**

Une caractéristique des approches à base de connaissances est l'utilisation des ressources lexicales. Un premier type de ressource classique qui peut être exploité sont les inventaires de sens, c'est-à-dire une ressource qui, à chaque mot, lie une liste de sens possibles comme, par exemple, un dictionnaire. D'autre part, des ressources telles que les thésaurus peuvent être utiles pour établir des liens entre les sens des différents mots. D'autre part, il existe des ressources telles que WordNet [15] qui sont des ressources lexicales structurées et qui, non seulement, joue le rôle d'inventaire de sens mais donne, également, accès à une hiérarchie de sens.

### **5.2. Outils d'extraction d'information**

Le processus de la recherche des réponses exige des analyses approfondies des passages qui peuvent comporter la réponse exacte. En effet, la reconnaissance des entités nommées peut être considérée comme un outil ou technique d'extraction d'informations.

### **5.3. Outils de recherche d'information**

Lors de la phase de recherche de documents, il est possible d'utiliser de nombreux moteurs de recherche comme des outils de recherche d'informations. Par exemple on peut utiliser le moteur de recherche Google, comme étant une source de données linguistiques, pour extraire les passages pertinents qui sont susceptibles de contenir la réponse précise à une question donnée ou à la consolidation de corpus qui permettrait de le faire.

### **5.4. Ressources sémantiques**

Consiste l'utilisation d'une ressource sémantique. La sémantique est l'étude du sens des expressions linguistiques. La quantité de recherche scientifique dans les modèles computationnels de la sémantique est beaucoup plus petite que d'autres domaines du TAL. Cela est peut-être dû à sa plus grande complexité et subtilité. Exp (WordNet arabe)

### 5.4.1. WordNet arabe

Qu'est-ce WordNet ? Un dictionnaire<sup>1</sup> ? Un thésaurus<sup>2</sup> ? Les dictionnaires contiennent généralement des connaissances sur des lexies<sup>3</sup> alors que les encyclopédies<sup>4</sup> contiennent des connaissances éparées, du monde, sur la surface de la terre. Quant aux thésaurus, leur structure est bâtie autour des concepts et aident l'utilisateur à acquérir l'unité lexicale la plus appropriée lorsqu'il a un concept à rechercher. WordNet n'est ni un dictionnaire classique ni un thésaurus : il est en fait, un arrangement des traits de chacune de ces deux ressources lexicales.

Le WordNet est une base de données lexicale lisible par machine qui regroupe les mots en ensemble de synonymes appelés synset. Chaque synset<sup>5</sup> peut être considéré comme représentant d'un sens de mot unique (sens ou concept). Le WordNet fournit généralement des définitions et des exemples généraux pour le synset et inclut les relations sémantiques entre elles.

Les relations sémantiques, qui incluent entre autres l'hyponymie<sup>6</sup> et l'hyperonymie<sup>7</sup>, permettent à un WordNet d'être interprété hiérarchiquement comme une ontologie/taxonomie lexicale. Plusieurs projets de WordNet ont été coordonnés pour inclure des liens croisés. Cela leur permet d'être utilisés non seulement comme thésaurus monolingues informatiques sophistiqués, mais aussi comme dictionnaires. La liste ci-dessous offre une sélection de définitions de la sémantique, avec des divergences importantes [16] :

- La sémantique est l'étude du sens des mots. [17]
- La sémantique est l'étude du sens. [18]
- La sémantique est l'étude du sens dans le langage. [19]
- La sémantique est l'étude du sens des mots, des phrases et des énoncés. [20]

---

<sup>1</sup> **Dictionnaire** Recueil des mots d'une langue, des termes d'une science, d'un art, rangés par ordre alphabétique, avec leur signification. [www.mediadico.com]

<sup>2</sup> **Thésaurus** une liste de termes sur un domaine de connaissances, reliés entre eux par des relations synonymiques, hiérarchiques et associatives. Le thésaurus constitue un vocabulaire normalisé. [www.fr.wikipedia.org]

<sup>3</sup> **Lexie** C'est une suite de caractères formant une unité sémantique, un mot, et pouvant constituer une entrée de dictionnaire. [www.fr.wikipedia.org]

<sup>4</sup> **Une encyclopédie** peut prendre la forme d'un livre ou plusieurs livres. Elle se présente souvent comme une collection d'articles traitant chacun un thème. [www.fr.wikipedia.org]

<sup>5</sup> **Synset** est un ensemble de mots qui ont la même signification (Synset = ensemble de synonymes), alors un synset représente un sens donné d'un mot. Les synsets sont reliés entre eux par différentes relations, soit lexicales ou taxonomiques.

<sup>6</sup> Y est un hyponyme de X si chaque Y est une sorte de X, c'est-à-dire le sens de Y est inclus dans le sens de X, exemple : le chien est un hyponyme de l'animal

<sup>7</sup> Y est un hyperonyme de X si chaque X est une sorte de Y, c'est-à-dire le sens de X est inclus dans le sens de Y, exemple : animal est un hyperonyme de chien

- La sémantique est l'étude du sens qui est véhiculé à travers la langue. [21]
- La sémantique linguistique est l'étude de la façon dont les langues organisent et expriment des significations. [22]
- La sémantique est la partie de la linguistique qui se préoccupe du sens. [23]

L'AWN<sup>1</sup> a été utilisé comme la référence lexicale pour évaluer les systèmes de désambiguïsation de sens de mot arabe (WSD : Word Sens Disambiguation). Dans WSD, les mots sont étiquetés avec leur signification particulière dans le contexte en utilisant des définitions de sens dans une ressource lexicale prédéfinie.

La base de données relationnelle établie pour leur AWN contient quatre entités principales :

- **Item** : Contient les entités conceptuelles, synsets, classe d'ontologie et des instances. Chaque entité a un identifiant unique et des informations descriptives comme le gloss.
- **Word** : Contient le sens du mot, où la forme du mot est relié à élément via son identifiant.
- **Form** : contient des informations lexicales (ex. variation flexionnelle, la racine et/ou la forme plurielle pour l'arabe).
- **Link (entre items)** : contient aussi le type de relation (ex. équivalence, subsumption...etc.)

Le projet Arabic WordNet (AWN) [24] a commencé en 2006 grâce à la collaboration de plusieurs universités et entreprises. AWN est basé sur la conception et le contenu de PWN [25]. Les synsets arabes sont jumelés à des synsets dans le WordNet de Princeton et sont capables au synset dans l'ontologie de haut niveau SUMO (Suggested Upper Merged Ontology). [26]

#### 5.4.2. Enjeux et limitations pour la construction de Wordnet

Les travaux mentionnés auparavant, qui ont tiré avantage de cette ressource sémantique (WordNet) soulèvent tout un ensemble de problèmes. En ce qui concerne les wordnets

---

<sup>1</sup> Disponible gratuitement dans le lien suivant : <http://www.globalwordnet.org/AWN/>

construits à partir de dictionnaires conventionnels, des problèmes spécifiques se posent : la circularité des définitions, les incohérences (par exemple lorsqu'une équivalence traductionnelle n'est proposée que dans un sens, et pas dans l'autre - ou lorsque les descriptions sémantiques sont de granularité plus fine en langue source qu'en langue cible, ou réciproquement), les problèmes de couverture (absence d'une entrée, absence de caractérisation d'une certaine acception), les problèmes de spécialisation (certains dictionnaires ne couvrent qu'un domaine de spécialité), etc.

Par ailleurs, la structure même des WN pose problème. La relation d'hyponymie, quand elle touche l'ensemble du lexique d'une langue, se heurte à certaines limites. Par exemple, dans l'organisation des verbes dans PWN, les liens hiérarchiques sont moins élaborés, et l'on passe rapidement d'un sens spécialisé à des sens très généraux. En outre, il n'y a aucune catégorisation hiérarchique actuellement définie pour les adjectifs et les adverbes présents dans réseau.

En ce qui concerne les noms, enfin, on trouve des hiérarchies déséquilibrées, puisque certains noms sont liés à une grande chaîne de sens finement gradués, tandis que, d'autres sont très proches des concepts les plus généraux.

Ensuite, de nombreuses définitions de sens sont manquantes, et certaines relations font défaut. Et pour cela nous avons construit notre propre ressource un corpus de texte qui contient 1273 fichiers (.txt) liée au domaine de la cybercriminalité et de plus un corpus de test de même domaine.

## **5.5. Techniques du traitement linguistique**

Les différentes techniques trouvées dans la littérature :

1. Suppression des nombres et chiffres du texte à traiter.
2. Suppression de la ponctuation.
3. Uniformiser la casse (rendre tout le texte en majuscule ou bien en minuscule).
4. La tokenisation signifie la segmentation du texte selon un caractère identifié (généralement les espaces simples sont les plus fréquemment utilisés).
5. Part-Of-Speech tagging qui est une méthode qui prend en considération le coté grammatical et le contexte (Nom, verbe, adjectif...).

6. Le stemming qui consiste à convertir le mot à sa racine ou bien l'élimination des suffixes, préfixes...
7. La suppression des mots vides ou stop-words (et, ou, avec, pour...).
8. Correction d'orthographe afin d'éviter les ambiguïtés et les erreurs de frappes.

## 5.6. Ressources NLP nécessaires

Il existe deux sources (ensembles) de données :

### 5.6.1. Corpus d'apprentissage :

Il sert à retirer un modèle ou un classement à partir d'un nombre suffisant d'information. Concernant la langue arabe, une variété de corpus est disponible en ligne. [27] a détaillé dans son article tous les corpus arabe existants pour des finalités d'avancement dans le domaine du traitement du langage naturel. Ce corpus se constitue d'un ensemble d'échantillons qui sont fournis pour l'apprentissage. L'approche d'apprentissage se base sur les mathématiques, en appliquant les modèles existants à la langue, cette méthode permet d'avoir des résultats rapides sans avoir besoin de modéliser les phénomènes linguistique. il est utilisé par des informaticiens sans formation en linguistique. On fournit un grand corpus avec un ensemble des textes et des analyses correctes, Qui sera balisé par un étiqueteur grammatical, par la suite l'algorithme calcule les fréquences dans ce corpus.

### 5.6.2. Dataset (jeux de donnée de test)

Ou encore le corpus de test sert à vérifier la qualité de l'apprentissage à partir du corpus d'apprentissage. On doit évaluer sa performance, précision et cohérence. C'est pour cette raison que les datasets sont conçus. La notation des réponses courtes peut également être formulée comme une tâche de similarité dans laquelle le score est attribué en fonction de la similarité des réponses des enseignants et des élèves.

Cependant, assez peu de dataset dans des domaines spécifiques sont librement disponibles pour la langue arabe, et parmi les datasets d'apprentissage qui existe on peut citer [28] [29] [30].

- ✚ Un système Text-to-Text (Texas) a été introduit dans [28]. Ici, le score est attribué en fonction d'une mesure de la similarité sémantique entre une réponse d'élève et d'une

réponse modèle lorsque plusieurs mesures sont utilisées, notamment la connaissance. Le système a été appliqué à un ensemble de données informatiques contenant 21 réponses et 610 réponses d'élèves, où la meilleure valeur de corrélation de Pearson entre l'automatique et les manuels était 0.47. Une version améliorée du système du Texas a été introduite à [29], qui utilisait des alignements de graphes de dépendance générés par apprentissage automatique. Le jeu de données utilisé dans cette version contenait 80 questions et 2273 réponses. La meilleure valeur de corrélation de Pearson était 0.518, et la meilleure erreur quadratique moyenne était de 0.978.

✚ [30], Cette recherche présente le premier ensemble de données en arabe pouvant être utilisé comme référence pour la notation automatique des réponses courtes en arabe. Les questions présentées dans le jeu de données couvrent un chapitre du programme officiel égyptien de sciences de l'environnement (ES), qui représente 25% du programme global. Le jeu de données contient 61 questions, 10 réponses pour chaque, avec un total de 610 réponses. La longueur moyenne des réponses des étudiants est de 2,2 phrases, 20 mots ou 103 caractères. Les marques avec des valeurs comprises entre 0 et 5 et ont obtenu un coefficient de corrélation de Pearson ( $r$ ) et une erreur quadratique moyenne (RMSE) de 0.83 et 0.75, respectivement.

Et pour remplir ce manque des datasets nous avons choisi de faire notre propre dataset afin d'évaluer les méthodes de réponse courte, il contient 48 questions et 2133 réponses étudiant. Les réponses ont été notées indépendamment par deux juges humains, en utilisant une échelle entière de 0 (complètement incorrect) à 5 (réponse parfaite). Les évaluations sont effectuées à l'aide de la méthode de coefficient de corrélation de Pearson mesuré par rapport à la moyenne des notes attribuées par l'homme pour chaque question. (Le corpus est détaillé dans chapitre 3).

## **6. Le Traitement automatique du langage naturel (TALN) et la langue arabe**

Le traitement automatique des langues naturelles (TALN) est un domaine à la frontière de la linguistique et l'informatique, il a pour objectif de développer des logiciels capables de traiter de façon automatique des données linguistiques exprimées dans une langue naturelle donnée et pour une application bien définie. Cet objectif passe nécessairement par l'explicitation

des règles de la langue puis les représente dans un formalisme calculable et enfin les implémenter à l'aide des programmes informatiques. De nos jours, il existe plusieurs applications du traitement des langues telles que la reconnaissance de l'écriture manuscrite (la détection de la langue), le résumé automatique, le traitement de la parole, l'annotation sémantique, l'indexation et la recherche de documents, l'extraction d'informations, la traduction, etc.

## 6.1. La langue arabe

La langue arabe appartient à la famille des langues sémitiques. Les langues sémitiques sont caractérisées par :

1. Un lexique construit principalement à partir de racines trilitères et quadrilitères<sup>1</sup>.
2. D'un système d'écriture de droite vers la gauche.
3. D'un alphabet de type Abjed<sup>2</sup>.

L'arabe est la première langue sémitique en nombre de locuteurs, elle est parlée et écrite dans le monde arabe par plus de 360 millions. Elle est classée dans les dernières années comme huitième langue mondiale en nombre d'utilisateurs dans l'Internet Utilisée par plus de 22 pays dans le monde arabe et par plus de 360 millions de personnes. Elle est classée dans les dernières années comme quatrième langue mondiale en nombre d'utilisateurs dans l'Internet Utilisée par plus de 22 pays dans le monde [37].

Les outils du traitement automatique de la langue Arabe, se caractérisent par leur diversité au niveau des langages de développement utilisés, des entrées/sorties manipulées, des représentations internes et externes des résultats, etc.

Par ailleurs, la diversité des populations arabes et de leurs cultures ont fait émerger différentes variantes de l'arabe allant de **l'arabe classique** utilisé dans le coran et les textes religieux, à **l'arabe moderne standard (ASM)** qui est une forme plus récente de l'arabe classique ; c'est la langue officielle du monde arabe actuellement, elle est utilisée dans l'enseignement et dans les médias [38].

---

<sup>1</sup> Racines composées de trois ou quatre lettres consonnes ; par fois appelées triconsonantiques et quadriconsonantiques. Les racines biconsonantiques et pentaconsonantiques ne génèrent que la catégorie nominale (Kouloughli 1994).

<sup>2</sup> Alphabet composé que des consonnes.

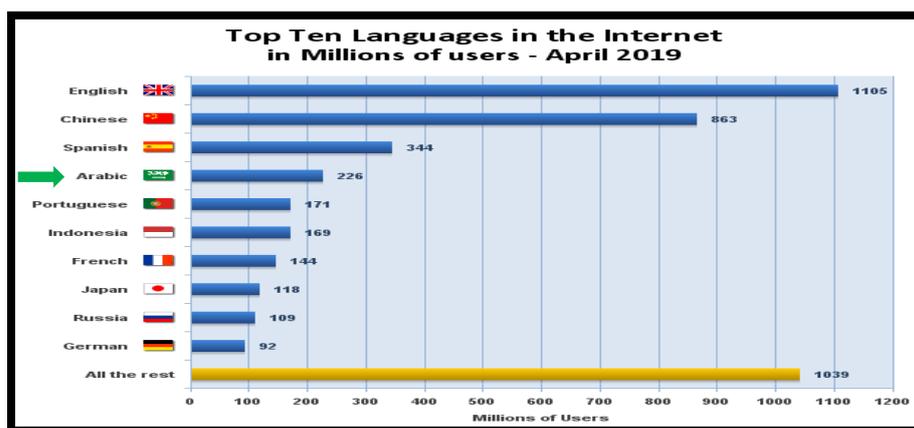


Figure 5 : Les dix langues les plus utilisées sur Internet [37]

## 6.2. Problèmes du traitement automatique de la langue arabe

Parmi les problèmes spécifiques à la langue arabe et à certaines autres langues sémitiques, nous citons l'absence des voyelles, l'absence des ponctuations régulières, les problèmes de flexion et d'agglutination.

### 6.2.1. Absence des voyelles

Le problème de la voyellation réside dans son absence dans les textes arabes. En effet, les signes de voyellation sont des signes diacritiques placés au-dessus ou au-dessous des lettres, qui apparaissent dans certains ouvrages scolaires pour débutants et dans le Coran. La non voyellation génère plusieurs cas d'ambiguïtés et des problèmes lors de l'analyse automatique [39].

### 6.2.2. Ordre des mots dans la phrase

En langue arabe, on met au début de la phrase le mot sur lequel on veut attirer l'attention ou le mot le plus riche en sens. Cet ordre provoque des ambiguïtés syntaxiques artificielles dans la mesure où il faut prévoir dans la grammaire toutes les règles de combinaisons possibles d'inversion de l'ordre des mots dans la phrase [40]. Ainsi par exemple, on peut changer l'ordre de la phrase (1) pour obtenir une autre phrase (2) ayant le même sens.

الطفل تعلم القراءة والكتابة في المدرسة / l'enfant a appris la lecture et l'écriture à l'école. (1)

في المدرسة الطفل تعلم القراءة والكتابة / à l'école l'enfant a appris la lecture et l'écriture. (2)

### 6.2.3. Segmentation de textes

Pour analyser un texte, nous devons procéder à sa segmentation en paragraphes, phrases et mots. Le problème principal réside dans le fait que cette segmentation est une source d'ambiguïté, vu que, d'une part, la ponctuation est rarement utilisée dans les textes arabes, et d'autre part, cette ponctuation, lorsqu'elle existe, n'est pas toujours déterminante pour guider la segmentation. De plus, certains mots outils peuvent marquer le début d'une nouvelle phrase, ce qui nécessite des analyses de surface afin de pouvoir segmenter le texte. Les phrases (3) et (4) illustrent ce problème :

«الطفل تعلم القراءة و الكتابة في المدرسة» / l'enfant a appris la lecture et l'écriture à l'école » (3)

Dans cette phrase, la particule 'و, w' ne joue pas le rôle de séparateur entre propositions mais une conjonction de coordination entre les mots " القراءة " et " الكتابة ", et donc ne segmente pas la phrase.

«الطفل تعلم القراءة و الكتابة في المدرسة وحفظ القرآن الكريم في المسجد» / l'enfant a appris la lecture à l'école, et il a appris le coran à la mosquée » (4)

Tandis que la particule " و, w " dans (4) joue le rôle de séparateur entre propositions et segmente la phrase en deux propositions. En général, l'espace reste le critère principal pour la segmentation de texte arabe.

### 6.2.4. Agglutination de mots

Contrairement à la plupart des langues latines, en arabe, les articles, les prépositions, les pronoms se collent aux adjectifs, noms, verbes et particules auxquels ils se rapportent. Un mot arabe peut parfois correspondre à toute une phrase. Par exemple, le mot arabe « أتستعملونها , est-ce que vous l'utilisez ? ». Cette caractéristique engendre des ambiguïtés morphologiques au cours de l'analyse. En effet, il est parfois difficile de distinguer entre un proclitique/enclitique et un caractère du mot en question. Par exemple, le caractère " و " dans le mot " وجمع " est un caractère qui fait partie de ce mot alors que dans le mot " وحصل ", il s'agit d'un proclitique.

### 6.2.5. Absence de ponctuation régulière

La langue arabe n'est pas appuyée principalement sur les signes de ponctuations et les marqueurs typographiques ; il est à noter que ces derniers ne sont pas utilisés de façon régulière

dans les textes arabes actuels, et même dans le cas où ils y figurent, ils ne sont pas gérés par des règles précises d'utilisation. Par ailleurs, nous pouvons trouver tout un paragraphe ne contenant aucun signe de ponctuation à part un point à la fin de ce paragraphe [41].

## **7. Travaux connexes à notre travail**

*Les travaux liés au même Projet entamés durant l'année 2018/2019* : 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, [61] présente une synthèse expérimentale des différentes approches de mesures de similarité syntaxique appliquées à des ensembles de données (Data Sets) exprimées dans la langue arabe.

[62] 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.

[65] 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.

De plus, il existe plusieurs systèmes de classement automatique menés en anglais, tels que E-rater [42], Automark [43], C-rater [46] et LVQ [47]. Par ailleurs, quelques études ont été réalisées en arabe [48], [49], [50].

Le travail de Gomaa et Fahmy [49], est le plus étroitement lié à notre travail car c'est la seule recherche qui a traité la langue arabe. Les auteurs introduisent un corpus de test qui traite des données de la science de l'environnement. Il contenait 61 questions, 10 réponses pour chacune, avec un nombre total de 610 réponses. Son objectif était de surmonter les difficultés rencontrées dans le traitement de texte arabe. Pour chaque essai, la corrélation de Pearson Le coefficient ( $r$ ) et l'erreur quadratique moyenne (RMSE) ont été calculé. Combinaison des mesures de différentes catégories atteint  $r = 0,83$  et  $RMSE = 0,75$ . Ces valeurs résultantes

étaient très proches des valeurs qui ont été marqués manuellement par deux annotateurs. En ce qui concerne la langue anglaise, beaucoup les systèmes de notation ont été appliqués en tant que tâche de similarité de texte à texte dans lequel le score est attribué en fonction d'une mesure de la similitude lexicale et sémantique entre une réponse d'élève et une réponse type lorsque plusieurs mesures sont utilisées, y compris basé sur les chaînes, basé sur la connaissance et basé sur corpus [51,28,29].

