
Matrice racine-schème pour le verbe de la langue arabe*

Youssef TAHIR

Département d'information & de communication
Faculté des lettres et des sciences humaines Saïss
Université Sidi Mohamed Ben Abdallah Fès – Maroc

Ytahir1@hotmail.com

Abstract: *We are interested in the implementation of an exhaustive database for Arabic language verbs. This database is represented as a matrix that the header of the rows contains all Arabic verbs' patterns and the header of the columns contains a list of all recognized roots in Arabic literature. The exhaustiveness of this list, which is one of the most important characteristics of our database, is due to the fact that we started with the generation of all possible combinations of three characters in Arabic language alphabet, then we have removed those which are not recognized in the literature. The intersection of each row and column of the matrix contains a code which indicates if the root and the pattern are compatible, and refers to a list of morpho-syntactic characteristics of the generated verb*

Résumé : *Cet article présente une base de données pour le verbe de la langue arabe sous forme d'une matrice dont l'en-tête de lignes contient les schèmes de verbes et l'en-tête de colonnes contient une liste exhaustive de toute racines reconnus dans la littérature arabe. L'exhaustivité de cette liste, qui est le point de force de notre base de données, est due au fait que nous avons commencé par la génération de toutes les combinaisons possibles à partir de l'alphabet de la langue arabe, ensuite nous avons supprimé celles qui ne sont pas reconnues dans la littérature. L'intersection de chaque ligne et colonne contient un code indiquant si la racine et le schème sont compatibles et référant à une liste de caractéristiques morpho-syntaxiques du verbe engendré par cette compatibilité.*

Keywords: *Linguistic resources, lexical database, matrix root-pattern, morphosyntax*

Mots clés : *Ressources linguistiques, base de données lexicale, matrice racine-schème, morphologie syntaxe, sémantique.*

**A matrix root-pattern for Arabic verbs*

1 Introduction

Les systèmes de Traitement Automatique des Langues Naturels (TALN) ne cessent d'évoluer, la réussite d'un tel système dépend principalement de sa base de données linguistique qui représente son unique base de connaissance. A cet effet, le besoin du TALN en ressources lexicales ne cesse de s'intensifier.

Pour les langues latines, plusieurs ressources morphologiques, syntaxiques et sémantiques existent déjà et sont à une étape assez avancée. Cependant, le traitement automatique de la langue arabe souffre encore du manque et de la faiblesse de ces ressources linguistiques. En effet, la plupart des analyseurs morpho-syntaxiques de la langue arabe repose sur la représentation racine-schème mais leurs base de données ne comportent pas une liste exhaustive des racines ainsi que des indicateurs pour exprimer l'homogénéité des racines et des schèmes. Ce qui explique le taux d'erreurs rencontrées lors de l'utilisation de certains systèmes de traitement automatique de la langue arabe.

Pour cette raison, nous sommes intéressés, par la réalisation d'une base de données linguistique (morphologique – syntaxique – sémantique) pour le verbe arabe. Cette base de données sera organisée sous forme d'une matrice en se basant sur la représentation racine-schème connue dans la littérature arabe.

Nous commencerons par présenter les différentes catégories de racines des verbes de la langue arabe. Ensuite, nous décrirons les étapes d'implémentation de notre matrice racine-schème, puis, la présentation statistique de notre base de données.

2 Classification des verbes arabes

Le verbe de la langue arabe est classé selon deux critères principaux : le schème et la racine [El-Dahdeh, 1999]. Dans ce travail, nous nous sommes intéressés au deuxième critère, celui-ci se base sur deux facteurs essentiels : le nombre et la nature (sains ou défectueux¹) de ces éléments.

Selon le nombre d'élément, on distingue deux classes de racines:

- Racine trilitère (الجزر الثلاثي) : c'est toute racine de trois éléments,
- Racine quadrilitère (الجزر الرباعي) : c'est toute racine de quatre éléments.

Selon la nature des éléments, on distingue deux classes principales :

- Racine saine (الجزر الصحيح) : formée des éléments non défectueux. Cette classe est encore répartie en trois sous-classes :
 - Racine saine qui ne contient pas la lettre /hamza/ et ses éléments ne sont pas dupliqués (الجزر السالم)
 - Racine saine contenant la lettre /hamza/ (الجزر المهموز)
 - Racine saine contenant un même élément dupliqué en deuxième et troisième positions (الجزر المضاعف). Cette racine sera représentée dans le verbe seulement par deux consonnes dont la deuxième surmontée par /šadda/.

