Une approche multicritère pour lever l'ambiguïté morphologique dans le texte arabe

Mohamed Amine Chéragui[1], Youssef Hceini[2] et Moncef Abbas[3]

[1]Ecole national Supérieure d'Informatique (E.S.I.), E-mails : m_cheragui@esi.dz
[2]Institut d'informatique, Université de Bechar, B.P. 417 Bechar, E-mails : y_hoceini@yahoo.fr
[3]USTHB, Faculté de Mathématiques, Dépt. R.O, E-mails : moncef_abbas@yahoo.com

Résumé : Nous essayons à travers cet article, de décrire le phénomène de l'ambiguïté morphologique dans le traitement automatique du texte arabe et de présenter une nouvelle approche de désambiguïsation morphologique différente de celles qui existent actuellement, comme l'approche par contraintes ou l'approche stochastique (probabiliste/statistique), où l'originalité de notre travail réside dans le fait de s'appuyer sur un formalisme mathématique solide qui est l'aide multicritère à la décision.

Mots clés : TALN[1]; TALA[2] ; Ambiguïté morphologique; désambiguïsation; étiquetage morphosyntaxique; AMD[3]; critère; agrégation; pondération.

1 Introduction

Depuis l'apparition de l'intelligence artificielle au milieu des années cinquante, plusieurs chercheurs ont eux l'idée de modéliser et reproduire, à l'aide d'ordinateur la capacité humaine à produire et à comprendre des énoncés linguistiques dans le but de communication. Donnant ainsi naissance à une nouvelle discipline connue sous l'acronyme TALN qui veut dire Traitement Automatique du Langage Naturel faisant intervenir plusieurs sciences comme : la linguistique, la logique, l'informatique théorique, les statistiques, la neuroscience…etc.

Aujourd'hui, avec le développement qu'a subit l'informatique que se soit en terme de vitesse de traitement ou de support de stockage, le traitement automatique du langage naturel est devenu un domaine à la fois technologique due à l'émergence d'un nombre important d'applications, tels que : les traducteurs automatiques, générateurs automatiques de résumé, correcteurs orthographiques d'erreurs, …etc. Mais aussi un domaine scientifique traitant des problématiques de plus en plus complexes comme celle de l'ambiguïté.

Dans la littérature l'ambiguïté est comparée à un état de confusion, cet embrouillement se manifeste sous différentes formes et selon les différents niveaux de traitements que se soit lexical, morphologique, syntaxique et même sémantique. L'une

[1] TALN : Traitement Automatique de la Langue Naturel.
[2] T.A.L.A : Traitement Automatique de la Langue Arabe.
[3] A.M.D : Aide multicritère à la décision.

des formes d'ambiguïté la plus persistant en traitement automatique de la langue arabe est l'ambiguïté morphologique.

Dans cet article, nous donnons une brève et complète description du phénomène de l'ambiguïté morphologique ainsi que les différentes approches de désambiguïsation existantes. Nous présenterons ensuite notre approche pour lever cette forme d'ambiguïté qui se base sur les techniques de l'aide multicritère à la décision que nous appliquerons à la langue arabe.

2 La morphologie : Principe et Objective

2.1 Principe

La morphologie est un domaine de la langue qui permet la description des règles régissant la structure interne des mots (unités lexicales), chez les grammairiens la morphologie est l'étude des formes des mots (flexion et dérivation), en d'autres termes, la morphologie est l'étude des mots considéré isolément (hors contexte) sous le double aspect de la nature et des variations qu'ils peuvent subir [1]. En langue arabe, l'analyse morphologique est d'autant plus importante que les mots sont fortement agglutinés[4], c'est-à-dire qu'ils sont formés dans leur majorité par assemblage d'unités lexicales et grammaticales élémentaires [2].

Ainsi Le traitement morphologique est considéré comme une introduction principale à la compréhension globale d'une langue naturelle ; il joue un rôle très important aussi bien du côté linguistique que du côté technique.

2.2 Objective

La plupart des études faites sur la morphologie arabe dans le passé ou bien aujourd'hui visent généralement à satisfaire les points suivants:

- La formation de nouveaux mots à partir des éléments lexicaux disponibles ;
- L'analyse des mots réellement existant ;
- Fournir les données nécessaires aux travaux des différents niveaux de traitement (syntaxe, sémantique et pragmatique) ;

3 L'ambiguïté morphologique dans le T.A.L.A.

Le traitement morphologique porte sur l'unité élémentaire identifiable, en l'occurrence le morphème, à ce niveau on s'intéresse à deux concepts essentiels qui sont :

- *la synthèse* qui permet de générer des mots ou des phrases par le biais d'un ensemble de règles de dérivation, de flexions et d'adaptations.

[4] Processus d'ajout d'affixes à un mot qui exprime ses différentes relations grammaticales.