Le système de Power-grading [52] était basé sur la formation d'une similitude métrique entre les réponses des élèves, puis en utilisant cette métrique pour regrouper les réponses en grappes et sous-grappes. Un excellent et un aperçu plus détaillé des travaux connexes peuvent être trouvés dans [53,54], tels que CarmelTC [55], C-Rater [46], Intelligent Assesment Technologies (IAT) [57] et Oxford-UCLES [58].

[50] ont proposé une méthode hybride basée sur un LSA modifié et une fonctionnalité syntaxique dans un essai de notation automatique de rédaction en arabe. Ils ont utilisé l'ensemble de données proposé dans [49]. Leur proposition s'est concentrée sur le marquage de partie de discours (POS) de manière à identifier la caractéristique syntaxique des mots dans la matrice de similarité. Leur étude visait à résoudre les inconvénients du LSA standard qui mettait l'accent sur l'analyse syntaxique limitée. Cependant, l'utilisation de la seule technique LSA dans leur étude ne garantissait pas un rapport de corrélation élevé.

[59] ont proposé un environnement d'examen en ligne en arabe facilitant l'interaction entre les étudiants et les enseignants. En fait, leur proposition n'est développée que pour la notation de questions objectives (non rédactionnelles). Pour la notation de l'essai, il n'ya rien de plus qu'un système de stockage dans lequel l'instructeur doit évaluer manuellement les écrits de l'élève.

[60] présentent un système automatique hybride combinant LSA et trois caractéristiques linguistiques : 1) l'origine des mots, 2) la fréquence des mots et 3) le nombre de fautes d'orthographe. Leur proposition détermine la dimensionnalité réduite optimale utilisée dans LSA pour évaluer la performance de leur système proposé. [48] présentent une méthode dépendant du synonyme et de l'origine des mots. Ils attribuent une pondération aux mots de réponse de l'instructeur afin d'améliorer le processus d'évaluation. Leur étude était irréalisable et n'avait ni ensemble de données ni résultat expérimental.

## **Conclusion**

La correction automatique des réponses courtes comme nous venons de voir est un domaine dont les racines sont ancrées et emmêlées avec pleins d'autres domaines tels que le traitement automatique de la langue, le calcul de similarité et l'évaluation automatique. Cet aspect multidisciplinaire offre une multitude d'axes de travail pour traiter le sujet. Dans le prochain chapitre nous exposerons la méthodologie que nous avons suivie dans notre travail et les techniques auxquelles nous avons eu recours.

# **Chapitre 3 : Système d'évaluation automatique des réponses courtes et l'élaboration d'un corpus de domaine et du Dataset**

## **Introduction**

De nos jours, les systèmes d'évaluation sont un sujet scientifique très important au sein de la communauté éducative. Les énormes quantités de tests et d'élèves ont entraîné la nécessité de systèmes de notation automatiques. Par conséquent, les chercheurs accordent plus d'attention à ces systèmes, qui sont devenus des éléments essentiels de la communauté éducative, car ils sont capables de réduire la charge de travail du personnel enseignant grâce aux examens des étudiants classés [49] - [50]. En fait, la plupart du temps des enseignants est perdu à cause de la notation manuelle. Les enseignants du monde entier souffrent du temps perdu à noter les notes des étudiants.

Le corpus est une ressource qui pourrait être très importante et utile pour faire avancer les diverses applications linguistiques. Cette ressource a beaucoup retenu l'attention en NLP ; mais il est difficile de trouver un corpus conçu pour le traitement du langage naturel et, plus spécifiquement, pour la langue arabe.

Travailler avec la construction de corpus devient une alternative intéressante aux différentes applications du traitement du langage naturel, telles que la réponse à une question, la traduction automatique, la recherche d'informations, etc. De même, avec les données hétérogènes et les demandes des utilisateurs pour des informations précises la nécessité du Web de mettre en évidence la construction du corpus. De plus, l'arabe n'a pas un nombre équivalent de corpus linguistiques par rapport à d'autres langues comme l'anglais. Dans ce chapitre, nous nous concentrons sur la construction de notre corpus de textes arabes.

## Méthodologie :

On commence par cette architecture qui résume les principales étapes de notre système que nous allons les traiter « voir la figure 6 » :

- Nous avons commencé par la construction du dataset.
- Après la construction du corpus de texte de même domaine avec le dataset.
- Ensuite nous avons fait une évaluation par rapport aux approches déjà développées.

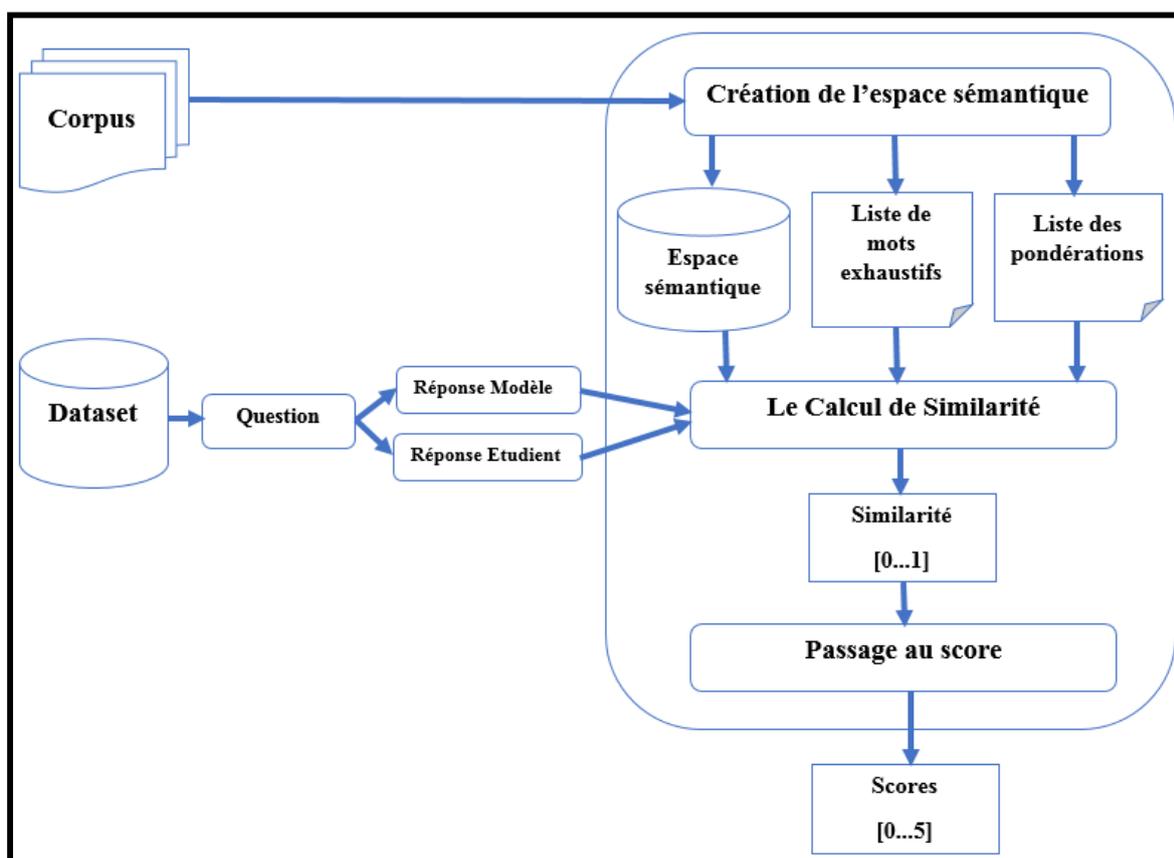


Figure 6 : Architecture de fonctionnement du système

### 1. Construction du DataSet :

Le processus de notation automatique (AS) aborde l'évaluation de la réponse de l'étudiant (RE) en effectuant une comparaison avec la réponse modèle. Plus la corrélation entre les réponses de l'élève et celle de l'enseignant est élevée, plus le système de notation est efficace.

Dans notre travail, la méthodologie générale sera utilisée pour faire avancer un système de notation des réponses courtes en arabe basé sur des méthodes de correspondance de similarité de texte. L'objectif ici est de recueillir des couples de (réponse modèle, réponse étudiant) avec

le score **manuel** attribué par l'annotateur humain. L'évaluation des approches consiste à comparer ce score au score attribuée automatiquement par le programme de notation.

Le Dataset « Cyber Criminality Arabic Short Answer Grading » (CCASAG Dataset) est construit sur la base des données d'examen recueillies lors de l'enseignement du module « Cybercriminalité » dispensé aux étudiants de Master 1 informatique (SSI, IL et SIR) du département d'Informatique de l'université de Blida pour l'année 2018/2019. Le cours est hébergé sur une plateforme de téléenseignement<sup>1</sup>. Le contenu du cours est disponible ICI<sup>2</sup>

Pour avoir un Dataset représentatif de la réalité il fallait passer par l'enseignement d'un module proprement dit en langue arabe. Environ 170 étudiants ont suivi le cours durant le premier semestre qui a été validé par un examen officiel écrit d'une durée de 1h30'.

- **Les questions :** L'ensemble de données contient 48 questions à réponses courtes proposée dans 5 types de questions :
  - « عرف » : Définir ?
  - « اشرح » : Expliquer ?
  - « ما النتائج المترتبة على » : Quelles sont les conséquences de ?
  - « علل » : Justifier ?
  - « ما الفرق » : Quelle est la différence ?

Pour chaque question, une réponse modèle est proposée par l'enseignant responsable.

- **Annotation des réponses des étudiants :**

Deux annotateurs humains ont corrigé les réponses des étudiants. Pour chaque réponse, deux scores humains sont disponibles pour refléter la nature subjective du processus d'évaluation déjà au niveau des enseignants. Chaque réponse est notée sur 5 points. Chaque annotateur n'était pas au courant de la note accordée par le deuxième annotateur.

Un total de **2133 réponses** étudiants et donc de couples de (réponse étudiant, réponse modèle enseignant) est enregistré.

Le tableau 1 représente des exemples de questions, de réponses types et de réponses courtes fournies par deux étudiants et les notes attribuées par deux correcteurs.

---

<sup>1</sup> <http://elearninginfo.univ-bouira.dz/test/moodle/course/index.php>

<sup>2</sup> <http://elearninginfo.univ-bouira.dz/test/moodle/course/view.php?id=4>

**Tableau 1 : Exemple sur les questions 1.6 et 2.11 du Dataset CCASAG**

السؤال 1.6	عرف مصطلح الجريمة المنظمة على الانترنت	
الإجابة النموذجية	هي عنف منظم تقوم به جماعات ترتكب أفعالا تخترق بها القانون للحصول على مكاسب مالية، بطرق وأساليب غير مشروعة تنفذ بعد تدبير وتنظيم	
<b>Note1</b>	<b>Note2</b>	<b>REPONSES ETUDIANTS</b>
5	5	هي عنف منظم تقوم به جماعات من اجل كسب الأموال وتعتمد على التنظيم و كسب الأموال.
4.5	4	هي سلوك غير قانوني تقوم على التنظيم بهدف رقة المعلومات او تغييرها.
السؤال 2.11	ما هي وظيفة راصد لوحة المفاتيح ؟	
الإجابة النموذجية	تسجيل جميع ضغطات المفاتيح وبالتالي كل كلمات المرور المدخلة على الكمبيوتر وأرقام البطاقات الائتمانية لغرض السرقة كما يستعمل أيضا لإثبات الهوية كقياس حيوي سلوكي.	
<b>Note1</b>	<b>Note2</b>	<b>REPONSES ETUDIANTS</b>
2.5	3.5	برنامج يثبت على اجهزة الحاسب يمكن من ارسال رمز المفتاح عند الضغط عليه للمخترقين فيسهل عليهم الحصول على كلمات السر و كذا رموز البطاقات الائتمانية و المصرفية.
2.5	2.5	يجمع و يعالج ضربات المفاتيح ليتمكن مرتكب الجريمة من سرقة الارقام السرية و التجسس عليه.

Le travail manuel de collecte des réponses est très fastidieux surtout qu'il fallait retaper toute les réponses des étudiants et gérer les styles d'écriture différents. En effet, faute d'absence de moyens (plateforme) pour réaliser le test directement en ligne, il nous a fallu faire l'examen en manuel en salle d'examen, ensuite retranscrire toute les réponses des étudiants sur machine. Notre Dataset (CCASAG) est construit dans les conditions naturelles.

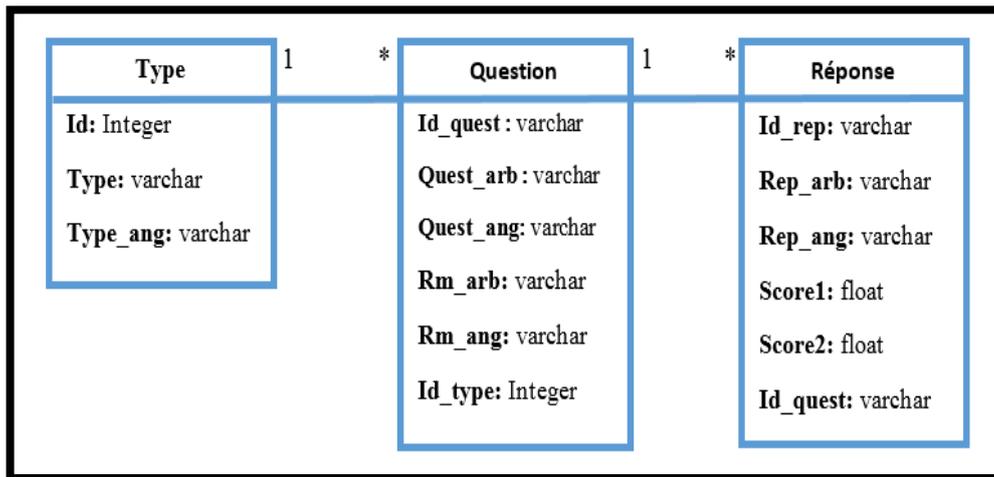
De plus la corrélation de Pearson(r) entre les notes accordées par les 2 annotateurs est calculée. Elle est de l'ordre de 83%. Nous avons considéré le score moyen des deux annotateurs comme un score manuel pour examiner la tâche de notation automatique. Un outil est développé pour manipuler le dataset de CCASAG et en faire des exportations dans différents formats. Le dataset est disponible<sup>1</sup> dans cinq formats : base de données, texte, XML, MOODLE XML et Word.

- **Structure du corpus de test (Dataset)**

Dans l'objectif de permettre la manipulation du dataset de CCASAG dans plusieurs versions et de pouvoir les extraire il est nécessaire de présenter son modèle dans le diagramme de classe « voir la figure 7 ».

<sup>1</sup> <https://drive.google.com/open?id=13DoT0HqDwrDnwcwTFucJsceM9rfK197C>

Le modèle accorde la possibilité à tout moment d'enrichir le corpus de test en ajoutant de nouvelles questions ou de nouvelles réponses étudiants :



**Figure 7 : Diagramme de classe de la base de données (Dataset CCASAG)**

Un examen plus approfondi des scores attribués par les deux annotateurs indique la sous-jacente subjectivité dans la notation des assignations à réponse courte déjà pour les annotations humaines. Par exemple dans le cas où les deux annotateurs étaient en désaccord (le **2eme** annotateur accordait la note la plus élevée) représente 38.2% des réponses ce qui signifie la notation du 2 ème annotateur été large par rapport au premier annotateur qui a été sévère. Le tableau suivant « **tableau 2** » présente les principales statistiques sur notre Dataset .

**Tableau 2 : Statistiques sur notre Dataset CCASAG**

Nombre de couples : 2133 Nombre de questions : 48 Corrélation de Pearson : 83 %		
Les cas possible	Nombre de réponses	Pourcentage de réponses%
Dans le cas où les 2 scores sont égaux	743	34.83%
Dans le cas où le score est différent de plus d'un point sur l'échelle à cinq points	235	11.01%
Dans le cas où le score est différent d'au moins 4 points	6	0.28%
Dans le cas où les annotateurs étaient en désaccord (le <b>2eme</b> annotateur accordait la note la plus élevée)	814	38.209%

Le score automatique est ensuite calculé en deux étapes. Tout d'abord, la similitude entre les réponses des étudiants et des modèles. Deuxièmement, les valeurs de similarité obtenues [0-1] sont mises à l'échelle sur l'échelle d'origine [0-5] pour faciliter la comparaison. L'étape de mise à l'échelle est essentielle et est étroitement liée à l'évaluation de la précision du système. La précision du système est évaluée en comparant les scores manuels et automatiques tout en tenant compte de deux facteurs, qui sont l'association et la taille de l'erreur. L'association est mesurée par le Pearson coefficient de corrélation ( $r$ ) pour indiquer l'exactitude du classement des notes. La taille de l'erreur est mesurée par la racine Erreur quadratique moyenne (RMSE) pour caractériser la précision de la prédiction automatique du score.

**Et parmi les plus importants avantages du dataset (CCASAG) :** utilisé dans la notion de paraphrase<sup>1</sup> particulièrement pour les réponses qui ont 5/5 (Dans le cas de deux phrases, le système doit déterminer si les deux phrases ont la même signification). Nous avons fait une extraction des réponses dont le score est compris entre **4 et 5**. Le résultat est présenté dans la même Url<sup>2</sup> du Dataset(CCASAG).

## **2. Approche de création du corpus de texte :**

Notre démarche de création du corpus de texte se base sur la collecte des URLs relatives à des termes clés de recherche et d'en extraire les informations qui vont constituer le corpus.

Plus la possibilité d'enrichir un corpus déjà créé et intégrer les résultats.

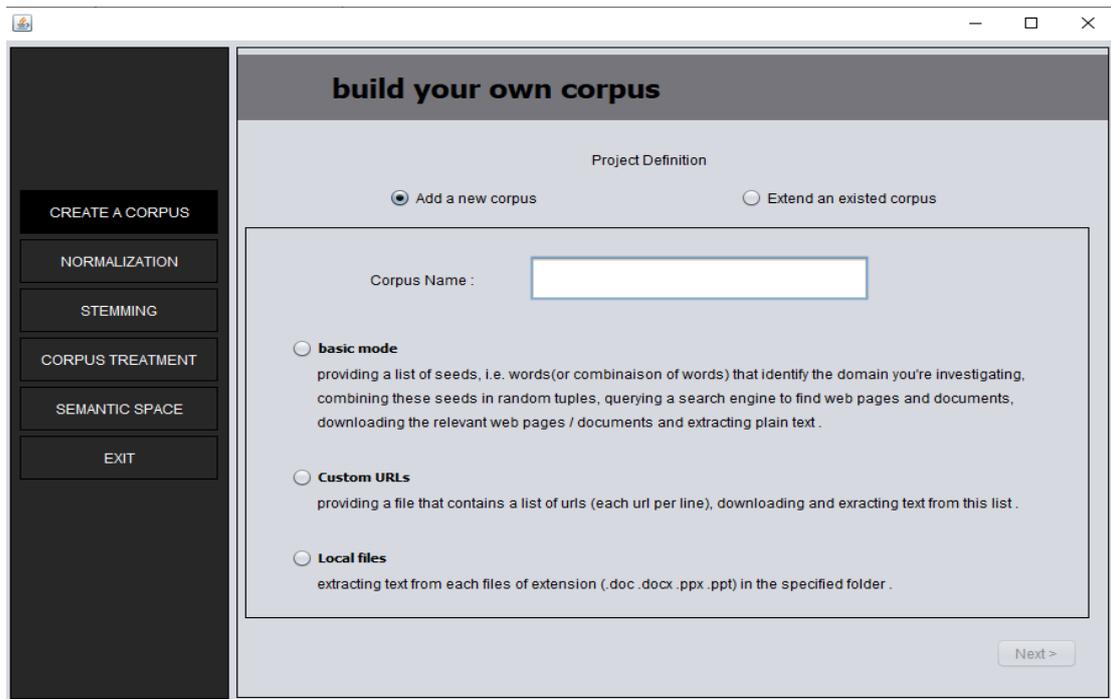
Nous avons construit 3 modes de création d'un corpus de texte possibles comme illustré dans **la figure 8 :**

- Basic mode (Mode de base).
- Custom URLs (URL personnalisées).
- Local files (Fichiers locaux).

---

<sup>1</sup> 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). Également, nous dirons que pour deux phrases qu'ont le même sens  $p_1$  et  $p_2$ , l'une est la paraphrase de l'autre et vice versa, puisque la relation de paraphrase est symétrique.

<sup>2</sup> <https://drive.google.com/open?id=13DoT0HqDwrDnwcwTFucJsceM9rfK197C>



**Figure 8 : L'interface de création de corpus**

### **2.1. Basic mode (Mode de base) :**

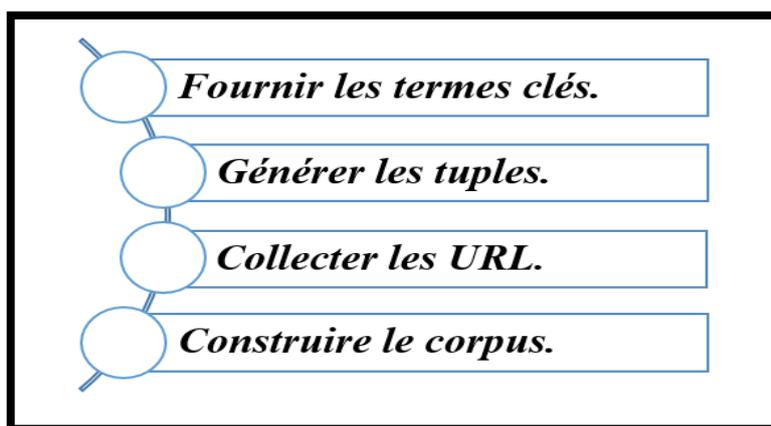
Lors de la première étape de la création de notre corpus de texte, l'utilisateur fournit une liste de termes comportant au moins 5 mots à utiliser comme éléments de base pour la collecte de texte. Ces termes donnent selon l'utilisateur la spécificité du domaine du corpus destination.