¹ L'alphabet de la langue arabe distingue trois lettres défectueuses sont : "ي", "و" et "ا"

- Racine défectueuse (الجزر المعتل) : qui a au moins un caractère défectueux. La présence de ces caractères peut causer des altérations morpho-phonologiques importantes au cours de la conjugaison, le verbe engendré par une telle racine risque de rencontrer des transformations et des situations de conjugaisons irrégulières. Nous distinguons plusieurs classes de défectuosité de la racine en fonction du nombre et de la position des caractères défectueux. Nous avons, alors, la classification suivante :
 - Racine trilitère défectueuse dont la défectuosité est au niveau de son premier élément (المثال)
 - Racine trilitère défectueuse dont la défectuosité est au niveau de son deuxième élément (الأجوف)
 - Racine trilitère défectueuse dont la défectuosité est au niveau de son troisième élément (الناقص)
 - Racine trilitère défectueuse dont la défectuosité est à la fois au niveau de son premier élément et son troisième élément (اللفيف المقروق)
 - Racine trilitère défectueuse dont la défectuosité est à la fois au niveau de son deuxième élément et son troisième élément (اللفيف المقرون)

3 Implémentation de la matrice racine-schème

Selon la classification ci-dessus, nous allons implémenter une base de données lexicale pour le verbe arabe sous forme d'une matrice en se basant sur la représentation racine-schème. Cette matrice contient également des codes indiquant les caractéristiques morpho-syntaxiques et des cellules contenant le sens des verbes engendrés.

Pour la réalisation de ce projet, nous nous sommes basés sur un ensemble de ressources linguistiques : les dictionnaires [معجم جهرة, أبو عبد الرحمن الخليل بن أحمد الفراهيدي] [معجم العين, ابن دريد أبي بكر محمد بن الحسن الأزدي البصري, 900] [اللغة أبي القاسم علي بن جعفر بن علي] [كتاب الأفعال] [ابن دريد أبي بكر محمد بن الحسن الأزدي البصري, 900] [اللغة جمال الدين عبد الله] [جمال الدين عبد الله] [ابن هشام الأنصاري, 2003] [ابن عصفور الإشبلي, 669], [ابن هشام الأنصاري, 2003]. Les deux premiers sont parmi les dictionnaires les plus anciens qui regroupent la plupart des entrées lexicales de l'arabe classique, par contre, le dernier est un nouveau dictionnaire comportant les entées lexicales de l'arabe moderne.

Nous présentons dans ce paragraphe les techniques et les étapes de la réalisation de cette matrice.

3.1 Matrice racine-schème pour les verbes trilitères sains

L'implémentation de la matrice des racines trilitères à passé par trois étapes ; Génération des combinaisons, Filtration phonologique et Filtration lexicale

3.1.1 Génération des combinaisons

Afin d'implémenter une base de données exhaustive comportant toutes les racines de la langue arabe, nous avons commencé par la génération de toutes les combinaisons de trois éléments possibles sur l'alphabet de la langue arabe. Pour cela, nous avons codé l'alphabet

par des caractères latins puis nous avons implémenté un algorithme en langage C pour la génération automatique des combinaisons.

La liste des combinaisons générées (en nombre de 15625 combinaisons) contient en plus aux racines recherchées d'autres combinaisons non utilisées dans littérature (combinaisons reconnues). L'étape suivante sera donc la filtration des racines (combinaisons reconnues) à partir de la liste obtenue en adoptant deux phases de filtration : filtration phonétique et filtration lexicale.

3.1.2 Filtration phonétique

L'enchaînement des unités acoustiques dans la langue arabe doit subir certaines règles phonologiques. Ceux-ci définissent une classification des lettres de l'alphabet arabe et caractérisent des contraintes de compatibilité et homogénéité pour l'enchaînement des lettres dans la composition des mots. A titre d'exemple, la règle phonologique relative à l'inharmonie des lettres de siffle [أبو الفتح عثمان بن جني, 1993] :

"حروف الصفيير وهي الصاد والسين والزاي لا تتركب بعضها مع بعض ليس في الكلام مثل: سصن وصسن وسنز"
Les lettres de siffle: "س", "ص" et "ز" ne se composent pas entre elles.

Une combinaison qui ne respecte pas cette règle phonologique ne peut pas constituer une racine parce qu'elle n'est pas acceptée phonologiquement dans la littérature arabe. Nous allons appeler telle combinaisons « *combinaisons non harmonieuses* », par contre nous allons appeler une « *combinaison harmonieuse* » toute combinaison qui respecte à la fois toutes les règles phonologiques de la langue arabe.

Nous avons collecté toutes ces règles phonologiques, ensuite, nous les avons programmées sous forme de *Requêtes SQL* pour supprimer les combinaisons non harmonieuses.

Nous présentons à titre d'exemple l'implémentation de la règle phonologique relative à l'inharmonie des lettres de siffle présentée ci-dessus :

```
DELETE racinesTrilitaires.fAe, racinesTrilitaires.cayn, racinesTrilitaires.lAm
FROM racinesTrilitaires
WHERE (((racinesTrilitaires.fAe)="z" Or (racinesTrilitaires.fAe)="s" Or
(racinesTrilitaires.fAe)="sad") AND ((racinesTrilitaires.cayn)="z" Or (racinesTrilitaires.cayn)="s"
Or (racinesTrilitaires.cayn)="sad")) OR (((racinesTrilitaires.cayn)="z" Or
(racinesTrilitaires.cayn)="s" Or (racinesTrilitaires.cayn)="sad") AND ((racinesTrilitaires.lAm)="z"
Or (racinesTrilitaires.lAm)="s" Or (racinesTrilitaires.lAm)="sad"; (
```

L'exécution de cette requête nous a permis de supprimer 414 combinaisons non harmonieuses. Ainsi que l'exécution de toutes les requêtes de suppression que nous avons implémentées à partir des règles phonologiques collectées nous a permis de supprimer 3683 combinaisons non harmonieuses et de réduire le nombre des cas que nous allons étudier dans l'étapes suivantes.

3.1.3 Filtrage lexical

La liste des combinaisons harmonieuses (en nombre de 11942 combinaisons), obtenue après étape de filtrage phonologique, contient en plus aux racines des combinaisons non reconnues dans la littérature arabe. En effet, une combinaison harmonieuse qui ne possède aucun sens ne peut pas constituer une entrée lexicale malgré qu'elle soit acceptée phonologiquement.

Nous avons alors besoin d'une autre phase de filtration a fin de trier les racines à partir de la liste des combinaisons harmonieuses, il s'agit de la phase de filtrage lexical.

L'objectif de cette phase est, non seulement de sélectionner les combinaisons qui représentent des entrées lexicales, mais aussi de chercher pour chaque combinaison reconnue l'ensemble des schèmes appropriés et d'affecter pour chaque compatibilité racine-schème les propriétés morpho-syntaxiques du verbe engendré.

Pour cela, nous avons attribué pour chaque paire "racine-schème" un code numérique variant de 0 à 7 :

- Le code «0» indique que la paire "racine - schème" n'est pas encore traitée.
- Le code «1» indique l'incompatibilité de la parie "racine - schème".
- un code supérieur ou égal à 2 est attribué dans le cas d'une compatibilité racine-schème, il indique les caractéristiques morpho-syntaxiques du verbe obtenu².

L'affectation de ces codes est faite en deux étapes :

La première étape est automatique, dans laquelle nous avons reprogrammé les règles phonologiques utilisées dans la phase de filtration phonologiques sous forme de requête de mise à jour, en prenant en considération la possibilité de l'enchaînement d'un élément de la racine et l'une des lettres d'augmentation³ (حروف الزيادة). Ceci nous a permis d'attribuer automatiquement le code 1 pour les paires "racine-schème" non compatibles pour des raisons phonologiques, par suite, de réduire un nombre de compatibilités racine-schème que nous allons étudier (manuellement) dans la deuxième étape.

Dans la deuxième étape, nous avons traité les paires "racine-schème" qui n'ont pas encore un code à partir de la première étape. Nous avons traité ces racine d'une manière manuelle, une par une, en se basant sur l'existence de l'entée lexicale dans les dictionnaires ابن دريد أبي بكر [معجم جهرة اللغة] et [أبو عبد الرحمن الخليل بن أحمد الفراهيدي] معجم العين محمد بن الحسن الأزدي البصري، 900

Une combinaison harmonieuses non compatible avec tous les schèmes sera considérée non reconnue et doit être supprimée de la liste de racines. Techniquement, nous avons implémenté une requête pour supprimer les combinaisons dont la paire "racine-schème" a le code 1 pour tous les schèmes.

En fin, nous avons obtenu une liste de toutes les racines trilitères saines reconnues dans la littérature arabe en nombre de 8794 racines. Nous l'avons présentée sous forme de matrice dont l'en-tête de lignes contient les schèmes de verbes et l'en-tête de colonnes contient les racines, à l'intérieur de la matrice, il ya le code indiquant les caractéristiques morpho-syntaxiques du verbe obtenu dans le cas de la compatibilité racine-schème et le code 0 dans le cas contraire.

² Les codes des racines reconnus sont attribués de la façon suivante : 2 → {لازم- مجرد- تام التصرف}، 3 → {لازم- مجرد- ناقص التصرف}، 4 → {تام التصرف}، 5 → {لازم- مجرد- ناقص التصرف}، 6 → {متعد - مجرد- تام التصرف}، 7 → {لازم- مزيد- تام التصرف}، متعد- {متعد-}، متعد- {لازم- مزيد- تام التصرف}، متعد- {متعد - مجرد- تام التصرف}، متعد- {لازم- مجرد- ناقص التصرف}، مجرد- ناقص التصرف،