Celles-ci sont ensuite combinées en «tuples» de différentes longueurs et envoyées sous forme de requêtes à un moteur de recherche, qui renvoie une liste d'URL potentiellement pertinentes. Les pages Web réelles sont ensuite récupérées, converties en texte brut et enregistrées au format "txt".

La procédure itérative d'extraction de terme / téléchargement de corpus est répétée autant de fois que nécessaire (par exemple, jusqu'à ce que le corpus atteigne une certaine taille).

On doit contrôler plusieurs paramètres importants, tels que le nombre de requêtes émises pour chaque itération, le nombre de mot clés utilisées dans une même requête, le nombre de pages à récupérer, etc.

Comme illustré dans **la figure 9** la démarche englobe 4 étapes :



**Figure 9 : Les étapes de création du corpus de texte**

### **2.1.1. Fournir nos termes choisis :**

Il s'agit de l'étape la plus importante du processus de création de corpus, nous fournissons ici les éléments de départ c'est-à-dire des mots (ou une combinaison de mots) qui identifient le domaine que nous étudions (Cybercriminalité) et qui seront utilisés pour générer les requêtes qui seront soumises au moteur de recherche.

L'utilisateur peut taper (ou copier / coller) la liste de termes qu'il a choisie dans la zone de texte. Le nombre minimum de termes que nous devons fournir est de 5.

### **2.1.2. Génération de tuples**

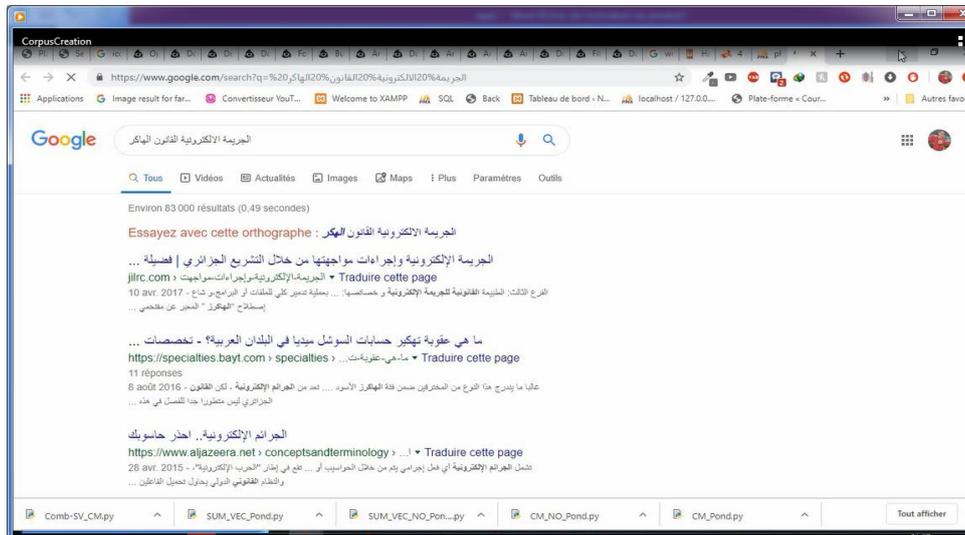
Les termes fournis à l'étape précédente seront regroupés de manière aléatoire pour former des n-tuples (une variété de combinaisons de nos termes). Ces tuples seront soumis en tant que requêtes au moteur de recherche.

Nous pouvons choisir le nombre de tuples à générer. Bien sûr, le nombre de combinaisons aléatoires possibles est fini et dépend du nombre de termes que nous avons fournis. Comme nous pouvons également désélectionner des n-tuples individuels si nous pensons qu'ils ne donneront pas de résultats intéressants.

### **2.1.3. Collecte d'URL**

Après la génération des tuples nous passons à la collecte des URLs. Il faut d'abord préciser le nombre d'url à extraire à partir chaque requête et puis commencer la génération des urls, ce processus est réalisé en récupérant le code html de chaque requête dans le navigateur web et se focaliser seulement sur les balises <a> dont le paramètre 'href' afin de récupérer les n premiers urls, une liste d'url sera enregistré dans un fichier texte. Cette phase

nous donne la possibilité de consulter les requêtes résultantes par la sélection d'une requête parmi une liste des requêtes pour l'ouvrir dans notre navigateur Web. La page ressemblera à ceci « **Voir la figure 10** » :



**Figure 10 : Exemple de la page web**

#### 2.1.4. Construire le corpus

Ceci est la dernière étape, dans laquelle les pages seront nettoyées automatiquement : Le code HTML sera supprimé et le texte sera extrait.

#### Exemple de création d'un corpus de texte avec basic mode :

Nous allons commencer la construction de notre corpus de texte à travers une liste de terme clés (un mot ou une combinaison de mots) qui sont liées au même domaine, dans notre cas domaine de cybercriminalité.

Liste des mots clés  الجريمة الإلكترونية، الهاكر، القانون، اختراق، قرصان الحاسوب

Ensuite nous allons faire une combinaison entre ces termes, pour notre exemple la longueur de chaque combinaison est égale à 2 et le nombre de combinaison ou bien de requête est égale à 3.

Combinaison des mots clés de longueur 2 (Génération de tuples)



الجريمة الإلكترونية ، الهاكر  
 الجريمة الإلكترونية ، القانون  
 الهاكر ، القانون

Par la suite, nous allons utiliser les trois combinaisons précédentes comme des requêtes au moteur de recherche Google, et là nous allons récupérer les n premiers URLs résultant .

L'extraction des URLs est effectuée avec la récupération des pages html de chaque requête et à ce niveau nous faisons l'extraction de ce qui est entre la balise « a » plus spécialement après le paramètre Href.



Après nous passons à l'extraction de texte à partir la liste d'url généré Dans cette étape , et après avoir obtenu la liste des URLs nous nous concentrons sur la pertinence des URLs , alors le seul cas qui ne pose pas problème c'est lorsque nous obtenons le code égale à 200 qui signifie que nous pouvons extraire cette url avec succès en visitant le code html de cette page web et récupérer seulement le texte entre les balises de paragraphe <p> ou titre<h1> jusqu'à <h6> ou la balise de liste triée <ol> ou non triée <ul> , dans les URLs qui reste qui nous donne un code égale à 3xx ou 4xx ou 5xx , l'extraction de ces URLs est impossible

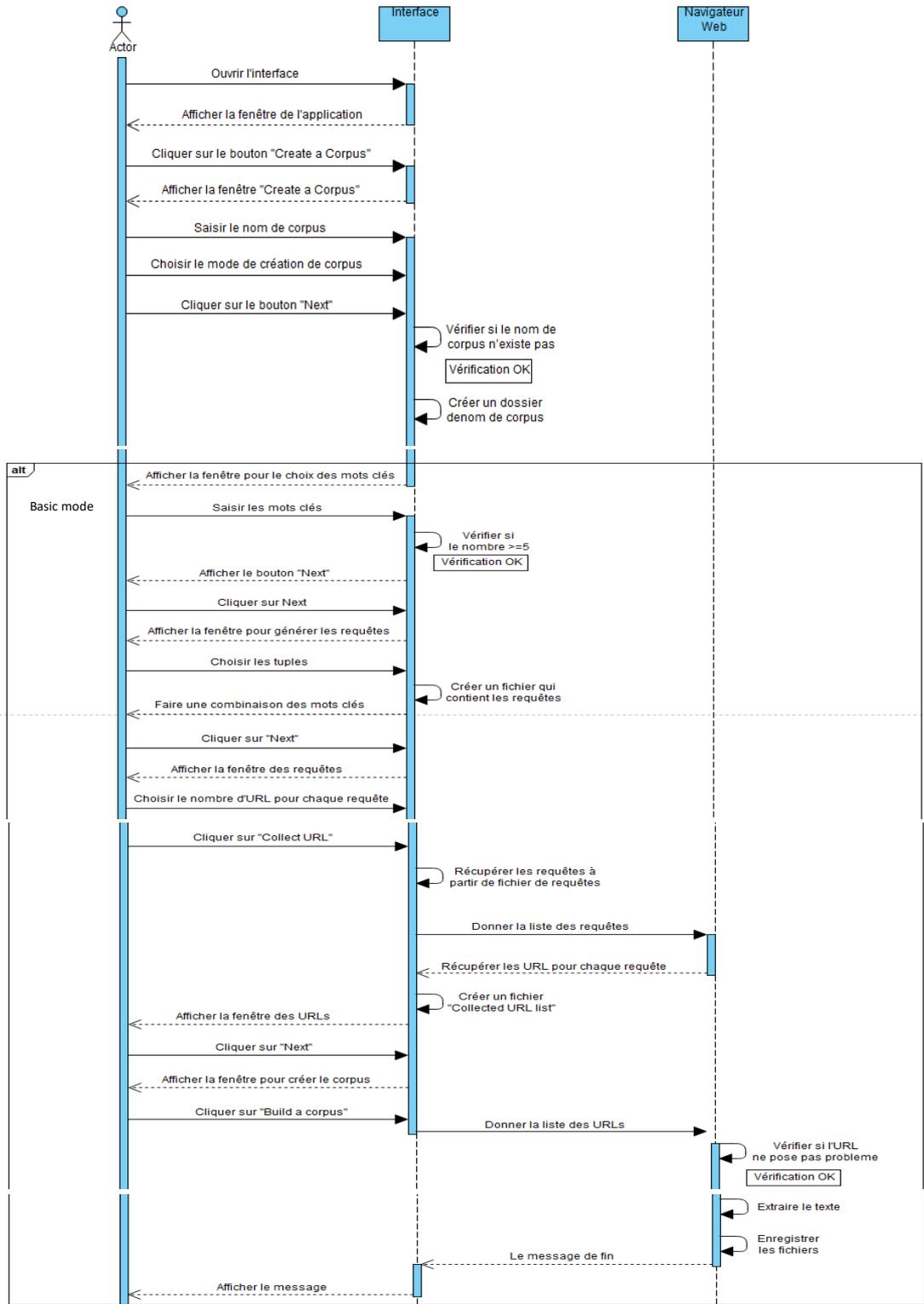
## 2.2. URL personnalisées

Dans ce mode, nous passons directement à la dernière étape, celle où le corpus est construit à l'aide d'une liste d'adresses Internet (ou URL) et se fait en donnant un fichier qui contient des URLs. Après nous devons vérifier si les URLs ne pose pas un problème pour pouvoir les téléchargés dans un seul dossier de format des fichiers (Doc, Docx, PPT, PPTx).Puis nous passons à la dernière étape c'est l'extraction du texte.

## 2.3. Fichiers locaux

Dans ce mode, nous aurons pas besoin de connexion internet car nous traitons des fichiers contenus dans un dossier sur notre ordinateur et qui sont de différent format (Doc, Docx, PPT, PPTx).

Le scénario global de l'ensemble des étapes de notre démarche est illustré par le diagramme de séquences suivant :



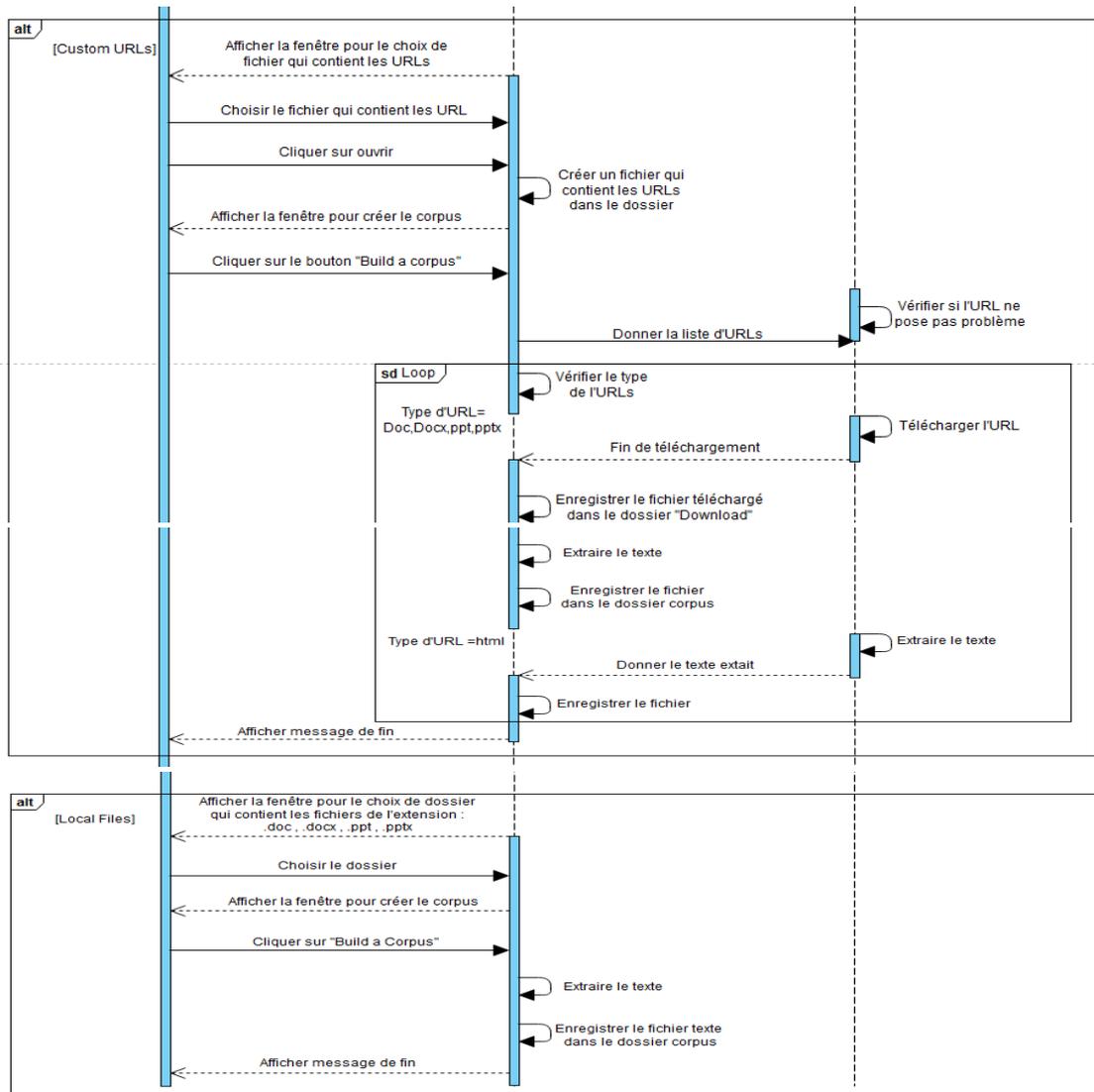


Figure 11 : Diagramme de séquence sur la création d'un corpus avec les 3 modes

### 3. Approche d'évaluation automatique déployée pour l'évaluation du DataSet (CCASAG)

Pour évaluer le dataset (CCASAG) et estimer l'impact du corpus de domaine sur l'évaluation automatique, nous avons choisi de déployer une approche basée sur l'espace sémantique définie dans [62]. Cette approche est basée sur des corpus et elle ne nécessite pas la compréhension du vocabulaire ou de la grammaire de la langue d'un texte. Comme illustré dans la figure 12, l'approche est composée de trois étapes et nous allons la retenir pour l'implémenter en vue de l'évaluation du dataset et du corpus de domaine.

Dans la première étape, il s'agit de créer l'espace sémantique après acquisition et traitement de corpus. Dans les étapes deux et trois, le calcul de similarité est effectué et une note est générée. Dans la suite nous développons ces trois étapes de manière plus détaillé :

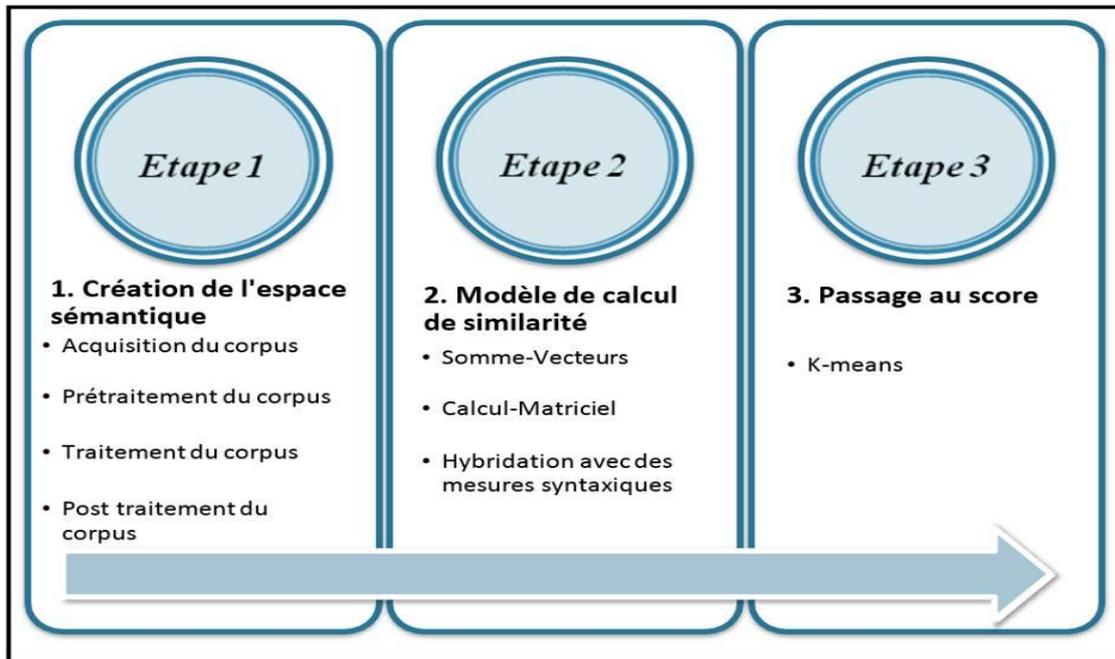


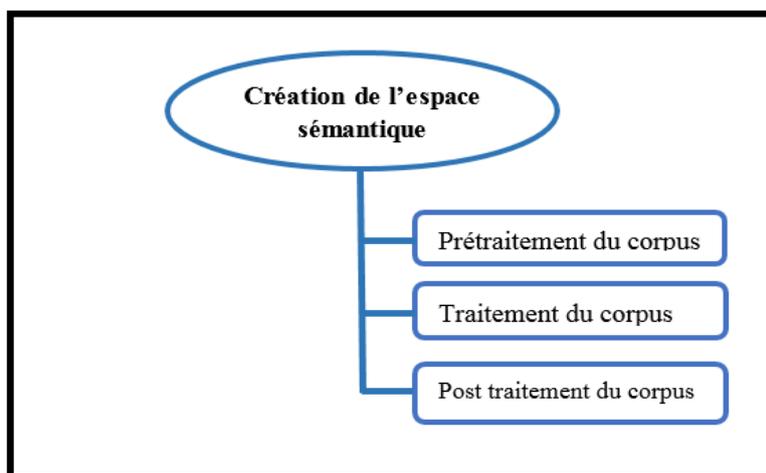
Figure 12 : Schéma des étapes principales [62]

## ETAPE 1 :

### 3.1. Construction de l'espace sémantique

Pour obtenir l'espace sémantique il fallait passer sur quatre étapes, la première c'est l'*Acquisition du corpus* nous avons éliminé cette étape car nous avons construit notre corpus de texte automatiques c'est déjà expliquer dans l'approche précédente (approche de construction de corpus de texte) alors que dans le travail de l'année passée ils ont analysé et traiter des corpus de texte qui existe déjà et qui sont disponibles gratuitement en ligne.

Ensuite nous passons au *prétraitement du corpus* pour avoir un corpus normaliser et stemmer en appliquant des outils de traitement automatique de la langue arabe sur les textes brutes après nous faisons le *traitement du corpus* ou nous allons générer l'espace sémantique après avoir choisir l'une des approches de similarité et au dernier le *post traitement du corpus* qui prend en considération l'importance des mots dans le corpus.



**Figure 13 : Phases de création de l'espace sémantique**

### **3.1.1. Acquisition du corpus du texte :**

Pour faire nous avons utilisé notre propre application avec ses différents outils qui permettant la construction d'un corpus de texte pour la langue arabe à partir du Web en interrogeant automatiquement le moteur de recherche. C'est au moteur de recherche (Google) d'identifier la liste des URL à partir d'une liste de mots. (comme nous avons détaillé dans l'approche précédente de construction de corpus de texte). Alors que dans l'approche basée sur l'espace sémantique de l'année passée ils ont analyser et traiter des corpus de texte qui existe déjà et qui sont disponibles gratuitement en ligne.

### **3.1.2. Prétraitement du corpus du texte :**

Après la collecte des documents textuels du corpus du texte, ce dernier doit être préparé pour les traitements ultérieurs. Cette phase est réalisée en se basant sur un ensemble d'étapes de prétraitement, afin de supprimer une certaine ambiguïté, réduire la quantité des traitement ultérieurs et adapter le corpus de texte à l'objectif final "extraction des candidats-termes ". (c'est le nettoyage des documents en utilisant les outils de traitement automatique de la langue arabe pour choisir le meilleure stemmer).

#### **a) Encodage uniques des textes**

L'encodage unique des textes en format standard, permet de représenter les textes sans aucune déformation au niveau de caractère lors de lecture. Tous les textes de notre corpus de texte sont représentés avec un encodage UTF-8.