³ Se sont les caractères qu'on ajoute à la racine pour former les schèmes augmentés, ils sont en nombre de dix caractères : "هـ"، "و"، "ي"، "ن"، "م"، "ت"، "ل"، "س"، "ر"، "ز"، "ح"، "ع"، "ا"

racine	فعل	sense facala
ب - ت - ر	5	بتر الشيء يبتره بترًا إذا قطعته
ب - ت - ك	5	بنتك الشيء يبتكه بتكا إذا قطعته
ب - ت - ل	5	بنت الشيء يبتله بتلا إذا قطعته
ب - ت - ع	1	rien
ب - ت - ق	2	بقق الماء إذا انفجر من حوض
ب - ج - ح	5	بجحت بالشيء أبحج وبجحت أيضا إذا فرحت به
ب - ج - د	5	بجد بالمكان يبجد بجودا إذا أقام به فهو باجد
ب - ج - س	5	بجست الشيء إذا شققته
ب - ج - ل	52	بجل الصباح إذا أضاء وبجلته بجلا إذا قطعت أجزائه وهو الأكل
ب - ج - م	2	بجم الرجل يجم بجمًا وجمومًا إذا سكت من عي أو هيبة فهو باجم
ب - ج - ن	1	rien
ب - ح - ت	5	بحثت عن الشيء أبحث بحثًا إذا كتفت عنه واستقصيت خبره
ب - ح - ر	52	بحرت الأذن إذا شققها وبحر الرجل إذا اجتهد في العذر
ب - خ - ر	2	بخرت القدر بخرا إذا سطع بخارها
ب - خ - س	5	بخسته حقه إذا ظلمته
ب - خ - ص	5	بخص عينه إذا أصاب بخصيتها أي إذا ادخل أصبعه فيها و البخص لحم العين
ب - خ - ق	5	بخق العين إذا عارها
ب - د - ر	5	بدرت الي الرجل تقدمت إليه
ب - د - ع	5	بدعت الشيء إذا أنشأته وبدعت الركي إذا استنبطتها
ب - د - غ	2	بدغ بدغا إذا جر البئنه على الأرض
ب - د - ل	2	بدل الرجل بدلا إذا وجعه يداه ورجلاه
ب - د - ن	2	بدن الرجل أي سمن
ب - د - هـ	5	بدهه يدهه بدها والمبادعة والبيبية هو أن يقاظك أمرا وتشي كلاما لم تستد له
ب - ذ - خ	52	بذخ بذخا إذا طل في كلامه وفقره وبذخ لسانه إذا فلقه وبذخ البجير إذا اشك هديره
ب - ذ - ر	5	بذر الرجل النبات من البذر

Figure 1: Partie de matrice racine-schème pour la racine trilitère saine

Nous avons ajouté à côté de chaque colonne d'un schème une autre colonne contenant le sens du verbe obtenu dans le cas de la compatibilité racine-schème. Exp. : Le code 5 dans la première ligne indique que :

- la racine "ب، ت، ر" est compatible avec le schème "فَعَلْ" et le verbe obtenu est "بَتَرَ"، son sens est : "بتر الشيء يبتره إذا قطعته"، c'est-à-dire "amputer une chose c'est la découper".
- le verbe obtenu a les propriétés linguistiques suivantes : (مجرد، متعد، تام التصرف)

Nous avons réalisé les matrices de la racine trilitère avec /hamza/, la racine trilitère avec /sadda/, la racine trilitère défectueuse en poursuivant les mêmes étapes.

3.2 Matrice de la racine quadrilitère

Nous distinguons deux types de racines quadrilitères saines ; la racine quadrilitère saine composé par une chaîne de deux lettres dupliquée (telle que زعزع) et la racine quadrilitère saine composée par une chaîne de quatre lettres différentes (telle que بعثر). Pour la première classe nous avons suivi la même démarche que celle utilisée pour la racine trilitère ; c'est-à-dire, générer automatiquement la liste de toutes les combinaisons possibles puis suivre les procédures de filtration utilisées précédemment. Ceci a été possible parce que le nombre des combinaisons générées ne dépasse pas 784 (28 × 28). Cependant, nous n'avons pas suivi cette démarche pour le cas des racines quadrilitères de

la deuxième classe pour deux raisons ; d'une part, parce que le nombre de combinaisons générées sera très grand ($614656 = 28 \times 28 \times 28 \times 28$), donc très difficile à traiter, d'autre part, parce que ce genre de racines est très rares dans la littérature arabe. Alors, nous avons décidé d'implémenter cette matrice en collectant ces racines à partir des dictionnaires de la langue arabe les plus connus et les intégrant dans la matrice principale.

racine	فعل	sense faclala
برقط	2	برقط إذا صعد في الجبل فسقط
برقع	5	برقعت الذابة والجارية البسببهما البرقع
بركع	2	بركع بركة إذا قام على أربع
بركل	2	بركل الرجل إذا مشى في الطين والماء
برنت	1 rien	
برنت	1 rien	
برنس	2	برنس الرجل إذا أسرع وتبخر
برهم	2	برهم الشجر إذا اجتمع ورقه وتمرد
بزعر	1 rien	
بزمخ	2	بزمخ الرجل إذا تكبر
بزنت	1 rien	
بسلم	2	بسلم الرجل إذا كره وجهه
بسمل	25	بسمل بسملة إذا أكثر من قول بسم الله
بعثر	5	بعثرت القبر وغيره إذا بددت ترابه
بعثق	1 rien	
بحتر	5	بحترني إذا نقصني
بحرص	1 rien	
بحزج	5	بحزج الشيء إذا فرقته
بعكر	5	بعكره بالسيف إذا ضربه
بعنق	1 rien	
بختر	1 rien	
بقتط	5	بقتط متاعه ويحتره
بكباك	5	بكباك الشيء إذا طرحته بعضه على بعض
بلاثر	52	بلاثر الرجل إذا حاد من فزع وبلاثر الرجل إذا أكل
بالأص	2	بالأص الرجل إذا حاد من فزع
بلجج	1 rien	
بلجم	5	بلجم البيطار الحمار إذا شد قوائمه من داء بصيبه