## b) Normalisation des textes

Afin de manipuler les variations du texte qui peuvent être représentées en arabe, nous appliquons plusieurs genres de normalisation sur le texte. Par exemple, dans l'arabe écrit, les voyelles sont souvent omises dans les textes, néanmoins, nous pouvons parfois trouver quelques voyelles présentes avec les mots. Alors, l'élimination de ces voyelles est nécessaire pour fin de normalisation. Une autre raison pour ce prétraitement est le fait que nous avons tendance fréquemment à mal écrire ces différentes formes de hamza. Ce genre d'erreurs est très répandu dans les textes arabes

Avant la lemmatisation, le document est normalisé comme suit [63]:

- Suppression des caractères spéciaux et les chiffres par exemple : 1, 2, 3 , ! , » , ( ...
- Supprimer les mots et les caractères latins : les caractères latins sont détectés par leurs graphiques.
- Remplacement de  $\bar{ا}$  ,  $اِ$  et  $أ$  avec  $ا$
- Remplacement de la lettre finale  $ي$  avec  $ى$
- Remplacement de la lettre finale  $ة$  avec  $ه$

Cette étape est nécessaire à cause des variations qui peuvent exister lors de l'écriture d'un même mot arabe. L'extraction se fait à partir du document original ce qui permet de préserver l'intégralité de l'information.

## c) Suppression des mots vides :

Cette étape consiste à éliminer tous les mots vides(en anglais : stop-word), comparer chaque mot reconnu avec les éléments de la liste des mots vides (voir annexe 1) qui contient tous les mots non-significatifs, si un mot en fait partie, il ne sera pas pris en considération. Il s'agit d'une liste de tous les mots d'outil, de liaison et d'articulations (pronoms (الضمائر المنفصلة), prépositions (حروف العطف), conjonctions (حروف الجر), etc.) par exemple : ، على ، عن ، لم في ، التي ، منذ ، ...من ، الى ، هذا ، هذه ، مع ، بعد ، الذي ، انه ، منذ ،

Généralement, les mots vides qui sont très courants (presque la moitié des occurrences des mots du texte) ne sont pas indexés car ils ne sont pas informatifs [64].

## d) Lemmatisation

Un des traitements les plus importants pour la langue arabe, en vue de la recherche d'information et d'indexation est la lemmatisation des mots. Son objectif est d'extraire la racine ou la radicale d'un mot, d'où les raciner sont des systèmes qui font l'extraction de racines des mots entrants et retourne comme résultat les préfixes, les suffixes, le schème<sup>1</sup> utilisé, ainsi la racine et la radicale, ce type d'analyse "simpliste", traite de façon identique affixes flexionnels et dérivationnels. La lemmatisation sert ainsi à la reconnaissance morphologique des mots d'un texte.

Les techniques utilisées pour procéder à la lemmatisation reposent généralement sur une liste d'affixes (suffixes, préfixes) de la langue considérée et sur un ensemble de règles de désuffixation construites à priori qui permettent, étant donné un mot, de trouver son lemme [67]. Un lemmatiseur est un programme de traitement automatique du langage qui permet de passer d'un mot portant des marques de flexion (pluriel, forme conjuguée d'un verbe...) à sa forme de référence (lemme ou forme canonique). Exemple : الحواسيب → "حسب"

En d'autres termes, la lemmatisation transforme différentes formes et variantes d'un certain mot en un seul mot canonique [68].

Il existe deux majeures approches de lemmatisation en langue arabe :

### Le stemming lourd

Le stemming lourd, également appelé « Root-Stemming » (Stemming à la racine), consiste à supprimer les préfixes et les suffixes bien connus pour extraire la racine réelle d'un mot et à identifier le motif en correspondance avec le mot restant.

---

<sup>1</sup> Le schème joue un rôle très important dans le processus de génération des formes dérivées à partir d'une racine ou d'extraire cette dernière à partir d'un mot. Qui est le même mot sauf les lettres de sa racine qui sont remplacées par les lettres de la racine 'فعل'. exp: mot = صالح ----> schème = فاعل

- **Le stem lourd (Heavy Stemming) :** son objectif est d 'extraire le Stem et la racine.



**Figure 14 : Exemple de heavy Stemming**

Parmi les stemmers lourds, nous mentionnons :

### **Khoja Stemmer<sup>1</sup>**

Khoja stemmer [65] est l'un des stemmers les plus proches de l'extraction manuelle de la racine. Les décisions prises par Khoja stemmer sont statiques, où il a une justification linguistique pour chaque décision. Mais, il ne capture pas la dynamique de la langue et n'explore pas toutes les possibilités linguistiques. Khoja stemmer [69] est l'un des premiers et plus puissants stemmer développés pour l'arabe [70], [71]. Il [61] a été le plus utilisé dans la recherche d'informations, suivi de la fouille et de la classification et enfin l'extraction de l'information.

Le Khoja stemmer suit cette procédure :

1. Supprimer les signes diacritiques représentant une voyelle.
2. Supprimez les mots vides, la ponctuation et les chiffres.
3. Supprimer l'article défini "ال"
4. Supprimez la conjonction inséparable «و».
5. Supprimer les suffixes.
6. Supprimer les préfixes.
7. Faites correspondre le résultat à une liste de modèles. Si une correspondance est trouvée, extrayez les caractères dans le motif représentant la racine.
8. Faites correspondre la racine extraite à une liste de racines "valides" connues.
9. Remplacez les lettres faibles "ا, ي, و" par "و"

<sup>1</sup> <http://zeus.cs.pacificu.edu/shereen/research.htm#stemming>

10. Remplacer toutes les occurrences de hamza “ء, ؤ, ئ, ة” par “أ”

11. Les racines de deux lettres sont vérifiées pour voir si elles doivent contenir un double caractère. Si oui, le caractère est ajouté à la racine.

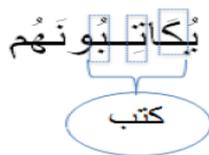
### ✚ Stemming léger :

Le stemming léger est un processus moins complexe, où le stemming est arrêté sur la suppression des préfixes et des suffixes, sans tenter d'identifier la racine réelle du mot.

De nombreux stemmers légers ont été recommandés pour la Langue arabe [72].

Pour atténuer l'impact de l'inconvénient majeur des algorithmes basés sur la racine, qui fait perdre la notion de sémantique, une lemmatisation légère provenant de l'arabe a également été proposée. Stemmers légers coupent quelques affixes tels que les terminaisons plurielles en anglais à partir des mots et sans effectuer d'analyse linguistique profonde. De ce point de vue, la majorité des approches tentent de supprimer les préfixes les plus fréquents (c'est-à-dire articles), les suffixes (c'est-à-dire les pronoms possessifs) et tous les préfixes ou suffixes cela peut être attaché au début ou à la fin des mots. Par exemple, les stemmers légers génèrent أعمال (ce qui signifie : œuvres) de أعمالهم parce que seuls les préfixes (y compris les antéfixes) et les suffixes (y compris les post fixes) sont supprimés. La décision de retirer cependant, les affixes sont généralement contrôlés par des règles heuristiques dérivées de l'utilisation commune de ces antéfixes [73].

➤ **Le stem léger (Light Stemming) :** s'intéresse seulement par le Stem.



**Figure 15 : Exemple de light Stemming**

Parmi les stemmers léger, nous mentionnons :

### **Tashaphyne stemmer<sup>1</sup>**

Tashaphyne est un stemmer arabe de poids léger (Zerrouki, 2010) [75]. Il commence par supprimer les lettres non liées dans le processus d'extraction de racine, tels que les signes diacritiques. Il utilise deux listes de préfixes et suffixes pour segmenter un mot donné. Tashaphyne fournit à la fois une tige légère ou une racine au mot d'entrée.

### **3.1.3. 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<sup>2</sup> est adopté qui ne prend pas en considération 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. Selon cette approche, on construit un tableau croisé «documents × Attributs» où les lignes correspondent aux Attributs et les colonnes correspondent aux documents. En outre, une fenêtre avec une taille prédéfinie est importante pour concrétiser la notion du voisinage.

La « **Figure 16** » montre le processus de génération de l'espace sémantique. Les détails de ce processus sont ci-dessous

---

<sup>1</sup> <https://pypi.org/project/Tashaphyne/>

<sup>2</sup> 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

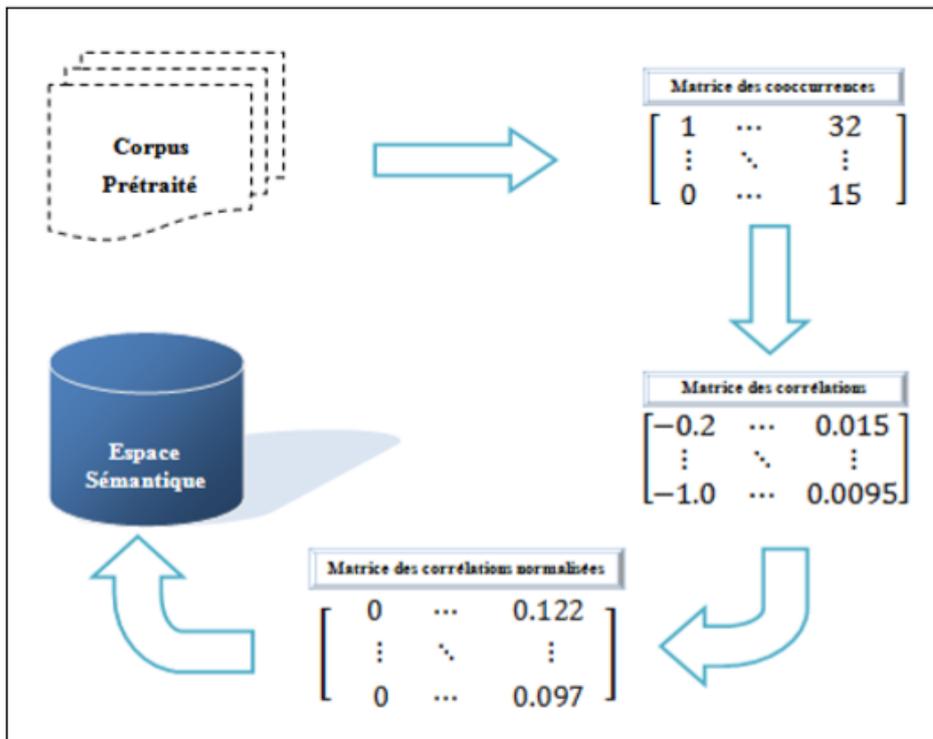


Figure 16 : le processus de génération de l'espace sémantique

- a) Le corpus est acquis et prétraité 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, nous appliquons une stratégie de normalisation. Nous allons alors construire une nouvelle matrice en appliquant la formule ci-dessous prise de l'algorithme COALS [56][35] pour chaque case/élément de la matrice des cooccurrences.

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

D'où :

- **a** et **b** sont les deux termes de la matrice de cooccurrences (ligne et colonne).
- **wa**, est 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.
- $\sum w_a$ , est la somme des colonnes de la ligne du terme a.
- $\sum w_i$ , est la somme des lignes de la colonne du terme b.
- $T = \sum \sum w_{i,j}$ , est la somme de tous les éléments de la matrice de cooccurrences.

Pour un large corpus, les valeurs de corrélations sont petites, alors, il est rare que la valeur de corrélation dépasse le **0.01**. De plus, la majorité des corrélations sont négatives (81.8%) [76].

#### 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 [76].

Cette étape représente la dernière étape de création de l'espace sémantique ou proprement dit : les vecteurs du contexte de chaque mot du corpus en entrée. Il est à noter que l'espace sémantique est généré qu'une seule fois. Il est donc stocké pour une utilisation ultérieure.

### 3.1.4. Post traitement du corpus du texte

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 [77] :

- 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.

#### ✚ **Fréquence du terme (Term Frequency " TF ") :**

Elle mesure l'importance d'un terme dans un document. cette mesure est proportionnelle à la fréquence du terme dans le document. L'idée sous-jacente est que plus un terme est fréquent dans un document, plus il est important dans la description de ce document. C'est le nombre d'occurrences de ce terme dans le document considéré.

$$\text{TF} = \left\{ \begin{array}{l} \text{Freq (t,d)} \\ \text{TF}_{\log} = -\log \left( \frac{\text{TFcount}}{n} \right) \\ \text{TF}_{\min-\max} = \frac{\text{TFlog}}{\max(\text{TFlog})} \\ \frac{\text{Freq(t,d)}}{\sum_{t' \in d} \text{Freq}(t',d)} \end{array} \right.$$

D'où :

- t : terme et d : document
- TFcount : le nombre d'occurrence du mot dans le corpus et n est le nombre total de mot dans le corpus
- TFlog=le nombre de fois que le mot apparaît dans le corpus [78]

Supposons que nous avons un corpus contenant n document (D1, D2,..., Dn) et un terme t. la valeur TF(t, D1) est le nombre d'occurrence du terme t dans le document D1 seulement.

## ✚ Fréquence inverse du document (Inverse Document Frequency "IDF") :

Mesure l'importance d'un terme dans toute la collection. Les termes qui apparaissent dans peu de documents de la collection sont plus représentatifs du contenu de ces documents que ceux qui apparaissent dans tous les documents de la collection. Cette mesure est utilisée pour désigner la pondération globale. Elle vise à donner un poids plus important aux termes les moins fréquents, considérés comme plus discriminants (par exemple les stop-words ont généralement une fréquence élevée alors qu'ils ne sont pas vraiment importants par rapport aux autres mots moins fréquent qu'eux).

$$IDF_i = \frac{|D|}{|\{dj : t_j \in dj\}|} \quad (1)$$

D'où :

- $|D|$  : nombre total de documents dans le corpus.
- $|\{dj : t_j \in dj\}|$  : nombre de documents où le terme  $t_j$  apparaît [79].

## ✚ TFIDF

La combinaison des deux mesures Tf-Idf(TF \* IDF) est une approximation de l'importance d'un terme dans un document. Ainsi, une valeur de Tf\*Idf élevée signifie que ce terme est important dans le document et apparaît peu dans les autres documents de corpus.

$$TFIDF = TF * IDF \quad (2)$$

Cette mesure a eu un succès limité dans les corpus de tailles très variables.

Mais L'inconvénient de la mesure TF \* IDF est qu'elle ne tient pas compte de la longueur de document, en effet, un terme dans un document long a plus de chances d'apparaître plusieurs fois qu'un autre terme dans un document court. Plus le document est long, plus les termes utilisés se répètent. La longueur des documents peut aussi induire l'utilisation d'un grand nombre de termes pour décrire un sujet. Pour pallier ce type de problèmes on a éliminé cette étape et la remplacer par un algorithme qu'on a créé pour diminuer la matrice de cooccurrences

d'un côté et réduire la complexité de notre application d'un autre cotés cet algorithme calcule le nombre d'occurrences d'un mot significatif dans un document.

Les formules adoptées au sein de ce travail sont :

- La formule  $TF_{minmax}$  pour le calcul des TF,
- La formule (1) pour le calcul des IDF,
- La formule (2) pour le calcul des TFIDF.

Après avoir construit notre espace sémantique, nous sauvegardons ce dernier dans un fichier « .txt ». Chaque ligne de ce fichier représente un vecteur d'un mot. Les éléments d'un vecteur du mot sont séparés par des espaces simples. Chaque élément est sous format chaîne de caractères « String ».

**ETAPE 2 :**

### **3.2. Modèles du calcul de similarité sémantique entre deux réponses courtes**

Après avoir construit l'espace sémantique, nous entamons la notion du calcul de similarité sémantique. Dans notre travail, nous avons considéré deux niveaux du calcul de similarité :

- a) Similarité entre phrases : nous avons réalisé ce concept par un modèle que nous nommons « somme vecteurs ».
- b) Similarité mot-à-mot : nous avons concrétisé ce concept par un modèle de calcul matriciel.
- c) Une hybridation des deux modèles précédents.

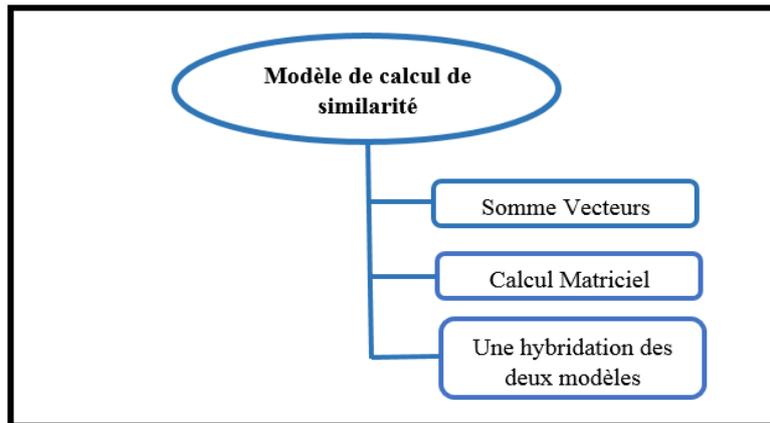


Figure 17 : Modèles du calcul de similarité sémantique entre deux réponses courtes

### 3.2.1. Le modèle somme-vecteurs (SV)

Soit deux phrases,  $RM = W_1, W_2, \dots, W_i$  et  $RE = W'_1, W'_2, \dots, W'_j$ , leurs vecteurs étant respectivement  $(V_1, V_2, \dots, V_i)$  et  $(V'_1, V'_2, \dots, V'_j)$ . Pour mesurer la similarité sémantique entre RM et RE, nous effectuons la somme (pondérée) des vecteurs de leurs mots. Nous avons utilisé deux méthodes proposées pour calculer la représentation vectorielle d'une phrase. « Voir la figure 18 »

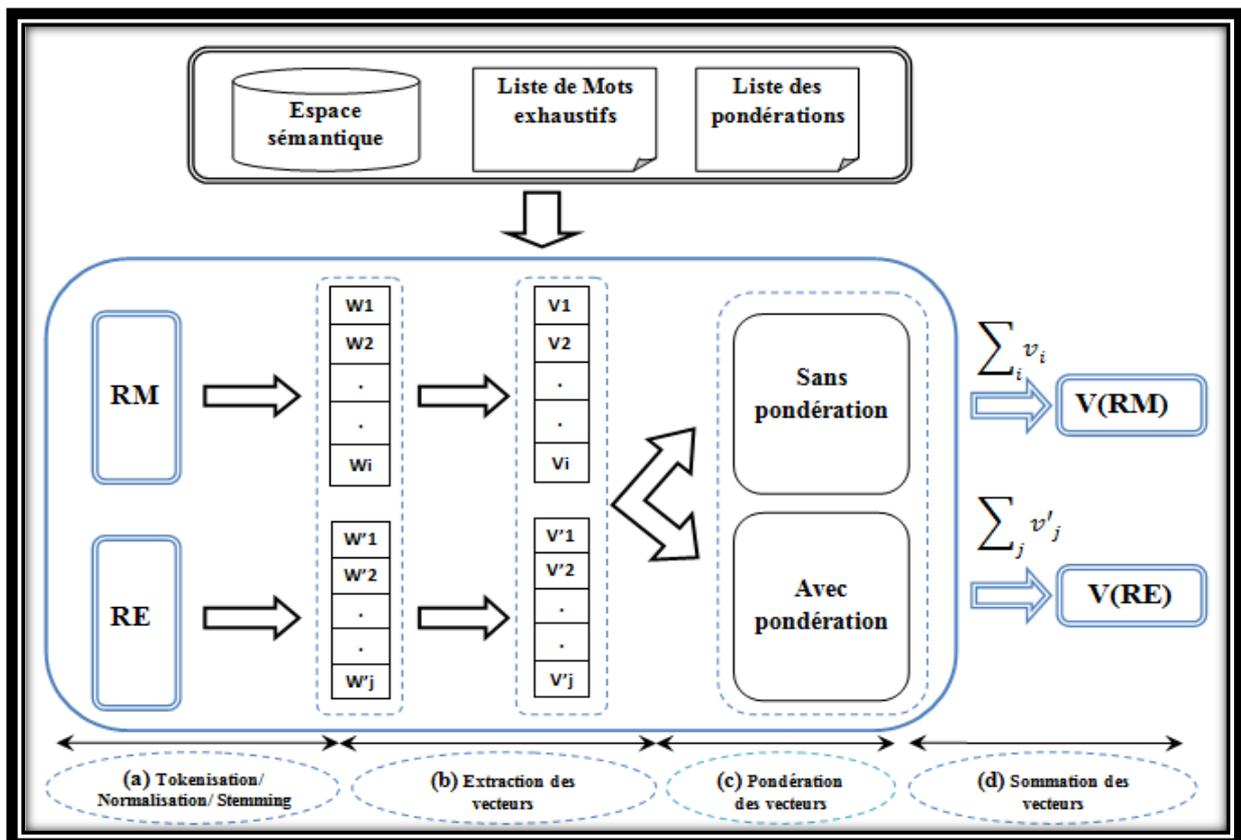


Figure 18 : Vue globale sur le fonctionnement du modèle SV

Nous détaillons les étapes du fonctionnement de la SV représentées au niveau de la « Figure 18» :

- a) Une étape de prétraitement (normalisation, stemming,...) est effectuée sur les deux réponses (RM et RE). Par la suite, chaque réponse est transformée en vecteur en appliquant une tokenisation. Deux vecteurs de mots résultent de cette étape.
- b) Maintenant, les mots de chaque réponse sont remplacés par leurs vecteurs de contexte. Cela est fait en récupérant le vecteur de contexte du mot à partir de l'espace sémantique construit et stocké précédemment. Un vecteur de vecteurs résulte à ce stade pour chaque réponse.
- c) Ensuite, nous allons récupérer des valeurs de pondération de chaque mot. Ici nous distinguons deux cas : avec une pondération et sans aucune pondération. Dans le cas de pondération, chaque vecteur de contexte du mot  $i$  est multiplié par la valeur de pondération qui représente l'importance du mot  $i$ .
- d) Chaque vecteur de vecteurs de contexte passe par une étape de sommation. Cela est obtenue en sommant les vecteurs de contexte de chaque réponse mot par mot ou proprement dit case par case.
- e) En arrivant au stade du calcul de la valeur de similarité entre les deux réponses. La mesure **Cosine** sur les vecteurs  $V(RM)$  et  $V(RE)$  est appliquée. Une valeur entre 0 et 1 est obtenue et qui représente la valeur de similarité entre les deux réponses.

### 3.2.2. Le modèle calcul-matriciel (CM)

Dans ce modèle nous construisons une matrice dont les termes de la 1<sup>ère</sup> réponse représentent les lignes tandis que les termes de l'autre réponse représentent les colonnes. Par la suite, chaque élément de la matrice représente une valeur de similarité entre les deux vecteurs de contexte du  $i^{\text{ème}}$  terme (réponse 1) et celui du *jème* terme (réponse 2).

### 3.2.3. Hybridation

Dans cette phase-là, nous visons à concrétiser la notion d'hybridation entre les mesures afin d'avoir de meilleurs résultats. Pour ce faire, nous allons adopter une combinaison, nous combinons les résultats des deux modèles de calcul précédents en deux manières :

#### 1) La moyenne :

Dans ce type, nous prenons la moyenne des deux valeurs de similarité (calculé par SV et CM) entre les deux réponses. Les résultats seront présentés dans le chapitre suivant.

#### 2) Le maximum :

Contrairement à la moyenne, dans ce cas, nous prenons la valeur maximum entre les deux valeurs de similarité. De même, les résultats dans le prochain chapitre.

### ETAPE 3 :

#### 3.3. Passage au score :

Après avoir mesuré la valeur de similarité entre les deux réponses, nous allons effectuer un passage au score de la valeur de similarité vers un note ou score selon le barème donné. Pour ce faire, nous adoptons la technique suivante :

#### Classifieur k-means [80] (Ou k-moyenne)

C'est une technique de classification non supervisée (est une méthode de partitionnement de données). Elle consiste à avoir un ensemble de données comme entrée (input) et elle donne en sortie ces données classifiées dans K classes. C'est un algorithme très populaire et facile dans sa mise en œuvre. L'inconvénient de cet algorithme réside dans l'exigence de la détermination du nombre de classe comme entrée.

Les couples de réponses dataset utilisés dans l'évaluation de cette approche sont notés sur 5 points. K est fixé à 11 pour avoir 11 classes de note (0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5) afin de prendre en compte la subjectivité du processus d'évaluation.

## **Conclusion**

Au cours de ce chapitre, nous venons de présenter l'approche de création d'un corpus de texte et la création d'un corpus de test (Dataset). Comme nous avons utilisé l'approche développé l'année passée pour créer notre espace sémantique ainsi que pour la génération des pondérations dans le but de tester son impact. Ensuite nous avons calculé la similarité qui sera convertis en score par la suite.

Les résultats des tests ainsi que ceux des hybridations seront présentés dans le chapitre suivant.

# Chapitre 4 : Résultats expérimentaux et Evaluation

Dans ce chapitre, nous montrons les expériences et les résultats des tests réalisés afin d'estimer la qualité de notre système et sa performance comparée aux autres systèmes qui existent, ainsi nous discutons les résultats obtenus.

## 1. Démarche expérimentale

### La démarche de construction de dataset :

La construction du dataset permet à un chercheur avantage d'exportation, pour construire une banque de question pour l'enseignement. Car notre outil de construction permet à faire des différents traitements :

- Ajouter des questions.
- Ajouter des réponses.
- Ajouter un nouveau type de questions .
- Extraire les fichiers xml, txt, doc.
- Calculer la similarité avec 2 possibilités :
  - Entre un couple de réponse (réponse modèle et réponse étudiant).
  - Entre deux fichiers texte l'un des réponses modèle et le deuxième des réponses étudiant. Pour l'évaluation du dataset.

### Outils développés :

Notre application englobe un système d'évaluation automatique des réponses courtes ainsi que d'autres outils du traitement du langage naturel que nous avons dû développer à cause du manque d'outils de traitement pour la langue arabe. Ce système se compose de différents outils intégrés entre eux présentés ci-dessous :

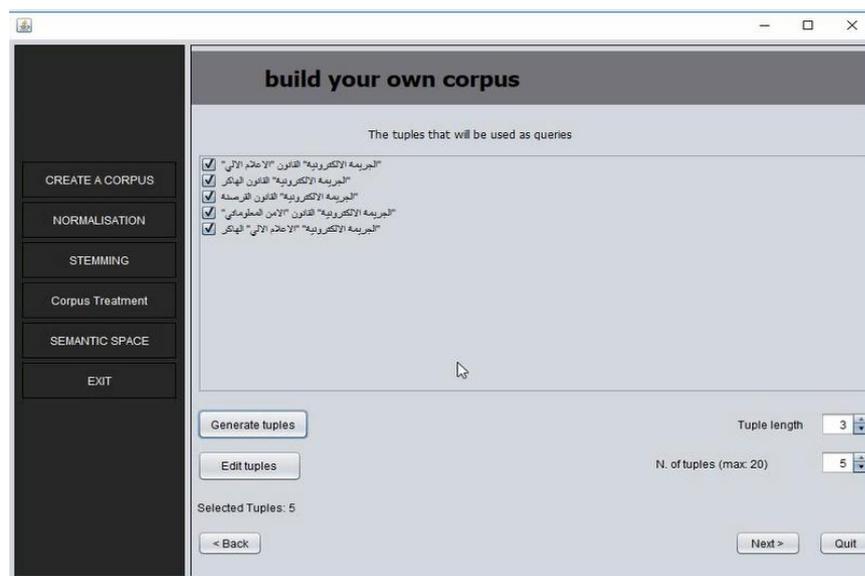
- Outil de création automatique de corpus de texte et d'espace sémantique.
- Outil de manipulation du dataset.
- Un outil d'évaluation.

### 1) Outil de création automatique de corpus de texte et d'espace sémantique :

Dans la création de corpus de texte nous avons construit 3 modes de création de corpus:

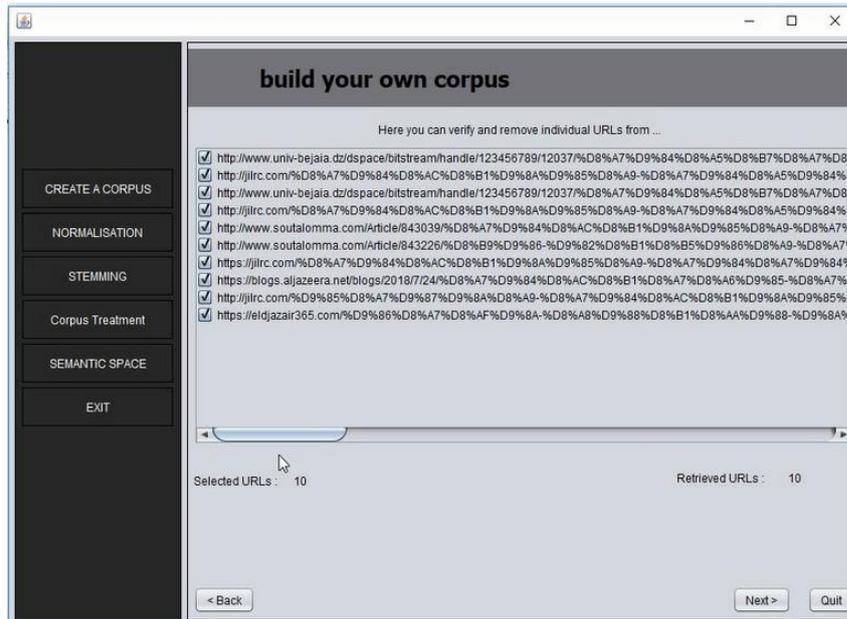
#### a) Mode de base :

Dans la première étape de création automatique du corpus nous devons entrer les termes clés (des mots ou une combinaison des mots) au minimum 5 termes, après nous passons à la 2ème partie qui représente la combinaison de ces termes clés (**voir la figure 19**)



**Figure 19 : La phase de combinaison des termes clés**

Après nous allons générer ces tuples et à l'étape suivante nous avons la possibilité de choisir parmi les requêtes résultantes pour l'ouvrir dans notre navigateur web. Ensuite et à l'étape de collecte des URLs nous pouvons choisir le nombre d'URLs pour chaque requête.



**Figure 20 : La phase de collecte des URLs**

Et après d’avoir récupérer la liste des URLs nous passons à la dernière étape de création de corpus de texte, dans laquelle les pages seront nettoyées automatiquement : Le code HTML sera supprimé, et le texte sera extrait.

**b) URL personnalisées**

Dans ce mode, nous passons directement à la dernière étape, celle où le corpus de texte est construit à l'aide d'une liste d'adresses Internet (ou URL).

**c) Fichiers locaux**

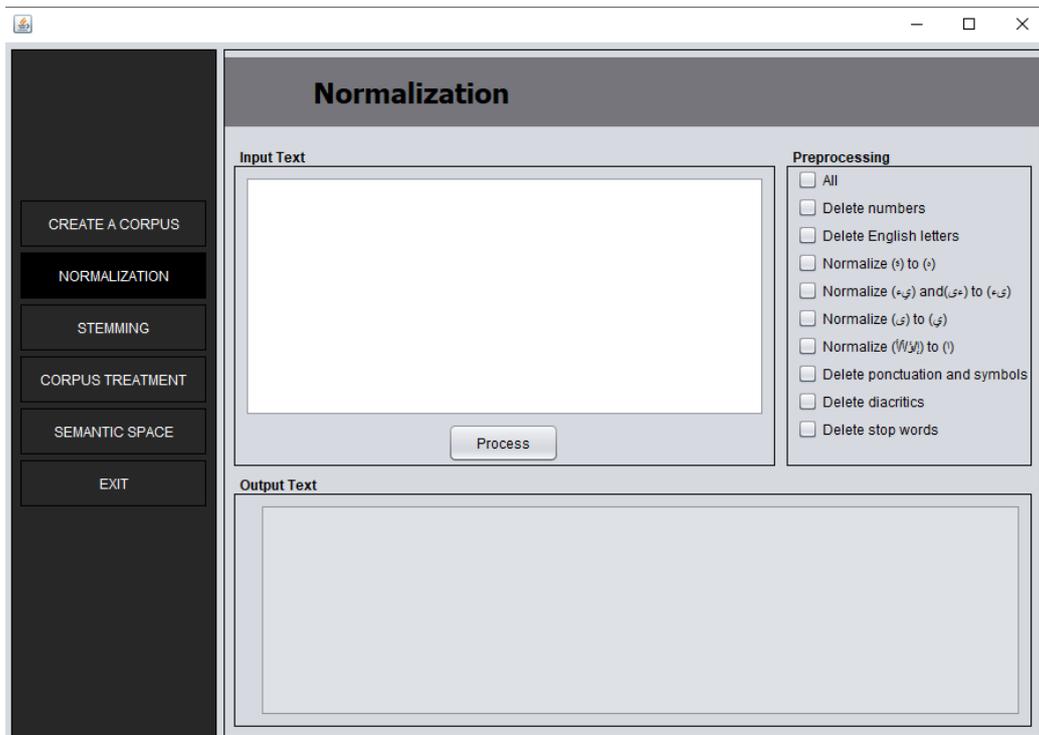
ce mode traitera tous les fichiers contenus dans un dossier sur notre ordinateur.

**De plus de la création d’un nouveau corpus de texte, notre outil nous donne la possibilité d’enrichir un des corpus déjà crée.**

Nous passons au traitement du corpus du texte pour créer l’espace sémantique :

➤ **Outils de Normalisation :**

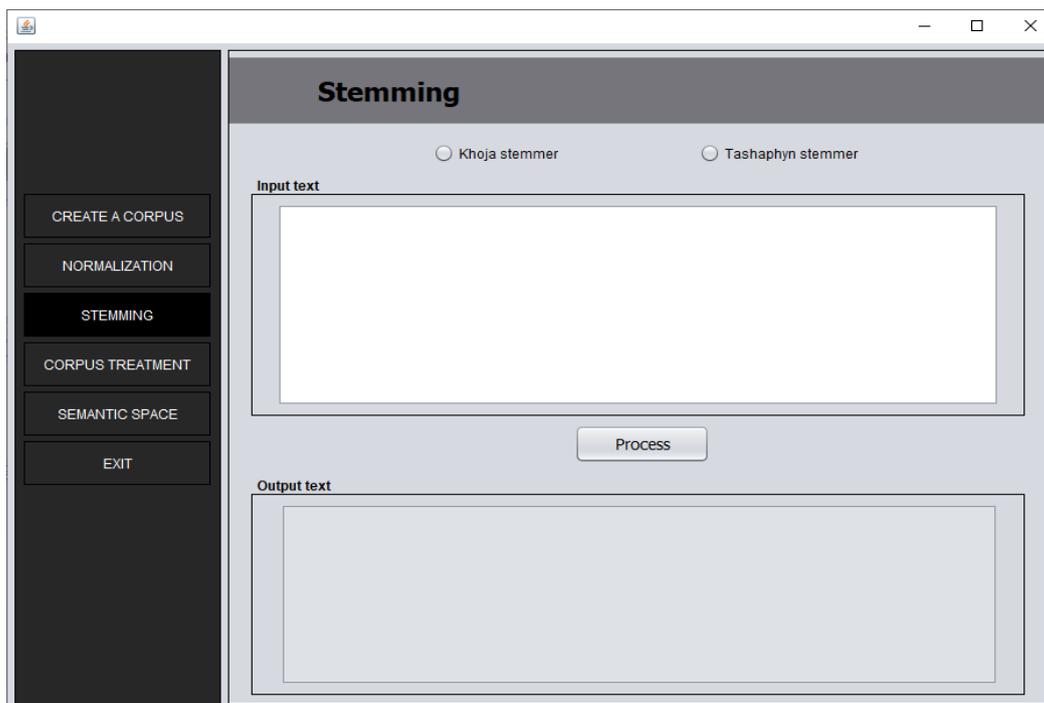
Dans cette partie, nous donnons la possibilité à l’utilisateur de percevoir le changement effectué sur un texte donné en entrée. L’utilisateur peut choisir l’opération de la normalisation en spécifiant les paramètres à effectuer parmi une liste de paramètres (suppression des numéros, suppression des diacritiques...), « voir Figure 21».



**Figure 21 : Outils de normalisation**

➤ **Outils de stemming :**

L'opération du stemming en spécifiant le stemmer (khoja ou Tashaphyne). « voir Figure 22 ».



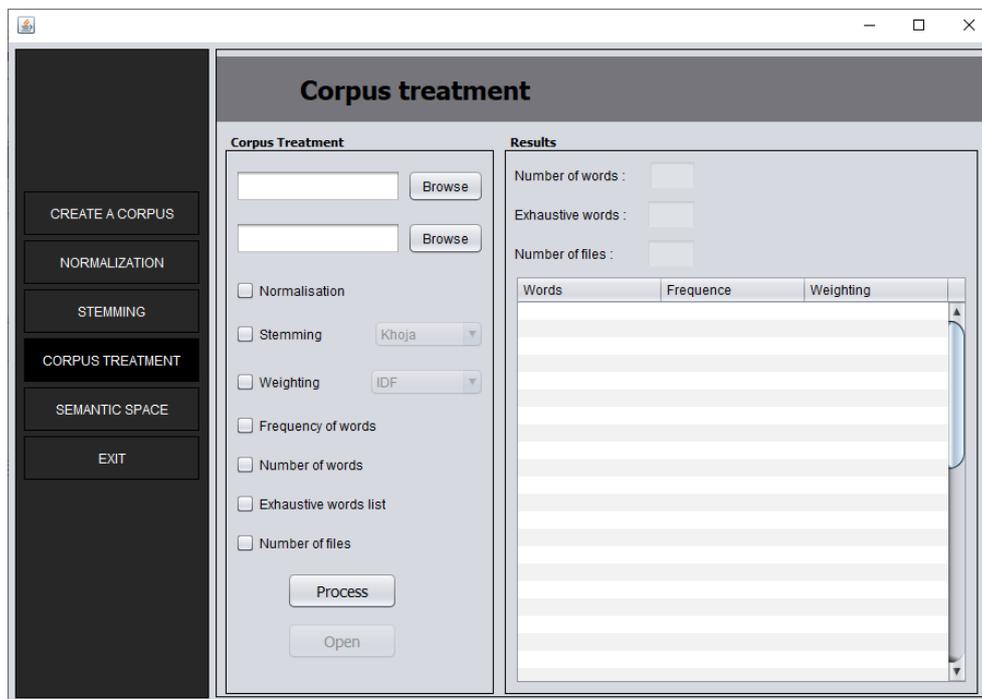
**Figure 22 : Outils de stemming**

➤ **Outils NLP**

Cette partie permet de faire différents traitements sur un corpus de texte choisi par l'utilisateur :

- 1) La normalisation,
- 2) Le stemming (khoja et tashapyne),
- 3) Les pondérations (TF, IDF, TF-MIN-MAX),
- 4) Fréquence des mots et leur nombre total,
- 5) Liste des mots exhaustifs.

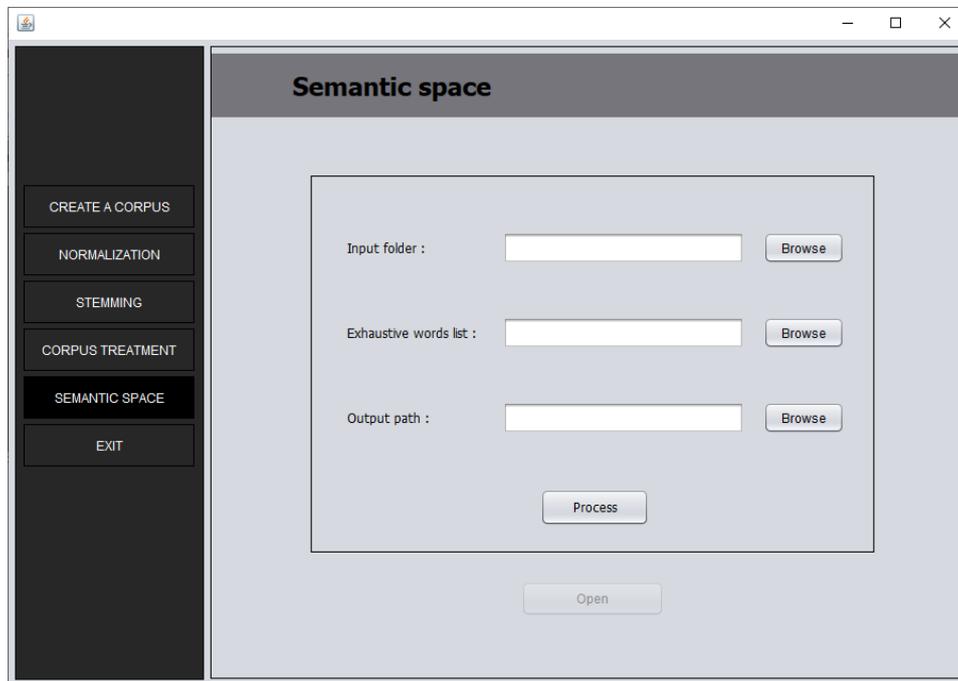
L'utilisateur doit donner en entrées le corpus de texte sous forme d'un dossier contenant que des fichiers texte, il doit davantage indiquer le chemin du dossier de sortie, et choisir un ou bien plusieurs traitement « voir Figure 23. ».



**Figure 23 : Outil NLP**

➤ **Outil de création de l'espace sémantique**

Cet outil permet de calculer l'espace sémantique d'un corpus de texte donné par l'utilisateur en entrée et il a le choix de travailler avec une liste des mots exhaustifs externe ou bien avec la liste générée automatiquement. L'utilisateur doit donner en entrée le chemin du dossier de sortie où les résultats vont être sauvegardés « voir Figure 24».

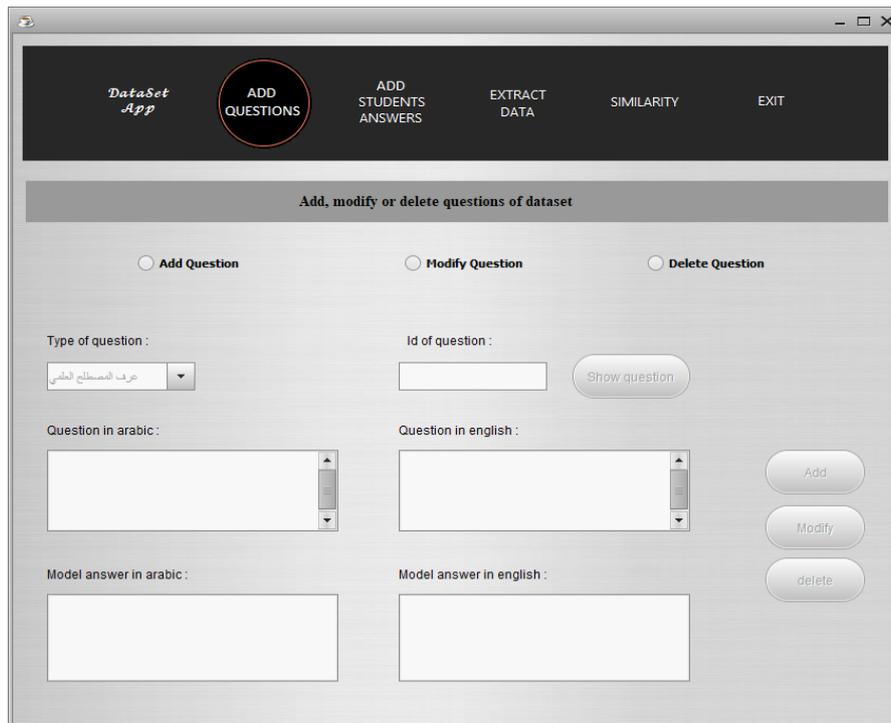


**Figure 24 : Outil de création de l'espace sémantique**

2) **Outil de manipulation du dataset :**

➤ **Add questions (Ajouter des question) :**

Dans cette partie, nous donnons la possibilité à l'utilisateur d'ajouter une question soit en arabe ou bien en anglais et choisir un type parmi la liste des types de question ou d'ajouter un autre type et donner pour chaque question son propre id, et pour l'opération de la suppression et modification l'utilisateur doit saisir le id de la question qui veut la modifier après cliquer sur le bouton « show question » pour obtenir l'affichage de cette dernière. « Voir la figure 25»

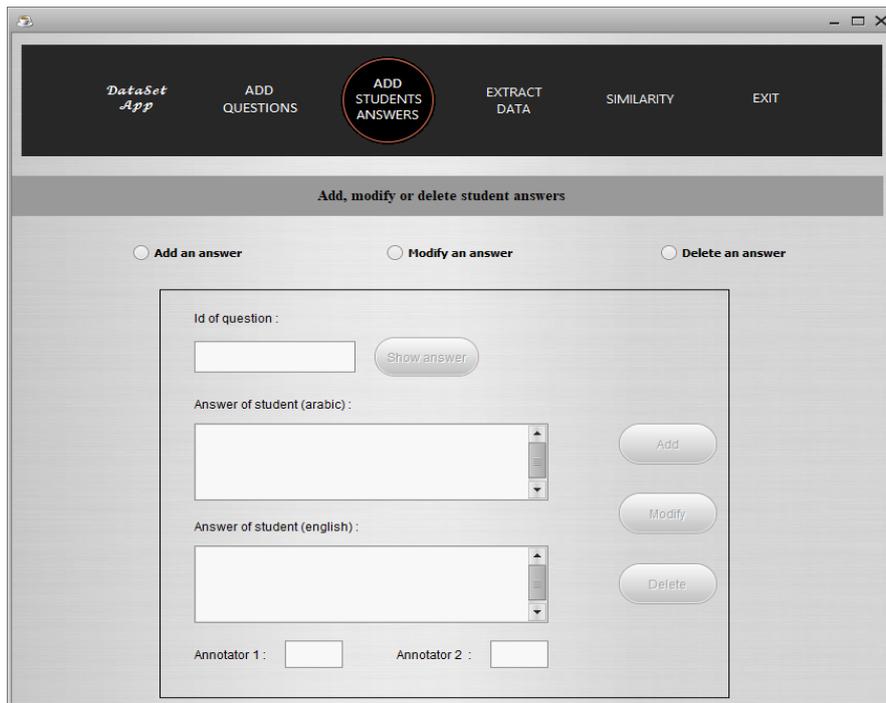


**Figure 25 : L'ajout des questions avec la réponse modèle sur le dataset**

✚ Comme nous avons la possibilité d'ajouter un type de question si la liste des types ne le contient pas.