Figure 2: Partie de matrice racine-schéme pour la racine quadrilitère

4 Données statistiques

Dans ce qui suit vous trouverez la représentation statistique de notre base de données

4.1 Racines trilitères saines sans /hamza/ et sans /šadda/

Elles sont au nombre de 3262 racines reconnues dans la littérature arabe, 2332 entre elles sont compatibles avec le schème "فَعَل", 1137 sont compatibles avec le schème "فعل" et 298 sont compatibles avec le schème "فَعْل". Cet écart peut être dû à la tendance chez les Arabes de parler avec fente (الفتحة) en raison de la facilité de la prononciation. Nous remarquons aussi que 761 racines sont compatibles à la fois avec les deux schèmes "فَعَل" et "فعل". Parmi ces racines il y'a certains qui engendrent deux verbes identiques en terme de caractéristiques morpho-syntaxique et de sens (exemple 1) et d'autres racines qui engendrent deux verbes totalement différents (exemple 2).

Exemples:

- 1- $\text{خَضِبَ} \rightarrow (\text{فعل}) + (\text{خ} - \text{ض} - \text{ب})$ } : "خَضِبَ الشَّجَرَ" et "خَضِبَ الشَّجَرَ" : deux verbes ayant le même sens et les mêmes caractéristiques morpho-syntaxiques
 $\text{خَضِبَ} \rightarrow (\text{فعل}) + (\text{خ} - \text{ض} - \text{ب})$
- 2- $\text{عَهَدَ} \rightarrow (\text{فعل}) + (\text{ع} - \text{ه} - \text{د})$: "عَهَدْتُكَ مَلْتَزِمًا" \rightarrow (verbe transitif), } deux verbes totalement différents
 $\text{عَهَدَ} \rightarrow (\text{فعل}) + (\text{ع} - \text{ه} - \text{د})$: "عَهَدْتُ إِلَيْكَ بِشَيْءٍ" \rightarrow (verbe non transitif)

Nous présentons ici un tableau qui dénombre la compatibilité de la racine trilitère saine sans /hamza/ et sans /sadda/ avec les schèmes augmentés :

Schème Racine	أَفْعَل	فَعَّل	فَاعَلَ	أَنْفَعَلَ	أَفْتَعَلَ	أَفْعَلَّ	تَفَعَّل	تَفَاعَلَ	اسْتَفْعَلَ	أَفْعُوْعَل	أَفْعُوْل	أَفْعَالٌ
Transitif	529	236	46	26	94	1	76	27	21	0	1	0
Non transitif	423	50	14	44	37	13	99	58	11	13	1	7
Total	952	286	60	70	131	14	175	85	32	13	2	7

Tableau 1: Compatibilité de la racine trilitère saine sans /hamza/ et sans /sadda/ avec les schèmes augmentés

4.2 Racines trilitères saines avec /sadda/

Elles sont au nombre de 470 racines, le tableau ci-dessous dénombre la compatibilité de ce type de racines avec les schèmes de verbes réduits et augmentés.

Schème Racine	فَعَلَ	فَعِلَ	فَعُلَ	أَفْعَلَ	أَنْفَعَلَ	فَاعَلَ	فَعَّلَ	أَفْتَعَلَ	تَفَعَّلَ	تَفَاعَلَ	اسْتَفْعَلَ	أَفْعُوْعَلَ	أَفْعُوْلَ	أَفْعَالًا
Transitif	103	14	0	57	2	0	2	2	3	2	1	3	0	0
Non transitif	172	11	1	64	2	4	9	5	7	1	5	0	0	0
Total	275	25	1	121	4	4	11	7	10	3	6	3	0	0

Tableau 2: Compatibilité de la Racine trilitère saine avec /sadda/ avec les schèmes de verbes

4.3 Racines trilitères saines avec /hamza/

Elles sont au nombre de 612 racines, réparties en trois sous-classes ; 126 racines avec /hamza/ à la première position, 186 racines avec /hamza/ en deuxième position, 300 racines avec /hamza/ en troisième position⁴. Les trois tableaux ci-dessous dénombrent la compatibilité de ces racines avec les différents schèmes

⁴ Nous avons trouver dix racines avec /hamza/ à la fois en première position et troisième position : "أ - ب - أ", "ج - أ", "أ - ي - أ", "أ - و - أ", "أ - ز - أ", "أ - ش - أ", "أ - ل - أ", "أ - ك - أ", "أ - ث - أ", "أ - ت - أ".