➤ **Add Student Answers (Ajouter les réponses des étudiants):**

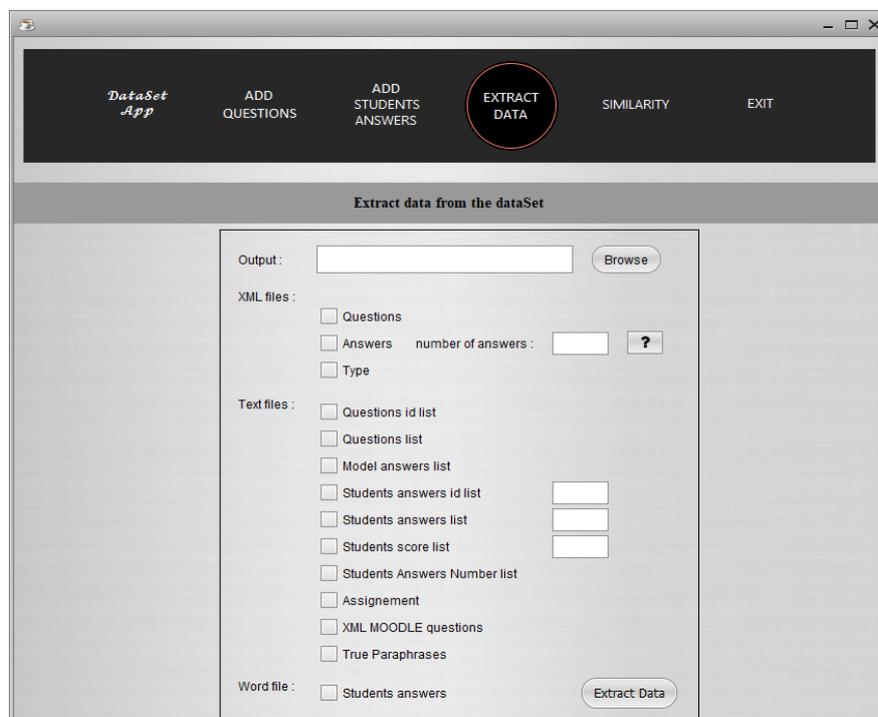
Dans cette opération l'utilisateur fait entrer la réponse d'étudiant avec l'identifiant de question qui convient qui est codifié à partir de la concaténation de l'identifiant de type avec l'identifiant de question en les séparant par point, et les scores du deux enseignant, par la suite l'identifiant de réponse sera attribué automatiquement qui est de même sera codifié en concaténant l'identifiant précédent avec celui-là. Plus la possibilité de la suppression et modification de telle réponse représentée par son identificateur. **(Voir la figure 26)**



**Figure 26 : L'ajout de réponses étudiantes avec leurs scores sur le Dataset**

➤ **Save as xml, txt ,doc (Enregistrer sous format XML, Txt, doc):**

C'est l'extraction des données (questions, réponses des étudiants, les scores des étudiants...) à partir du Dataset et les sauvegarde soit en fichier xml, fichier txt ou document word. « Voir la figure 27 »

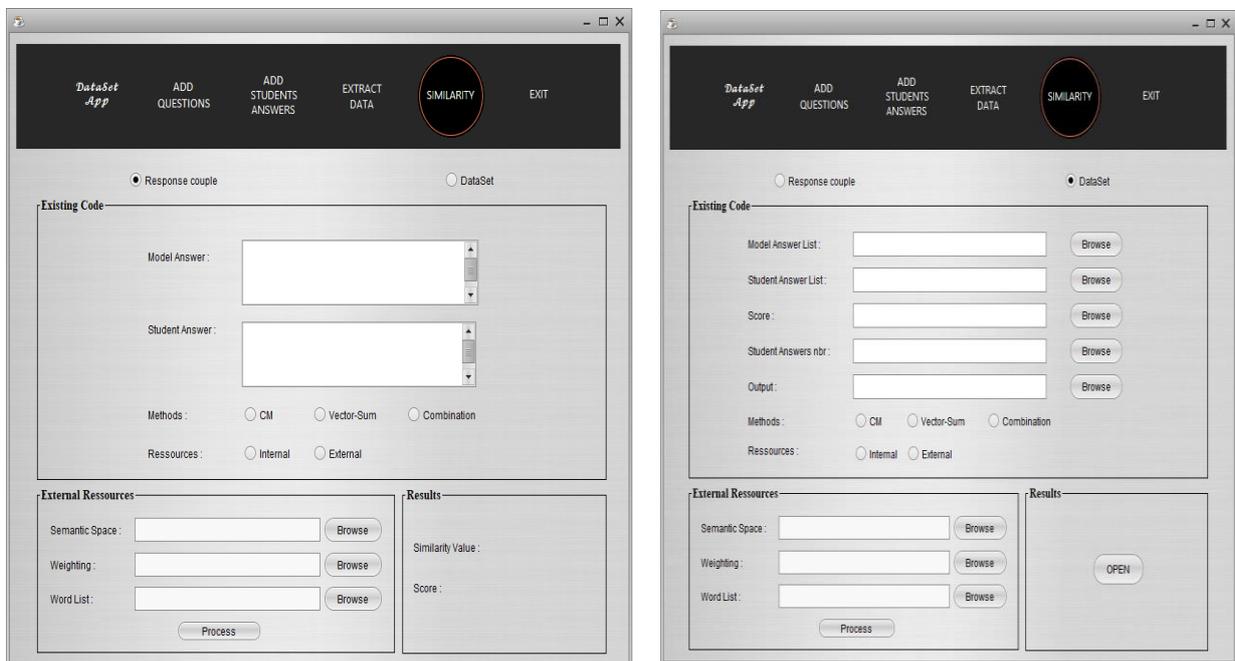


**Figure27 : L'enregistrement de différentes versions des fichiers**

### 3) Outils d'évaluation automatique des réponses courte (Similarity) :

Ce module nous permet de calculer la similarité entre deux réponses courtes ainsi qu'un DS avec nos modèles et leurs combinaisons détaillées dans le chapitre précédent. Il permet aux utilisateurs de choisir de faire le calcul avec des ressources internes (Espace sémantique et pondération déjà calculés) ou bien de choisir des ressources externes « voir Figure 28».

Du côté DS, l'utilisateur a la possibilité d'évaluer un DS de son choix c'est à dire généré les notes automatiques, il doit avoir en entrée deux fichiers texte, l'un présente les réponses modèle et l'autre les réponses des étudiants en choisissant le barème ainsi il doit choisir un dossier de sortie. Tandis que pour le côté des deux réponses, l'utilisateur doit saisir deux réponses (RM et RE) et choisir le reste des paramètres. Le résultat s'affiche directement dans l'interface.



**Figure 28 : Outil d'évaluation automatique des réponses courtes**

## 2. Le module d'évaluation :

L'évaluation d'un système implémenté ou d'une approche proposée est indispensable pour estimer le succès d'une recherche. Il devient primordial d'accorder un rôle central aux métriques d'évaluation qui consiste à comparer un résultat produit avec des résultats corrects attendus. L'analyse de plusieurs situations d'évaluation dans notre cas, illustre l'importance d'un choix cohérent des métriques et de l'utilisation conjointe de plusieurs métriques. En essayant d'analyser les résultats de ce travail, nous avons été confrontés à la détermination de la métrique à utiliser pour évaluer les scores obtenus par rapport aux scores manuels fournis.

La corrélation de Pearson [44] est la métrique la plus fréquemment utilisée par les recherches dans ce domaine. Bien qu'elle ne soit pas citée et utilisée dans la majorité des travaux connexes, nous avons choisi d'inclure conjointement au coefficient de Pearson, l'erreur quadratique moyenne (Root Mean Squared Error (RMSE)[45]) pour quantifier la différence (ou le décalage) entre le résultat (score) obtenu par le système et celui obtenu par l'expert humain, et à les comparer avec les résultats obtenus en utilisant d'autre corpus de texte (CNN, BBC, khaleej)

### 2.1. Coefficient de Pearson (CP)

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 **CP**. 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.

Le tableau 3 représente les détails sur la corrélation de Pearson :

**Tableau 3 : Valeurs de corrélation de Pearson**

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

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 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.

En conclusion, l'évaluation de nos approches correspond à trouver la meilleure minimisation de l'erreur quadratique avec une maximisation du coefficient de corrélation.

De plus les tests qu'on a fait avec notre corpus de cyber on a utilisé et traiter aussi des corpus disponibles gratuitement en ligne :

- ✚ **CNN Arabic Corpus (2010)**<sup>1</sup> Ce corpus a été collecté du site web CNN Arabic *cnnarabic.com*, il comprend 5070 documents texte. Chaque document texte appartient à l'une des six catégories (Business 836, Divertissement 474, Nouvelle de Moyen-Orient 1462, Science et technologie 526, Sports 762, Nouvelle du monde 1010). Le corpus contient 2.241.348 (2.2M) mots et 144.460 mots-clés de district après le retrait des mots vides [34].

---

<sup>1</sup> <https://sourceforge.net/projects/ar-text-mining/files/Arabic-Corpora/>

✚ **BBC Arabic Corpus (2010)**<sup>1</sup> BBC Arabic corpus a été collecté du site web BBC Arabic bcarabic.com, le corpus comprend 4 763 documents texte. Chaque document texte appartient à l'une des sept catégories (Nouvelles du Moyen-Orient 2356, Nouvelles du monde 1489, Business et Economie 296, Sports 219, Presse internationale 49, Science et technologie 232, Art et culture 122). Le corpus contient 1.860.786 mots (1.8M) et 106.733 mots-clés de district après le retrait des mots vides [34]

✚ **Al Khaleej Corpus (2004)**<sup>1</sup>[66] Ce corpus a été réalisé pour des expériences sur thème de l'identification des sujets pour la langue arabe. Il a été extrait de milliers d'articles téléchargés à partir d'un journal en ligne. Le corpus se constitue de quatre catégories (Nouvelles Internationales 953, Nouvelles Locales 2398, Sports 1430, Economie 909) et contient plus de 5000 articles qui correspondent à près de 3 millions de mots. La ponctuation a été intentionnellement omise [76].

### **3. Résultats et discussion :**

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

- L'impact du corpus de domaine spécifique.
- Qualité du dataset.

#### **3.1. Les résultats des deux modèles de similarité sémantique :**

Dans le tableau 4 nous présentons les résultats obtenus pour les modèle SV et CM avec et sans pondération de fréquences sur l'espace sémantique cyber. Le meilleur résultat (CP=**68.31%**) étant obtenu avec le modèle SV sans pondération pour une dimension 17225. « **Voir le tableau 4** »

---

<sup>1</sup> <https://sourceforge.net/projects/arabiccorpus/files/khaleej-2004corpus/>

**Tableau 4 : Résultats du modèle SV et CM avec l'espace sémantique Cyber**

				Corpus cyber			
				SUM_VEC Sans Pond	SUM_VEC Avec Pond	CM Sans Pond	CM Avec Pond
Khoja	Dimension des vecteurs	17225	CP(%)	68.31	67.42	62,83	62,42
			RMSE	1.06	1.08	1.44	1.33
Tashapyne		23715	CP(%)	68.18	67,33	62.53	62.18
			RMSE	1.05	1.07	1.64	1.64

Dans le tableau 5 nous présentons les résultats obtenus pour les modèle SV et CM avec et sans pondération de fréquences sur l'espace sémantique khaleej. Le meilleur résultat (CP=67.62%) étant obtenu avec le modèle SV avec pondération pour une dimension 14908. « Voir le tableau 5 »

**Tableau 5 : Résultats du modèle SV et CM avec l'espace sémantique Khaleej**

				Corpus khaleej			
				SUM_VEC Sans Pond	SUM_VEC Avec Pond	CM Sans Pond	CM Avec Pond
Khoja	Dimension des vecteurs	14908	CP(%)	67.01	67.62	63.23	62.16
			RMSE	1.07	1,07	1.33	1,35
Tashaphyne		18630	CP(%)	65.79	66.06	63.16	62.18
			RMSE	1.14	1,17	1,59	1,64

Dans le tableau 6 nous présentons les résultats obtenus pour les modèle SV et CM avec et sans pondération de fréquences sur l'espace sémantique CNN. Le meilleur résultat (CP=67.60%) étant obtenu avec le modèle SV avec pondération pour une dimension 16752. « Voir le tableau 6 »

**Tableau 6 : Résultats du modèle SV et CM avec l'espace sémantique CNN**

				Corpus cnn			
				SUM_VEC Sans Pond	SUM_VEC Avec Pond	CM Sans Pond	CM Avec Pond
Khoja	Dimension des vecteurs	16752	CP(%)	66.18	67.60	62.93	62.16
			RMSE	1.08	1.09	1.46	1.35
Tashaphyne		21032	CP(%)	65.82	66.13	63.09	62.18
			RMSE	1.13	1.15	1.6	1.64

Dans le tableau 7 nous présentons les résultats obtenus pour les modèle SV et CM avec et sans pondération de fréquences sur l'espace sémantique cyber. Le meilleur résultat (CP=68.20%) étant obtenu avec le modèle SV avec pondération pour une dimension 24230. « Voir le tableau 7 »

**Tableau 7 : Résultats du modèle SV et CM avec l'espace sémantique BBC+CNN**

				Corpus bbc+cnn			
				SUM_VC Sans Pond	SUM_VC Avec Pond	CM Sans Pond	CM Avec Pond
Khoja	Dimension des vecteurs	24230	CP(%)	66.36	68.20	62.71	62.16
			RMSE	1.09	1.05	1.33	1.35
Tashaphyne		28062	CP(%)	66.02	65.87	63.20	62.19
			RMSE	1.12	1.2	1.57	1.64

Nous constatons Pour les 4 espaces sémantiques même si le meilleur résultat est obtenu pour l'espace cyber, les résultats pour les autres restent très proches et comparables.

➤ **Combinaison des modèles (CM et SV) : « corpus cyber »**

Comme le montre le « Tableau 8 », l'hybridation des deux modèles de similarité a permis d'améliorer nettement les résultats obtenus séparément et ceci pour notre corpus cyber . Remarquons aussi que la combinaison par la moyenne donne un meilleur résultat dans ce cas. « Voir le tableau 8 »

**Tableau 8: Résultats de la combinaison des modèles CM et SV**

		CP(%)	RMSE
Stemmer Khoja	Moy	68.67	1.11
	Max	68.31	1.06
Stemmer Tashaphyne	Moy	67.61	1.17
	Max	68.34	1.05

### **3.2. Les résultats des similarités syntaxiques :**

Nous avons testé toutes les similarités syntaxiques et puis procéder à des combinaisons, pour pouvoir extraire la meilleure corrélation de Pearson et erreur quadratique.

Le classifieur Kmeans avec  $K=11$  a donné en général meilleur résultat c'est pour cette raison que nous ne présentons que le calcul de score utilisant Kmeans.

Dice a été notre meilleure approche avec une corrélation de 69,562% en utilisant un stem Léger, le fonctionnement de Dice qui se base sur les mots communs, le cas des réponses des étudiants avec réponses de l'enseignant ou le vocabulaire est presque le même d'où vient cette valeur importante de Pearson. « **Voir le tableau 9** »

Catégories	Algorithmes	Sans Stemming		Stemmer Khoja		Light Stemmer	
		CP(%)	RMSE	CP(%)	RMSE	CP(%)	RMSE
Similarités syntaxiques	Cosine	65,649	1,765	65,911	1,323	65,02	1,54
	STS	63,547	1,5	64,246	1,408	66,374	1,429
	Dice	69,101	1,179	67,245	1,249	69,562	1,18
	LCS	53,162	1,42	55,091	1,478	57,889	1,354
	Needlman	53,337	1,615	53,294	1,585	56,66	1,469
	Bgram	33,893	2,417	40,213	2,059	45,34	2,133
	Trigram	34,294	2,828	38,846	1,408	66,374	1,429
	Jaro	54,12	1,228	58,768	1,251	58,303	1,226
	NLCS	53,49	1,81	55,488	1,79	53,231	1,88
	Jaccard	50,454	1,271	54,169	1,231	54,932	1,233
	Euclidienne	23,842	1,508	42,764	1,382	36,93	1,406
	Damerau-levenshtein	50,013	1,547	52,956	1,612	52,658	1,624

**Tableau 9 : Résultats des similarités syntaxiques**

#### ➤ **Les combinaisons entre approches syntaxiques**

Nous allons entamer des combinaisons entre nos approches syntaxiques afin d'apporter des améliorations au niveau de la corrélation Pearson et l'erreur quadratique. Pour se faire nous avons appliqué des combinaisons avec une similarité moyenne, ainsi que similarité maximale :

### Similarité moyenne :

Cette méthode de combinaison non supervisée calcule simplement une moyenne de K scores de similarité par paires :

$$S_{cmb} = \frac{1}{k} \sum_{k=1}^k S_k \leftrightarrow s_{ij}^{cmb} = \frac{1}{k} \sum_{k=1}^k s_{ij}^k$$

### Similarité maximale :

Cette méthode de combinaison non supervisée calcule un maximum de K similarités par paires:

$$s_{ij}^{cmb} = \max(s_{ij}^1, \dots, s_{ij}^k)$$

#### a. Combinaison STS, Dice

Dans le tableau 10 nous retrouvons les résultats de l'application de la combinaison STS et Dice. la combinaison entre les meilleurs algorithmes a amélioré la corrélation de Pearson elle est devenue **69,011**.

#### b. Combinaison Dice, Cosine

Dans le tableau 10 nous retrouvons les résultats de l'application de la combinaison Dice, Cosine, la combinaison a amélioré la corrélation de Pearson elle est devenue 69,106%.

**Tableau 10 : Résultats de la combinaison de la similarité syntaxique**

Catégories	Algorithmes	Sans Stemming		Stemmer Khoja		Light Stemmer	
		CP(%)	RMSE	CP(%)	RMSE	CP(%)	RMSE
Combinaison des algorithmes syntaxiques	STS (wf:0), Dice(Max)	67,139	1,2	68,496	1,171	69,011	1,174
	STS(wf:0,1), Dice(Max)	66,402	1,279	68,397	1,87	68,81	1,201
	STS(wf:0,2), Dice(Max)	65,655	1,182	67,809	1,212	66,921	62,521
	DiceCosine	68,724	1,44	68,563	1,313	69,106	1,486

### Hybridation des modèles SV et CM sémantique avec la mesure syntaxique (Dice) :

La combinaison du modèle CM sémantique et la mesures syntaxique Dice a donné le meilleure résultat (CP=68.65%) avec le stemmer Tashaphyne et avec WF=0.0. Les résultats sont reportés dans le « **Tableau 11** ».

**Tableau 11 : Résultats de la combinaison du modèle CM avec la mesure syntaxique (Dice)**

	(CM sémantique, Dice) WF=0.0		(CM sémantique, Dice) WF=0.1	
	CP(%)	RMSE	CP(%)	RMSE
<b>Stemmer Khoja</b>	68.21	1.26	66.20	1.38
<b>Stemmer Tashaphyne</b>	<b>68.65</b>	1.26	68.56	1.28

La combinaison du modèle SV avec la mesure syntaxique a donné le meilleur résultat (CP=70.37%) avec le stemmer Tashaphyne. Les résultats sont reportés dans le « **Tableau 12**» en pondérant avec les TFmin-max.