Schème Racine	فَعَلَ	فَعِلَ	فَعُلَ	أَفْعَلَ	أَنْفَعَلَ	فَاعَلَ	فَعَّلَ	أَفْتَعَلَ	تَفَعَّلَ	تَفَاعَلَ	اسْتَفْعَلَ	أَفْعُوَعَلَ	أَفْعُوَلَّ	أَفْعَالًا
Transitif	18	16	0	1	0	1	3	1	3	0	1	0	0	1
Non transitif	44	9	0	5	0	0	12	1	9	1	2	0	0	0
Total	62	25	0	6	0	1	15	2	12	1	3	0	0	1

Tableau 3: Compatibilité de la racine trilitère saine avec /hamza/ en première position avec les schèmes de verbes

Schème Racine	فَعَلَ	فَعِلَ	فَعُلَ	أَفْعَلَ	أَنْفَعَلَ	فَاعَلَ	فَعَّلَ	أَفْتَعَلَ	تَفَعَّلَ	تَفَاعَلَ	اسْتَفْعَلَ	أَفْعُوَعَلَ	أَفْعُوَلَّ	أَفْعَالًا
Transitif	26	18	7	13	0	0	3	2	2	4	1	0	0	1
Non transitif	56	8	0	19	0	6	12	1	2	3	0	0	0	0
Total	82	26	7	32	0	6	15	3	4	7	1	0	0	1

Tableau 4: Compatibilité de la Racine trilitère saine avec /hamza/ en deuxième position avec les schèmes de verbes

Schème Racine	فَعَلَ	فَعِلَ	فَعُلَ	أَفْعَلَ	أَنْفَعَلَ	فَاعَلَ	فَعَّلَ	أَفْتَعَلَ	تَفَعَّلَ	تَفَاعَلَ	اسْتَفْعَلَ	أَفْعُوَعَلَ	أَفْعُوَلَّ	أَفْعَالًا
Transitif	33	19	14	17	0	0	0	1	1	2	2	0	0	10
Non transitif	98	15	3	42	0	12	14	5	7	1	4	0	0	1
Total	131	34	17	59	0	12	14	6	8	3	6	0	0	11

Tableau 5: Compatibilité de la Racine trilitère saine avec /hamza/ en troisième position avec les schèmes de verbes

4.4 Racines trilitères défectueuses

Nous avons obtenu 1892 racines défectueuses, ces racines sont réparties en plusieurs classes selon leur nature de défectuosité. Les tableaux ci-dessous dénombrent la compatibilité de ces racines avec les différents schèmes.

Schème Racine	فَعَلَ	فَعِلَ	فَعُلَ	أَفْعَلَ	أَنْفَعَلَ	فَاعَلَ	فَعَّلَ	أَفْتَعَلَ	تَفَعَّلَ	تَفَاعَلَ	اسْتَفْعَلَ	أَفْعُوَعَلَ	أَفْعُوَلَّ	أَفْعَالًا
Transitif	51	55	25	34	0	0	4	0	2	0	3	0	0	0
Non transitif	94	11	1	57	0	12	20	0	7	1	3	0	0	0
Total	145	66	26	91	0	12	24	0	9	1	6	0	0	0

Tableau 6: Compatibilité de la Racine trilitère défectueuse dont le premier élément est "و" avec les schèmes de verbes

Schème Racine	فَعَلَ	فَعِلَ	فُعِلَ	أَفْعَلَ	انْفَعَلَ	فَاعَلَ	فَعَّلَ	افْتَعَلَ	تَفَعَّلَ	تَفَاعَلَ	اسْتَفْعَلَ	افْعُوْ عَلَ	افْعُوْلَ	افْعَالًا
Transitif	4	3	7	7	0	0	0	0	0	0	0	0	0	0
Non transitif	8	1	1	6	0	1	0	0	1	0	1	0	0	0
Total	12	4	8	13	0	1	0	0	1	0	1	0	0	0

Tableau 7: Compatibilité de la Racine trilitère défectueuse dont le premier élément est "ي" avec les schèmes de verbes

Schème Racine	فَعَلَ	فَعِلَ	فُعِلَ	أَفْعَلَ	انْفَعَلَ	فَاعَلَ	فَعَّلَ	افْتَعَلَ	تَفَعَّلَ	تَفَاعَلَ	اسْتَفْعَلَ	افْعُوْ عَلَ	افْعُوْلَ	افْعَالًا
Transitif	76	37	0	28	4	3	6	1	13	6	2	0	0	0
Non transitif	108	4	0	54	1	10	73	1	12	6	4	0	0	0
Total	184	41	0	82	5	13	43	2	25	12	6	0	0	0

Tableau 8: Compatibilité de la Racine trilitère défectueuse dont le deuxième élément est "و" avec les schèmes de verbes