**Tableau 12 : Résultats de la combinaison des modèles SV et Dice syntaxique**

	(SV sémantique, Dice)			
	Sans Pondération		Avec Pondération	
	CP(%)	RMSE	CP(%)	RMSE
<b>Stemmer Khoja</b>	70.13	1.02	69.68	1.03
<b>Stemmer Tashaphyne</b>	<b>70.37</b>	1.04	69.56	1.05

### Exemple d'attribution des notes avec les deux approches SV et CM :

Comme le montrent les figures 29 et 30, la corrélation entre les algorithmes appliqués et la notation manuelle (la moyenne de deux annotateurs) sont présentés pour la même question. Les degrés vont de 0 à 5 et le nombre des étudiants dix. Les chiffres montrent que SV peut être utilisé comme l'algorithme le plus proche de la notation manuelle dans le cas sémantique « voir figure 29 », alors que dans le cas syntaxique c'est Dice. « Voir la figure 30 »

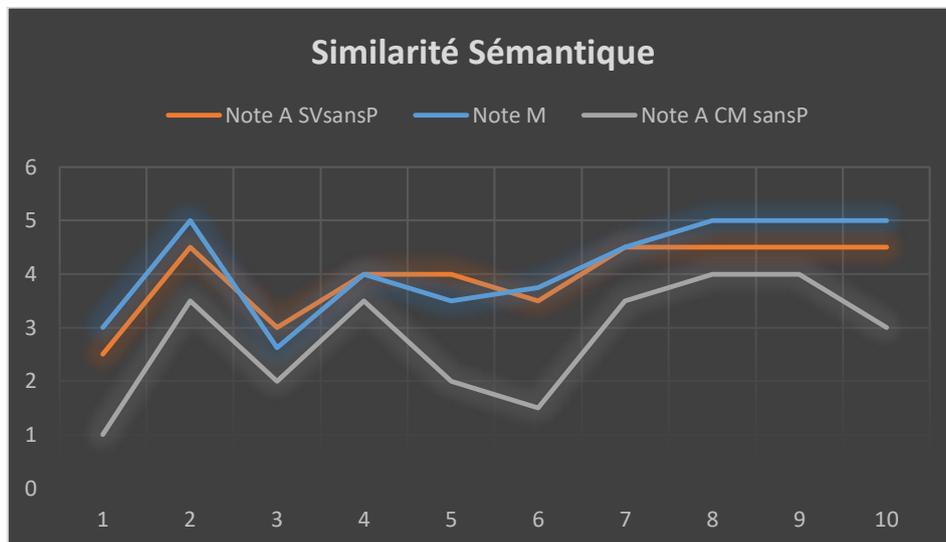


Figure 29 : Les notes attribuées par l'enseignant, SV, CM : pour la réponse de 10 élèves à la question " عرف مصطلح الجريمة الإلكترونية "

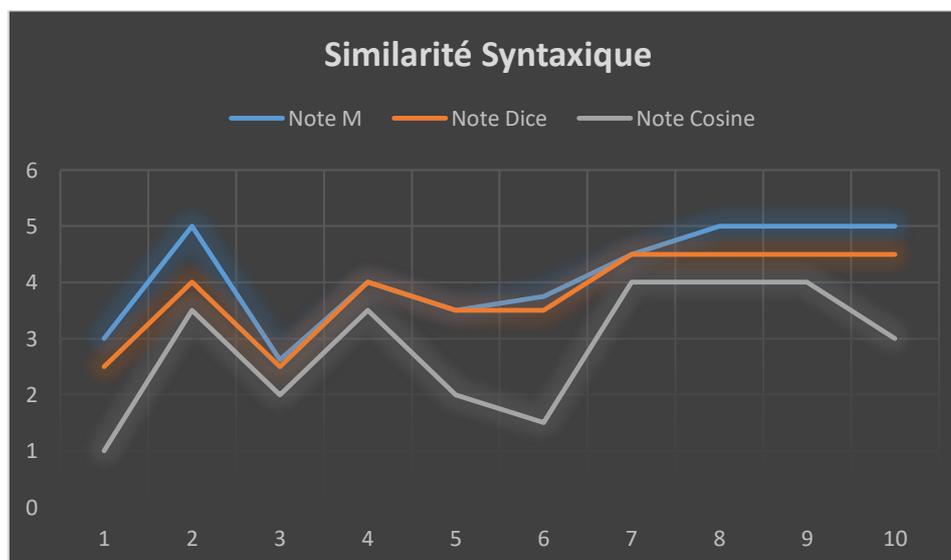


Figure 30 : Les notes attribuées par l'enseignant, DICE, COSINE : pour la réponse de 10 élèves à la question " عرف مصطلح الجريمة الإلكترونية "

c. Les résultats en utilisant les WE :

A la place de l'espace sémantique, nous avons utilisé une deuxième représentation des mots, celles utilisant les WE. Plusieurs approches ont été déjà explorées dans un travail précédent [65]. Nous profitons de l'acquisition des WE de Zahran [84] et de AraVec [85] pour évaluer le dataset, les seuls disponibles pour la langue arabe. Nous avons utilisé les WE de Zahran dans les modèles : CBOW (Continuous Bag Of Word) [82] et SkipGram [83].

**CBOW ou SKIPGRAM ?**

En général SkipGram présente de meilleurs résultats particulièrement dans les approches qui calculent la similarité basée sur la somme des vecteurs. On peut expliquer ceci par la nature même du modèle qui à partir d'un mot génère le contexte et donc la sommation des vecteurs de mots capture mieux les contextes.

Les tableaux qui suivent reportent les résultats de SOMVEC en utilisant les WE de **Zahran** et **ARavec**. Le meilleur résultat (CP=67.37%) est obtenu avec **Zhran SkipGram** et avec **Khoja stemmer** :

**Tableau 13 : Résultats des WE "Aravec CBOW Kmeans"**

	SomVect NonPond	SomVect Idf	SomVect Tfminmax	SomVect Tflog	SomVect TFIDF
<b>Sans Stem</b>	CP: 50,66 EQ: 1,34	CP: 50,43 EQ: 1,37	CP: 49,47 EQ: 1,37	CP: 45,29 EQ: 1,38	CP: 50,31 EQ: 1,37
<b>Khoja</b>	CP: 57,54 EQ: 1,25	CP: 58,04 EQ: 1,47	CP: 57,44 EQ: 1,24	CP: 57,44 EQ: 1,24	CP: 47,22 EQ: 1,45
<b>Tashaphyne</b>	CP: 56,04 EQ: 1,36	CP: 53,90 EQ: 1,49	CP: 56,98 EQ: 1,41	CP: 57,02 EQ: 1,41	CP: 49,35 EQ: 1,54

**Tableau 14 : Résultats des WE "Aravec SkipGram Kmeans"**

	SomVect NonPond	SomVect Idf	SomVect Tfminmax	SomVect Tflog	SomVect TFIDF
<b>Sans Stem</b>	CP: 49,76 EQ: 1,36	CP: 48,63 EQ: 1,34	CP: 49,17 EQ: 1,36	CP: 48,61 EQ: 1,34	CP: 47,27 EQ: 1,36
<b>Khoja</b>	CP: 63,44 EQ: 1,18	CP: 55,04 EQ: 1,23	CP: 62,46 EQ: 1,13	CP: 62,46 EQ: 1,13	CP: 51,13 EQ: 1,29
<b>Tashaphyne</b>	CP: 64,25 EQ: 1,09	CP: 61,75 EQ: 1,13	CP: 63,56 EQ: 1,13	CP: 63,07 EQ: 1,11	CP: 59,22 EQ: 1,17

**Tableau 15 : Résultats des WE "Zahran CBOW Kmeans"**

	SomVect NonPond	SomVect Idf	SomVect Tfminmax	SomVect Tflog	SomVect TFIDF
Sans Stem	CP: 41,81 EQ: 1,47	CP: 42,29 EQ: 1,42	CP: 38,36 EQ: 1,52	CP: 45,82 EQ: 1,39	CP: 51,96 EQ: 1,31
Khoja	CP: 65,54 EQ: 1,13	CP: 57,24 EQ: 1,23	CP: 64,33 EQ: 1,13	CP: 64,25 EQ: 1,13	CP: 53,92 EQ: 1,34
Tashaphyne	CP: 65,11 EQ: 1,13	CP: 59,36 EQ: 1,18	CP: 64,75 EQ: 1,14	CP: 65,25 EQ: 1,13	CP: 57,62 EQ: 1,21

**Tableau 16 : Résultats des WE "Zahran SkipGram Kmeans"**

	SomVect NonPond	SomVect Idf	SomVect Tfminmax	SomVect Tflog	SomVect TFIDF
Sans Stem	CP: 37,52 EQ: 1,60	CP: 45,96 EQ: 1,48	CP: 38,07 EQ: 1,59	CP: 41,69 EQ: 1,56	CP: 45,27 EQ: 1,39
Khoja	CP: 62,81 EQ: 1,30	CP: 67,37 EQ: 1,20	CP: 62,67 EQ: 1,18	CP: 62,63 EQ: 1,18	CP: 52,91 EQ: 1,28
Tashaphyne	CP: 63,48 EQ: 1,22	CP: 57,54 EQ: 1,25	CP: 62,04 EQ: 1,25	CP: 62,26 EQ: 1,25	CP: 54,31 EQ: 1,29

### Hybridation des modèles SV et CM sémantique avec le modèle WE :

Nous avons combiné nos résultats (des modèles CM et SV) avec ceux obtenus pour les modèles utilisant les WE (les modèles CBOW et SkipGram). Nous ne présentons dans le « Tableau 17. » que celles ayant apporté une amélioration significative des résultats.

**Tableau 17 : Résultats des combinaisons WE**

Modèle	DS (CP//RMSE)
CM Sémantique + Best WE	63.2 // 1.30
SV+Best WE	68.3 // 1.03
(CM Sémantique+SV)+Best WE	68.72 / 1.09

#### **4. Récapitulation des résultats et discussion**

Au terme de cette synthèse expérimentale, nous aboutissons à plusieurs constatations que nous discutons à travers les points suivants :

1) Une amélioration des résultats est bien marquée avec les combinaisons (interne ou externe). Ce qui confirme bien que les modèles hybrides donnent de meilleurs résultats (Le best de nos résultats est obtenue avec les combinaisons externes).

2) Le meilleur résultat avec le DS de cyber est obtenu avec la combinaison du modèle SV et la méthode Dice (70.37%). Ceci est prévisible puisque il y a plus de chance dans ce DS de trouver plusieurs mots en commun dans la réponse de l'étudiant et celle de la réponse modèle.

3) D'un autre côté, le stem lourd a été très efficace dans la similarité sémantique alors que le stem léger a été efficace avec la similarité syntaxique.

Le « Tableau 18 » résume les meilleurs résultats obtenus pour les différentes combinaisons :

**Tableau 18 : Récapitulatif des résultats**

Les meilleurs résultats avec DS Cyber	
Modèle	DS (CP // RMSE)
<b>Combinaison Sémantique</b>	
SV	68.31 // 1.06
CM	62.83 // 1.44
SV + CM (Moy)	68.67 // 1.11
SV + CM (Max)	68.34 // 1.05
<b>Combinaison Syntaxique</b>	
Dice	69.56 // 1.18
CM + Dice	68.65 // 1.26
SV + Dice	70.37 // 1.04
<b>Combinaison WE</b>	
SV + Best WE	63.2 // 1.30
CM + Best WE	68.3 // 1.03
Comb(SV+CM) + Best WE	68.72 // 1.09

**Examen approfondi :**

**a- Entre la note automatique et la note manuelle :**

L'analyse approfondie pour le dataset Cyber (CCASAG) en considérant les moyennes des deux scores attribués par les 2 annotateurs (moyennes des deux scores) et les scores manuelle attribué automatiquement.

**Tableau 19 : Analyse approfondie entre la note automatique et la note manuel pour le DS Cyber**

Score manuel Et score automatique	Nombre	Pourcentage (%)
Egaux	252	11.81 %
Inférieur ou égale à 1 (<=1)	1252	58.69 %
Inférieur ou égale à 2 (<=2)	475	22.26 %
Inférieur ou égale à 3 (<=3)	130	6.09 %
Supérieur à 3 (>3)	24	1.12 %

Si on veut analyser le tableau nous constatons que les résultats sont très satisfaisants puisque :

- 70.5 % des réponses présentent un écart d'au maximum de 1 point ce qui est intéressant.
- 92.76 % des réponses présentent un écart d'au maximum de 2 points. Ce qui est très raisonnable.

C'est une interprétation plus approfondie que le coefficient de Pearson qui nous permet d'examiner au détail les écarts entre l'être humain et la machine et elle confirme bien la qualité du dataset cyber et de l'approche d'évaluation. Ce sont des écarts comparables aux écarts humains entre annotateurs.

**b- Entre les 2 scores manuels :**

**Tableau 20 : Analyse approfondie entre les deux scores manuels pour le DS Cyber**

Score manuel 1 Et score manuel 2	Nombre	Pourcentage (%)
Egaux	743	34.83 %
Inférieur ou égale à 1 (<=1)	1155	54.14 %
Inférieur ou égale à 2 (<=2)	189	8.86 %
Inférieur ou égale à 3 (<=3)	38	1.78 %
Supérieur à 3 (>3)	8	0.37 %

### Comparer le comportement humain avec celui de la machine :

Si on compare les deux lignes concernant l'écart de points d'au maximum un point : 58.69 % et 54.14 %. C'est pratiquement la même chose.

Les écarts entre machine sont proches que ceux des humains et ce qui explique la qualité de l'approche malgré la subjectivité du processus d'évaluation lui-même.

## **5. Ressources matérielles et logicielles utilisées lors du développement**

Nous avons rencontré un problème matériel dans la phase principale de notre approche « Traitement du corpus du texte ». Ce dernier n'a pas abouti sa fin à cause de l'insuffisance de la mémoire de nos PC portables (8 Go de RAM). Pour cela, nous avons eu la chance de travailler sur un serveur à distance « Voir Figure 31 » fournit par l'université de Bouira et c'est la configuration minimale requise pour notre travail.

### **HP 470065-652 ProLiant ML350p Gen8 Intel Xeon**

<b>Processeur:</b>	E5-2620 / 2 GHz
<b>RAM:</b>	16 Go
<b>OS:</b>	Linux

**Figure 31 : Références du serveur à distance**

Nous travaillons avec le langage de programmation **Python 3.6** sur nos machines personnelles et le serveur. Pour les interfaces, planifiées en java, nous avons utilisé l'environnement de développement **Netbeans 8.2** et **PyCharm Community 2019.1.3** pour Python.

**A propos notre corpus de texte de domaine « Cybercriminalité » :**

**Tableau 21 : Des informations sur le corpus Cyber**

Corpus Cyber		
<b>Taille</b>		20.4 Mo
<b>Nombre de fichier</b>		1273 fichiers (.txt)
<b>Langage</b>		Arabe
<b>Nombre total de mots</b>	<b>Khoja</b>	1342677
	<b>Tashaphyne</b>	1227656
<b>Nombre de mot unique</b>	<b>Khoja</b>	17225
	<b>Tashaphyne</b>	23715

## **Conclusion :**

Au terme de cette synthèse expérimentale, nous aboutissons à plusieurs constatations que nous discutons à travers les points suivants :

-D'après l'analyse approfondie entre la note automatique et la note manuel pour le DS Cyber :

- 70.5 % des réponses présentent un écart d'au maximum de 1 point ce qui est intéressant.
- 92.76 % des réponses présentent un écart d'au maximum de 2 points. Ce qui est très raisonnable.

-Les combinaisons des approches sémantiques syntaxiques et WE ont amélioré beaucoup les résultats :

On a pu atteindre une corrélation de Pearson de 70.37% pour notre dataset cyber et notre corpus cyber qui est le meilleur résultat pour la combinaison de similarité syntaxique avec CM sémantique, alors que avec la combinaison similarité sémantique (SV+CM) était 68.67% et 68.72% pour WE avec la combinaison des modèles sémantique (SV+CM).

# Conclusion et perspectives :

L'évaluation joue un rôle important dans l'enseignement et à des effets importants sur l'apprentissage. Néanmoins, cette tâche présente une charge contraignante pour les enseignants en termes de temps, de concentration et de précision.

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 ainsi qu'une analyse sur les méthodes utiliser pour la création de corpus de texte à partir le web. L'analyse faite nous a permis de mettre l'accent sur les enjeux de l'utilisation de la langue arabe dans le domaine de l'évaluation automatique et des approches de similarité.

Nous avons ensuite présenté notre système développé qui intègre : un outil de manipulation de dataset (le corpus de test) qui est basé sur un enseignement pédagogique sur un contenu de programme diffusé en langue arabe suivi par un test utilisant des questions aux réponses courtes. Pour une variété d'utilisation, le data set est élaboré dans les cinq versions : Texte, XML et base de données. Plus un outil de création automatique de corpus de domaine, à partir du web, par interrogation automatique le moteur de recherche et nous avons intégré avec cet outil l'outil de création d'espace sémantique. Et enfin un outil d'évaluation pour évaluer le dataset (le corpus de test) par rapport au corpus de texte qui ont été construit de même domaine pour étudier l'impact du domaine commun sur les résultats obtenus.

Nous avons réussi au meilleur résultat (**CP=70.37%**) en utilisant la similarité hybridée par la combinaison des modèles SV et Dice syntaxique », alors que dans la similarité sémantique seule a donné un meilleur résultat (**CP=68.67%**) « De la Combinaison des modèles (CM et SV) », et avec les WE en obtenus le meilleur résultat (**CP=67.37%**) « Zahran SkipGram ».

En perspectives, nous proposons de considérer entre autres les aspects suivants :

- Nous aurons aimé de faire la consolidation des différents outils développés sur une plateforme en ligne accessible via une interface web.
- Enrichir l'évaluation à la prise en compte de la spécification des remarques de l'enseignant et du retour de feedback à l'apprenant permettant de lui spécifier les parties de sa réponse qui nécessitent plus d'investissements de sa part.
- Ajouter une étape de correction grammaticale automatique au début de l'étape du traitement du texte afin de corriger les erreurs de saisies et par la suite éviter les ambiguïtés.

# Bibliographie

- [1] S. Burrows, I. Gurevych, et B. Stein, The eras and trends of automatic short answer grading, vol. 25, no 1. 2015.
- [2] D. Callear, J. Jerrams-Smith, V. Soh, D. J. Jerrams-smith, et H. P. Ae, « CAA of Short Non-MCQ Answers », in In Proceedings of the 5th International CAA conference, 2001.
- [3] C. Leacock et M. Chodorow, « C-rater: Automated Scoring of Short-Answer Questions », Comput. Hum., vol. 37, no 4, p. 389-405, 2003.
- [4] S. Jordan, « Investigating the Use of Short Free Text Questions in Online Assessment », Final Proj. report, Cent. Open Learn. Math. Sci. Comput. Technol. Open Univ. Milt. Keynes, United Kingdom, 2009.
- [5] J. Sukkarieh et J. Blackmore, « c-rater: Automatic Content Scoring for Short Constructed Responses. », FLAIRS Conf., p. 290-295, 2009.
- [6] J. Z. Sukkarieh et S. Stoyanchev, « Automating Model Building in C-rater », in Proceedings of the 2009 Workshop on Applied Textual Inference, p. 61-69, 2009.
- [7] Sukkarieh, J. Z., Pulman, S. G., and Raikes, N. (2003). Auto-marking: Using Computational Linguistics to Score Short, Free Text Responses. In Proceedings of the Twenty-Ninth Annual Conference of the International Association for Educational Assessment, pages 1–15, Manchester, United Kingdom.
- [8] Sukkarieh, J. Z., Pulman, S. G., and Raikes, N. (2004). Auto-marking 2: An update on the UCLES-Oxford University Research into using Computational Linguistics to Score Short, Free Text Responses. In Proceedings of the Thirtieth Annual Conference of the International Association for Educational Assessment, Philadelphia, Pennsylvania.
- [9] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, et I. H. Witten, « The WEKA Data Mining Software: An Update », SIGKDD Explor. Newsl., vol. 11, no 1, p. 10-18, 2009.
- [10] Bloom taxonomie : A taxonomy for learning, teaching, and assessing : a revision of Bloom's taxonomy of educational objectives. New York : Longman, c2001
- [11] Bouleknadel et al. 2008S. Bouleknadel, B.Daille et D. Aboutajdine. A multi-word term extraction program for Arabic language. In the 6th international Conference on language resources and evaluation LREC, pp. 1485-1488, 2008.
- [12] Grefenstette, G. 1994. Exploration in Automatic Thesaurus Discovery. Kluwer Academic Press, Boston.
- [13] Benzater Nebia ; « Analyse morphologique du Texte Arabe Pour son Indexation Sémantique », Thèse de Magistère, soutenue, en 2015.
- [14] [Blachère et Gaudefroy, 1975] R. Blachère et M. Gaudefroy-Demombynes, Grammaire de l'arabe classique (morphologie et syntaxe), G.P. Maisonneuve & Larose, Editeurs à Paris, 508 p. (1975).
- [15] Miller G. 1995. Wordnet: A lexical database. Actes de ACM 38, pp. 39-41.
- [16] Georges Leboss, Abd El Salam Al Hajjar, Mohammad Hajjar et Gilles Bernard. "Semantics information extraction of Arabic words :Methods and Resources". In : The 5th International Conference on Information Systems and Technologies (ICIST). Istanbul, Turkey, mar. 2015.
- [17] Pierre Guiraud. La sémantique. 655. Puf, 1975.
- [18] John Lyons. Semantics (vols I & II). Cambridge University Press, 1977.

- [19] James R Hurford, Brendan Heasley et Michael B Smith. *Semantics : a coursebook*. Cambridge University Press, 2007.
- [20] Pierre Lerat. *Sémantique descriptive*. Classiques Hachette, 1983.
- [21] hn I. Saeed. *Semantics*. Blackwell Publishing, 1997.
- [22] Charles W Kreidler. *Introducing english semantics*. Routledge, mai 1998.
- [23] Sebastian Lobner. *Understanding Semantics*. Arnold, 2002.
- [24] Elkateb, S., Black, W., Rodriguez, H., Alkhalifa, M., Vossen, P., Pease, A., and Fellbaum, C., (2006). *Building a WordNet for Arabic*. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation*, Genoa, Italy.
- [25] Elkateb, S. "Design and implementation of an English Arabic dictionary/editor", PhD thesis, Manchester University (2005).
- [26] Niles I., Pease A. 2003. *Linking Lexicons and Ontologies: Mapping WordNet to the Suggested Upper Merged Ontology*. Actes de 2003 International Conference on Information and Knowledge Engineering (IKE '03), Las Vegas, Nevada.
- [27] Wajdi Zaghouani. "Critical survey of the freely available Arabic corpora". In: *Proceeding of the Workshop on Free/Open-Source Arabic Corpora and Corpora Processing Tools*, LREC.2014, p. 1-8.
- [28] Mohler, M., & Mihalcea, R. (2009). *Text-to-text semantic similarity for automatic short answer grading*. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 567-575). Association for Computational Linguistics.
- [29] Mohler, M., Bunescu, R. C., & Mihalcea, R. (2011). *Learning to Grade Short Answer Questions using Semantic Similarity Measures and Dependency Graph Alignments*. In *ACL* (pp. 752-762).
- [30] Gomaa, W.H. & Fahmy, A. A. (2012). *Short Answer Grading Using String Similarity and Corpus-Based Similarity*. In *International Journal of Advanced Computer Science and Applications (IJACSA)*, Vol. 3, No.11.
- [31] Islam, A. and Inkpen, D. (2008). *Semantic text similarity using corpus-based word similarity and string similarity*. *ACM Trans. Knowl. Discov. Data* 2, 2 (Jul. 2008), 1–25
- [32] E. Negre, « Comparaison de textes: quelques approches... », 2013.
- [33] « Indice et distance de Jaccard », 2018. [En ligne]. Disponible sur: [https://fr.wikipedia.org/wiki/Indice\\_et\\_distance\\_de\\_Jaccard](https://fr.wikipedia.org/wiki/Indice_et_distance_de_Jaccard). [Consulté le: 17-juin-2018].
- [34] M. Saad et W. Ashour, « OSAC: Open Source Arabic Corpora », 6th Int. Conf. Electr. Comput. Syst. (EECS'10), Nov 25-26, 2010, Lefke, Cyprus., no May 2014, p. 118-123, 2010.
- [35] 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.
- [36] Mihalcea, R., Corley, C. & Strapparava, C. (2006). *Corpus based and knowledge-based measures of text semantic similarity*. In *Proceedings of the American Association for Artificial Intelligence*.(Boston, MA).
- [37] Top Ten Internet Languages - World Internet Statistics [www.internetworldstats.com/stats7.htm](http://www.internetworldstats.com/stats7.htm)
- [38] Attia, M., Pecina, P., Toral, A., Tounsi, L., & Van Genabith, J. *A lexical database for modern standard Arabic interoperable with a finite state morphological transducer*. In C. Mahlow, & M. Piotrowski (Eds.), *Proceeding of Second International Workshop, SFCM Systems and*

Frameworks for Computational Morphology, Zurich, Switzerland,, August 26, 2011 2011 (pp. 98-118): Springer

[39] Fathi Debili, Hadhemi Achour et Emna Souissi. De l'étiquetage grammatical à la voyellation automatique de l'arabe. In in Correspondances, volume vol 71, pages 10–26, IRMC : Institut de Recherche sur le Maghreb Contemporain, Tunis, 2002. (Cité en pages xiii, 49, 52, 53 et 58.)

[40] Lamia Hadrich Belguith, Leila Baccour et Mourad Ghassan. Segmentation de textes arabes basée sur l'analyse contextuelle des signes de ponctuations et de certaines particules. Actes de la 12ème conférence sur le Traitement Automatique des Langues Naturelles TALN'2005 - Dourdan France, vol. Vol. 1, pages 451–456, 2005. (Cité en pages xiii, 50, 51, 54, 55, 56, 82, 96, 99 et 122.)

[41] Lamia Hadrich Belguith, Chafik Aloulou et Abdelmajid Ben Hamadou. MAS-PAR : De la segmentation à l'analyse syntaxique de textes arabes. In CÉPADUÈS-Éditions, éditeur, Revue Information Interaction Intelligence I3, volume 7, pages 9 – 36, <http://www.revue-i3.org/>, mai 2008. 2008. ISSN : 1630-649x. (Cité en pages 51, 59 et 82.)

[42] Burstein, J., "The e-rater scoring engine: Automated Essay Scoring with natural language processing", In M. D. Shermis and J. C. Burstein (Eds.), Automated Essay Scoring: A cross disciplinary approach (pp. 113–121). Mahwah, NJ: Lawrence Erlbaum Associates, 2003.

[43] Mitchell, T., Russel, T., Broomhead, P., & Aldridge N., "Towards robust computerized marking of free-text responses", In M. Danson (Ed.), Proceedings of the Sixth International Computer Assisted Assessment Conference, Loughboroug University, Loughborouh, UK., 2002.

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

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

[46] Leacock C., Chodorow M. C-rater: Automated Scoring of Short- Answer Questions. Computers and Humanities 37, 2003.

[47] A. Shehab, M. Elhoseny, and A. E. Hassanien, "A hybrid scheme for Automated Essay Grading based on LVQ and NLP techniques," in 2016 12th International Computer Engineering Conference (ICENCO), 2016, pp. 65-70.

[48] K. M. Nahar and I. M. Alsmadi, —The automatic grading for online exams in Arabic with essay questions using statistical and computational linguistics techniques,|| MASAUM Journal of Computing, vol. 1, no. 2, 2009.

[49] Gomaa, Wael Hassan, and Aly Aly Fahmy. "Automatic scoring for answers to Arabic test questions." Computer Speech & Language 28.4 (2014): 833-857

[50] R. Mezher and N. Omar. || A Hybrid Method of Syntactic Feature and Latent Semantic Analysis for Automatic Arabic Essay Scoring. || Journal of Applied Sciences ,16(5): 209-215(2016).

[51] Gomaa, W.H. & Fahmy, A. A. (2012). Short Answer Grading Using String Similarity and Corpus-Based Similarity. In International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 3, No.11.

[52] Basu, S., Jacobs, C., & Vanderwende, L. (2013). Powergrading: a Clustering Approach to Amplify Human Effort for Short Answer Grading, In Transactions of the ACL (TACL).

[53] Gomaa, W.H. & Fahmy, A.A. (2011). Tapping Into The Power of Automatic Scoring. the eleventh International Conference on Language Engineering, Egyptian Society of Language Engineering (ESOLEC '2011).

[54] Ziai, R., Ott, N. & Meurers, D. (2012). Short answer assessment: establishing links between research strands. In: Proceedings of the Seventh Workshop on Building Educational Applications Using NLP, June. Association for Computational Linguistics, pp. 190–200.

- [55] Rosé, C. P., Roque, A., Bhembe, D., & VanLehn, K. (2003). A hybrid approach to content analysis for automatic essay grading. In Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology: companion volume of the Proceedings of HLT-NAACL 2003--short papers Volume 2 (pp. 88-90). Association for Computational Linguistics.
- [56] Rohde, D. L., Gonnerman, L. M., and Plaut, D. C. (2006). An improved model of semantic similarity based on lexical co-occurrence. *Communication of the ACM*, 8:627–633
- [57] Mitchell, T., Russell, T., Broomhead, P. & Aldridge, N. (2002). Towards robust computerized marking of freetext responses. In: Proceedings of the Sixth International Computer Assisted Assessment Conference. Loughborough University, Loughborough, UK.
- [58] Pulman, S.G. & Sukkariéh, J.Z. (2005). Automatic short answer marking. In: Proceedings of the Second Workshop on Building Educational Applications Using NLP, June. Association for Computational Linguistics, pp. 9–16.
- [59] Rashad, Magdi Z., et al. "An Arabic web-based exam management system." *International Journal*
- [60] Alghamdi, M., M. Alkanhal, M. Al-Badrashiny, A. Al-Qabbany, A. Areshey, A. Alharbi. A hybrid automatic scoring system for Arabic essays. *J. AI Commun.*, 27:103-111.
- [61] A. Abdallah et K. Garoudja. « Mesures de similarité syntaxique pour un système d'évaluation automatique des réponses courtes : Application à la langue arabe ». 2 Mémoire master USDB 1. 2018/2019
- [62] A. Benayad et Y. Atoub. « Mesures de similarité sémantique pour un système d'évaluation automatique des réponses courtes : Application à la langue arabe ». Mémoire master USDB 1. 2018/2019.
- [63] M. Mustafa, A. S. Eldeen, S. Bani-Ahmad, and A. O. Elfaki, "A Comparative Survey on Arabic Stemming: Approches and Challenges," *Intell. Inf. Manag.*, vol. 9, no. 02, p. 39, 2017.
- [64] Vergne J, (2004) « Découverte locale des mots vides dans des corpus bruts de langues inconnues, sans aucune ressource », 7eme Journées internationales d'analyse statistique des données textuelles, GREYC – University of Caen.
- [65] M. Hannoufi et H. Henniche. « Les Word-Embedding pour l'évaluation automatique des réponses courtes en apprentissage en ligne : Application à la langue arabe ». Mémoire master USDB 1. 2018/2019.
- [66] M. Abbas et K. Smaili, « Comparison of Topic Identification methods for Arabic Language », in *International Conference on Recent Advances in Natural Language Processing - RANLP 2005*, no 14-17, 2005.
- [67] Larkey et al. 2002 L. Larkey, L. Ballesteros, and M. Connell. Improving Stemming for Arabic Information Retrieval: Light Stemming and Co-occurrence Analysis, *Proc. of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 275 - 282, 2002.
- [68] I. Retrieval, *Introduction to Information Retrieval*. 2008.
- [69] S. Khoja and R. Garside, "Stemming arabic text," Lancaster, UK, Computing Department, Lancaster University, 1999.
- [70] I. A. Al-Sughaiyer and I. A. Al-Kharashi, "Arabic morphological analysis techniques: A comprehensive survey," *Journal of the American Society for Information Science and Technology*, vol. 55, no. 3, pp. 189–213, 2004.

- [71] L. S. Larkey and M. E. Connell, "Arabic information retrieval at UMass in TREC-10," DTIC Document, 2006.
- [72] K. Abainia, S. Ouamour, and H. Sayoud, "A novel robust Arabic light stemmer," *J. Exp. Theor. Artif. Intell.*, vol. 29, no. 3, pp. 557–573, 2017.
- [73] M. Mustafa, A. S. Eldeen, S. Bani-Ahmad, and A. O. Elfaki, "A Comparative Survey on Arabic Stemming: Approches and Challenges," *Intell. Inf. Manag.*, vol. 9, no. 02, p. 39, 2017.
- [74] F. Issac, T. Hamon, L. Bouchard, L. Emirkanian, C. Fouqueré, extraction informatique de données sur le web : une expérience, in *Multimédia, Internet et francophonie : à la recherche d'un dialogue*, Vancouver, Canada, mars 2001.
- [75] T. Zerrouki, "Tashaphyne, Arabic light Stemmer/segment." 2010.
- [76] M. Abbas, « Corpus arabe - Mourad Abbas ». [En ligne]. Disponible sur: <https://sites.google.com/site/mouradabbas9/corpora>. [Consulté le: 19-juin-2018].
- [77] C. Tambellini et C. Berrut, « Pondération des données incertaines dans les systèmes de recherche d'informations : une première approche expérimentale . », p. 247-261, 2006.
- [78] 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.
- [79] D. Callear, J. Jerrams-Smith, V. Soh, D. J. Jerrams-smith, et H. P. Ae, « CAA of Short Non-MCQ Answers », in *In Proceedings of the 5th International CAA conference*, 2001.
- [80] 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.
- [81] P. Resnik, *Parallel strands: A preliminary investigation into mining the web for bilingual text*, in *conference of the association for machine translation in the Americas*, 1998.
- [82] T. Mikolov, W. Yih, and G. Zweig, "linguistic Regularities in Continuous Space Word Representations," no. June, pp. 746-751, 2013.
- [83] T. Mikolov, K. Chen, G. Corrado, and J. Desn, "Distributed Representation of Words and Phrases and their Compositionality," pp. 1-9.
- [84] 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.
- [85] Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient Estimation of Word Representations in Vector Space. In: *ICLR: Proceeding of the International Conference on Learning Representations Workshop Track*, Arizona, USA, pp. 1301–3781 (2013).
- [86] GHOUL, Dhaou. Construction d'un corpus arabe à partir du Web dans le but d'identifier les mots-outils ou tokens. *JADT'14*, Paris, France, 2-6 juin 2014.

# Annexe

## A. Liste des stop-words :

هَلَا	كان	امسى	عدة	س	ساء	أضحى	ثم	يوليو	والتي	لَسُنْ	بشكل
حاء	بدلا	مازال	جوان	ألقى	أولئك	اما	برس	خمسين	وقد	سنوات	اثنين
سنتيم	التي	لكن	بعد	كأين	شمال	اربعون	كأنما	الاولى	ماذا	تسعمئة	شبن
كيت	كي	ثم	هكذا	أبريل	مهما	خاء	اللتين	مثل	ولم	حي	أي
جيم	صه	تحت	إلا	أمامك	شخصا	أنى	قد	طَق	هؤلاء	راء	لدي
نا	أهلا	من	دولار	إليكم	أنتم	أبو	ظل	لَيْسَا	ذواتي	مايو	علم
نوفمبر	اعادة	هل	سبع	انبرى	نفسه	لايزال	عدا	لَسُنْتُ	كم	مقابل	أَيَّان
عيانا	شَتَانْ	سبعين	فاء	أوشك	ثمانين	ساءما	بس	لكنه	بل	ذ	الذين
يومنذ	هَاتِي	ما	إذ	معاذ	عشر	لات	مليون	هَيْهَاتَ	أخ	عمّا	ك
إلي	ذَيْن	مليار	سيما	كلم	بهما	أَيْتَ	إحدى	عندما	م	خمس	حمو
لازال	عن	خمسائة	لما	وهذا	د	هَذِي	صبرا	كيفما	ستكون	اذ	أل
ثان	أعطى	سنة	قاطبة	إليك	شيكل	ان	خامس	وهي	لن	أقل	إن
بنا	هَاتَانْ	تلکم	كاف	تسعمائة	صار	بان	فانه	همزة	كى	لعل	حاليا
والذي	انقلب	هناك	بهم	صفر	ثلاثمئة	فى	عنه	يناير	لَسُنْتُ	الي	مانفك
فيم	فلان	وليس	حول	فوق	أربعمئة	أها	ميم	تاء	ثمانون	حيثما	لوما
أغسطس	إزاء	نَبَا	ترك	واضاف	ين	كانت	الاخيرة	إذا	كل	آناء	بات
لهم	أمد	الان	منذ	بمن	مذ	سين	هلم	ة	بلى	وحتى	لام
عل	هذا	ومنذ	أحد	عنها	ساء	عام	عليك	علق	ن	بطان	هَاتِه
الثانية	سرا	غير	هيئة	ياء	ويكأن	نعم	مه	عشرة	فكان	فرادى	هَذَيْن
خال	فمنهم	ليت	إذا	طاء	فبراير	وجد	نحن	أولاء	الف	إمّا	باسم
مكائنك	واحد	ذينك	تم	ي	نفس	وعن	واوضح	وكانت	أقبل	غين	سبعون
أب	لم	بي	و	بك	فيفري	تلك	ب	من	منها	ث	الماضي
لَسُنْتُ	أجمع	مايزال	ز	بَلَه	وا	ليرة	أف	مافتى	ذلك	أفريل	رُب
ذيت	احد	أصلا	اتخذ	لكن	نيف	أجل	إى	ها	هؤلاء	اليها	واها
خاصة	بما	تينك	أن	أمس	ع	اكذ	وفي	هن	خلال	برح	استحال
اثر	لو	سرعان	هذا	ى	بكم	سوى	فيما	أوة	صراحة	خلا	حزيران
جميع	أمام	لك	هَاتَيْن	مائة	لنا	أين	ستين	ذِي	ق	مكانك	لعل
خمسون	تموز	عاشر	ثمنمئة	ستون	هي	فلس	نَحْ	تسعة	حوالى	كأَي	إليه
ولا	سبعمئة	تعسا	عاد	بد	ايار	ساءمّا	ثم	حسب	ل	أه	أن
ذواتا	بن	فمنذ	جدا	جويلية	لهما	اللواتي	اللواتي	لولا	يوم	قط	ريث
ثلاثون	خَبْر	فيها	يمكن	رابع	سبعة	إنّما	كيف	هنالك	كأن	علي	حبيب
ايام	عدم	هيا	ظن	اثنان	أن	هم	جعل	شباط	معه	ر	أربعة
إن	ابتدا	الا	قلما	طفق	حار	اثنى	أب	إيانا	وأبو	حادي	لكم
وهذه	الألاء	ض	خَذَار	اخلوق	لئن	لبيك	رُب	ء	عل	أصبح	حمدا
أنفا	أفعله	اجل	ذال	وأن	أض	زعم	بعض	لأن	عسى	الا	او

أيلول	عدد	الاول	وَي	خلف	ولكن	ضحوة	وهب	رويدك	اى؟	ح	أيا
تان	وُشْكَانَ	ثالث	نِعْمًا	في	ظَنَّ	لوكالة	اللذان	بهذا	ستة	منه	خ
إياه	حَبْدًا	ديسمبر	لَعَلَّ	وتلك	نيسان	تلقاء	إياهن	إلَيْكَ	ذا	ارتدَّ	جمعة
عين	ابين	بكما	أَيَّ	أَمِينٍ	مكانكما	ذانك	أب	ضمن	دون	له	لقاء
ن	اللتى	ثُمَّ	اف	غادر	لَيْسُوا	واو	ألا	اليه	قوة	أَنْتِ	عليه
أَيَّان	كلتا	حَمَّ	ذان	وكان	كَلَّا	ثلاثين	إِنَّ	دال	ثلاثمائة	خلفا	أيار
صباح	باء	كسا	ذه	ظاء	وُشْكَانَ	لي	أَفِ	دواليك	تَيْنِ	خمسمة	لَسْتُمْ
لَكُنْمَا	ريال	قاف	الذين	ثلاث	أمامك	ثمان	فمنها	هذه	كَانَ	درهم	شَتَّانَ
ستمئة	حيَّ	إما	حم	أي	ظ	اخرى	جانفي	لكي	لاسيما	تخذ	يفعلون
بئس	اصبح	سبحان	ليس	ست	عما	إِنَّ	حاشا	اضحى	تسعين	إياكن	كذا
أنت	اليوم	يورو	هَيَّهَات	فهو	بَسَّ	جير	صدقا	كليهما	كلا	سبت	دينار
إليكما	لكنَّ	هلم	ج	ف	جلل	يونيو	كذلك	طرا	راح	كلَّما	إياها
حَتَّى	وراءك	سبتمبر	لهذا	وبين	الحالي	فإن	سادس	بخ	أه	سابع	أَفِ
أنتن	كلها	هذه	صه	كما	حرى	ثلاثاء	الثاني	كلَّا	وكذلك	تفعلان	ثمَّ
ذو	لَيْسَ	انفك	أو	تعلم	اربعة	إلى	جنيه	مَنْ	غداة	خميس	عليها
غدا	أولئكم	ثلاثة	زاي	ذلكن	واضافت	و	متى	مكانكم	عَدَسُ	الذي	ص
تكون	هاؤم	أنا	اثنا	تلكما	هما	أل	بها	هيَّا	تي	حتى	و6
ضاد	لها	التي	وقالت	تاء	مليم	صاد	سمعا	لَيْسَتْ	كثيرا	كن	وثي
بعدا	لذن	إي	اكثر	نهاية	حقا	إياهم	تِه	لكما	ومع	للامم	نون
قال	إذما	وله	أخو	ألف	فان	ظَلَّ	اللتيا	عند	بَسَّ	ذلكما	هو
فقط	إياكما	أمسى	لَعَلَّ	غالبا	أ	أيضا	هَذَانِ	فو	تَانِكَ	رجع	الآن
بؤسا	تشرين	الى	لَيْسَتْ	مَرَّة	لقد	يوان	تحوّل	اول	بين	عشرون	أفعل
فضلا	إياك	كانون	ممن	ا	غ	لَسْتُمْ	عاما	أول	وعلى	ورد	أربع
تجاه	هنا	حبذا	يفعلان	أربعاء	فيه	انه	زيارة	إِذَا	لعمر	كأين	حاي
أنتما	لَكِنَّ	اعلنت	أكتوبر	طالما	درى	إيأي	ستمائة	يكون	كلاكما	حين	ذات
تفعلون	كرب	ومن	وهو	اربعين	ت	ثمانمة	أخ	ذلكم	اللاني	ايضا	الذى
يمين	ثاني	ثمَّ	مساء	الوقت	أَنَّ	هلالة	هَبَّ	أولالك	أه	اطار	حيث
بماذا	عدا	خمسة	كاد	،	أم	نعم	ط	ثمانية	ذاك	إلَيْكَ	هاء
هاهنا	وان	مابرح	بغثة	اللتى	ذوا	أينما	أخذ	ثماني	مع	امس	أى
الذاتي	بضع	دونك	إياهما	الذي	به	أما	لذلك	سوف	ولايزال	أكثر	لا
قرش	أربعمائة	تفعلين	تارة	لَسْتِ	مادام	شبه	طاق	الألى	شَتَّانَ	عوض	
إذن	بسبب	سحقا	لَسْنَا	كليكما	عجبا	بيد	مارس	غَدًا	مئة	مما	
عشرين	حجا	هَجَّ	فقد	ماي	سبعمائة	كخ	لهن	تسع	وقف	ثامن	
بهن	السابق	على	يلى	أبدا	لدى	أَفِ	إييه	قبل	ه	ئ	
اللاواتي	انها	لكيلا	أَيَّ	أنه	علَّ	أ	امام	حدث	بكن	ء	
مئتان	لما	أوت	نحو	بان	كانَّ	تاسع	مئتين	عامة	تسعون	إياكم	
أنشأ	أي	المقبل	أما	آذار	ضد	هيهات	ش	كلاهما	اللتان	هاك	