Schème Racine	فَعَلَ	فَعِلَ	فُعِلَ	أَفْعَلَ	انْفَعَلَ	فَاعَلَ	فَعَّلَ	افْتَعَلَ	تَفَعَّلَ	تَفَاعَلَ	اسْتَفْعَلَ	افْعُوْ عَلَ	افْعُوْلَ	افْعَالًا
Transitif	74	7	0	36	3	2	4	2	5	2	2	0	0	0
Non transitif	71	4	0	19	2	5	20	2	10	0	3	1	0	0
Total	145	11	0	55	5	7	24	4	15	2	5	1	0	0

Tableau 9: Compatibilité de la Racine trilitère défectueuse dont le deuxième élément est "ي" avec les schèmes de verbes

Schème Racine	فَعَلَ	فَعِلَ	فُعِلَ	أَفْعَلَ	انْفَعَلَ	فَاعَلَ	فَعَّلَ	افْتَعَلَ	تَفَعَّلَ	تَفَاعَلَ	اسْتَفْعَلَ	افْعُوْ عَلَ	افْعُوْلَ	افْعَالًا
Transitif	62	7	7	57	2	0	2	2	3	2	1	3	0	0
Non transitif	102	3	1	64	2	4	9	5	7	1	5	0	0	0
Total	164	10	8	121	4	4	11	7	10	3	6	3	0	0

Tableau 10: Compatibilité de la Racine trilitère défectueuse dont le troisième élément est "ز" avec les schèmes de verbes

Schème Racine	فَعَلَ	فَعِلَ	فَعُلَ	أَفْعَلَ	انْفَعَلَ	فَاعَلَ	فَعَّلَ	افْتَعَلَ	تَفَعَّلَ	تَفَاعَلَ	اسْتَفْعَلَ	أَفْعُوْعَلَ	أَفْعُوْلَ	أَفْعَالًا
Transitif	47	73	0	26	2	0	2	2	3	2	1	3	0	0
Non transitif	94	14	1	67	2	4	9	5	7	1	5	0	0	0
Total	141	87	1	93	4	4	11	7	10	3	6	3	0	0

Tableau 11: Compatibilité de la Racine trilitère défectueuse dont le troisième élément est "ي" avec les schèmes de verbes

Schème Racine	فَعَلَ	فَعِلَ	فَعُلَ	أَفْعَلَ	انْفَعَلَ	فَاعَلَ	فَعَّلَ	افْتَعَلَ	تَفَعَّلَ	تَفَاعَلَ	اسْتَفْعَلَ	أَفْعُوْعَلَ	أَفْعُوْلَ	أَفْعَالًا
Transitif	4	3	0	3	0	0	0	0	0	1	0	0	0	0
Non transitif	8	1	0	7	0	1	3	0	0	0	0	0	0	0
Total	12	4	0	10	0	1	3	0	0	1	0	0	0	0

Tableau 12: Compatibilité de la Racine trilitère défectueuse dont le premier élément est "و" et le troisième élément est "ي" avec les schèmes de verbes

Schème Racine	فَعَلَ	فَعِلَ	فَعُلَ	أَفْعَلَ	انْفَعَلَ	فَاعَلَ	فَعَّلَ	افْتَعَلَ	تَفَعَّلَ	تَفَاعَلَ	اسْتَفْعَلَ	أَفْعُوْعَلَ	أَفْعُوْلَ	أَفْعَالًا
Transitif	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Non transitif	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Total	2	0	0	0	0	0	0	0	0	0	0	0	0	0

Tableau 13: Compatibilité de la Racine trilitère défectueuse dont le deuxième élément est "و" et le troisième élément est "ي" avec les schèmes de verbes

Schème Racine	فَعَلَ	فَعِلَ	فَعُلَ	أَفْعَلَ	انْفَعَلَ	فَاعَلَ	فَعَّلَ	افْتَعَلَ	تَفَعَّلَ	تَفَاعَلَ	اسْتَفْعَلَ	أَفْعُوْعَلَ	أَفْعُوْلَ	أَفْعَالًا
Transitif	3	6	0	0	0	0	1	0	0	0	0	0	0	0
Non transitif	8	2	0	3	0	2	2	1	1	1	1	0	0	0
Total	11	8	0	3	0	2	3	1	1	1	1	0	0	0

Tableau 14: Compatibilité de la Racine trilitère défectueuse dont le deuxième élément est "ي" et le troisième élément est "و" avec les schèmes de verbes

4.5 Racines quadrilitères

Elles sont au nombre de 1306 racines : 313 racines quadrilitères saines gémées, 873 racines quadrilitères saines dégémées, et 120 racines quadrilitères défectueuses.

Schème Racine	فعل	فعل	فعل	فعل	فعل	تمفعّل	تفعّل	تفعّل	تفعّل	تفعّل	افعلّل	افعلّل	افعلّل	افعلّل	تفوعّل
Transitif	107	0	0	0	0	10	0	0	20	0	0	0	0	0	0
Non transitif	133	0	0	0	0	17	0	0	26	0	0	0	0	0	0
Total	240	0	0	0	0	27	0	0	46	0	0	0	0	0	0

Tableau 15: Compatibilité de la Racine quadrilitère saine gémée avec les schèmes de verbes

Schème Racine	فعل	فعل	فعل	فعل	فعل	تمفعّل	تفعّل	تفعّل	تفعّل	تفعّل	افعلّل	افعلّل	افعلّل	افعلّل	تفوعّل
Transitif	299	0	0	0	0	5	0	2	45	0	42	132	0	19	0
Non transitif	249	0	0	0	0	3	0	0	29	0	3	13	0	0	0
Total	584	0	0	0	0	8	0	2	74	0	45	145	0	19	0

Tableau 16: Compatibilité de la Racine quadrilitère saine dégémée avec les schèmes de verbes

Schème Racine	فعل	فعل	فعل	فعل	فعل	تمفعّل	تفعّل	تفعّل	تفعّل	تفعّل	افعلّل	افعلّل	افعلّل	افعلّل	تفوعّل
Transitif	45	0	0	0	0	0	0	0	6	0	42	4	0	1	0
Non transitif	14	0	0	0	0	0	0	0	5	0	3	0	0	0	0
Total	59	0	0	0	0	0	0	0	11	0	45	4	0	1	0

Tableau 17: Compatibilité de la Racine quadrilitère défectueuse avec les schèmes de verbes

5 Conclusion

Dans ce travail, nous avons implémenté une base de données exhaustive pour le verbe de la langue arabe sous forme d'une matrice en se basant sur la représentation racine-schème. Chaque schème est représenté par deux colonnes ; la première contient un code indiquant sa compatibilité avec la racine ainsi que les caractéristiques morpho-syntaxique du verbe engendré par cette compatibilité, la deuxième contient le sens du verbe engendré tel qu'il est dans les dictionnaires (forme textuelle).

Dans le cadre de la représentation du sens en TALN, nous envisageons de compléter notre travail par l'affectation d'un code indiquant le sens. Cette base de données lexicale sémantique va nous servir pour implémenter un automate morpho-syntaxico-sémantique de l'arabe standard.

6 Références

- [1] أبو الفتح عثمان بن جني، 1955، الخصائص، تحقيق محمد علي النجار الجزء II، الناشر دار الكتاب العربي بيروت لبنان.
- [2] أبو الفتح عثمان بن جني، 1993، سر صناعة الإعراب، دراسة وتحقيق د. حسن هندواوي الجزء I و II، الناشر دار القلم، دمشق.
- [3] جمال الدين عبد الله بن هشام الأنصاري، 2003، شرح شذور الذهب، الطبعة الأولى، الناشر دار بن كثير، دمشق.
- [4] عبد الرحمن جلال الدين السيوطي، المزهر في علوم اللغة وأنواعها، شرح وتعليق محمد المولى بك- محمد أبو الفضل- إبراهيم علي - محمد البجاوي، الجزء II، منشورات المكتبة العصرية بيروت.
- [5] ابن عصفور الإشبلي، 669، الممتع في التصريف، تحقيق الدكتور فخر الدين قباوة، الجزء I، دار المعرفة، بيروت- لبنان.
- [6] ابن دريد أبي بكر محمد بن الحسن الأزدي البصري، 900، معجم جهرة اللغة، الجزء I و II.
- [7] أبو عبد الرحمن الخليل بن أحمد الفراهيدي، 100، معجم العين، تحقيق الدكتور مهدي المخزومي، د. إبراهيم السامرائي الأجزاء 1 و 2 و 3 و 4 و 5 و 6 و 7 و 8.
- [8] أبي القاسم علي بن جعفر بن علي السعدي، المعروف بابن القطاع الصقلي، 2003، كتاب الأفعال، تحقيق إبراهيم شمس الدين، منشورات محمد علي بيضون، دار الكتب العلمية بيروت لبنان، الطبعة الأولى.
- [9] خديجة الحديثي، 2003، أبنية الصرف في كتاب سيويوه معجم ودراسة، الطبعة الأولى، مكتبة لبنان ناشرون، بيروت.
- [10] مصطفى الغلايني، 2006، جامع الدروس العربية، دار الفكر للطباعة والنشر والتوزيع، الطبعة الأولى، بيروت.
- [11] بيبب عبد العزيز عتيق، مدخل إلى علم النحو والصرف، دار النهضة العربية للطباعة والنشر.
- [12] **Richard Timon Daly**, 1974, Application of the Mathematical Theory of Linguistic, Monton The Hague paris.
- [13] **Y. TAHIR**, 2006, Modélisation A Objets d'une Base de Données Morpho. Syntaxiques pour la Langue Arabe, Thèse de Diplôme de Doctorat, Spécialité Informatique.
- [14] **Jhon Von Neunnan**, Théorie Générale et Logique Des Automates, Traduit de Langlais par Jean- Poul Auffrand Précède de la Pensée et les Machines: le Mécanisme Algorithme de Jhon Von Neunnan pas Gerard Ghazal Collection Milieux Champ Volon.