- *l'analyse* qui a pour rôle d'associer à un mot graphique un ensemble d'informations décrivant les unités morphologiques et grammaticales entrant dans sa composition (proclitiques, préfixes, base, suffixes, enclitique) [3]. C'est à ce stade, que l'ambiguïté morphologique se manifeste ; lorsque l'analyse assigne à une unité lexicale plusieurs informations (ou l'inverse) ce qui génère la notion de combinatoire [4].

  Exemple :

  Prenons comme exemple, le mot «أحْـمَـدُ », ce dernier peut prendre plusieurs interprétations (voir figure 1) et celà est dû au contexte.

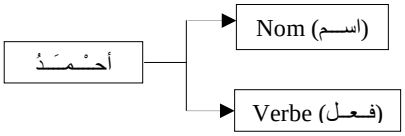

Figure 1. Exemple d'une ambiguïté morphologique.

4 Etiquetage Morphosyntaxique pour lever l'ambiguïté morphologique

4.1 Qu'est ce que le Tagging?

Le Tagging (étiquetage ou marquage) est le fait d'assigner une étiquette ou un tag à un mot. Le tag contient des informations morphosyntaxiques sur le mot, c'est-à-dire des informations concernant la forme et la fonction du mot. Elles comprennent notamment la catégorie grammaticale, le genre, le nombre, le temps et le mode [5].

4.2 Principe de Base

Pour mettre en place un étiqueteur morphosyntaxique de la langue arabe, on doit construire trois (03) modules qui seront complémentaires, ces modules sont :

- *Module de segmentation :* lorsqu'on évoque la segmentation dans le traitement automatique des langues dites naturelles, on parle le plus souvent de trois (03) niveaux de segmentations :
  - Segmentation au niveau du texte ;
  - Segmentation au niveau de la phrase ;
  - Segmentation au niveau du mot.
- *Module d'analyse morphologique :* Le but principal de ce type d'analyse est de vérifier l'appartenance d'un mot donné au domaine linguistique choisi et de pouvoir disposer ainsi de tous les renseignements le concernant pouvant servir à l'analyse syntaxique.
- *Module de désambiguïsation :* La désambiguïsation est une étape cruciale dans le processus d'étiquetage morphosyntaxique, à ce niveau du traitement si un mot est

mal étiqueté, les règles de la grammaire s'appliqueront mal ou pas du tout. Cependant la phase de désambiguïsation n'est pas toujours nécessaire ou obligatoire au bon déroulement du processus d'étiquetage. Il faut dire que le module de désambiguïsation rentre en jeu dans un seul cas de figure, celui où l'unité lexicale (mot) reçoit plus d'une étiquette (plus d'une information morphosyntaxique), ce qui va générer une situation de confusion ou ambiguïté. C'est à ce stade que notre contribution va apparaître en présentant une nouvelle démarche de désambiguïsation différente de celle qui existe actuellement basée sur l'aide multicritère à la décision.

5 Méthodes déjà existantes pour lever l'ambiguïté morphologique arabe

Dans la littérature, les approches de désambiguïsation se répartissent en deux (02) catégories et chaque catégorie englobe une ou plusieurs techniques pour lever l'ambiguïté morphologique. La figure 2 ci-après donne une indication sur les différentes approches et les techniques qui vont avec :

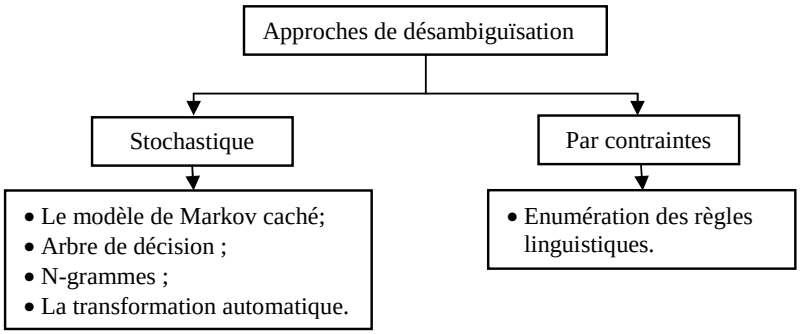

Figure 2: Approches et techniques de désambiguïsation.

5.1 Approche par contraintes

L'approche de désambiguïsation par contraintes (ou par règles) est la méthode la plus ancienne qui a été mis en place pour remédier au problème de l'ambiguïté morphologique. Cette approche se base principalement sur l'intervention d'un linguiste (ou un grammairien) afin d'établir une liste de règles permettant de lever l'ambiguïté. Ces règles sont généralement classées en trois (03) catégories [6], qui sont :

- Les règles contextuelles ou de principes ;
- Les heuristiques ;
- Les règles non contextuelles.

Parmi les étiqueteurs qui adoptent l'approche par contraintes on peut citer : Freeman's Arabic Tagger [7].

5.2 Approche stochastique (probabiliste / statistique)

Lever l'ambiguïté morphologique en adoptant une approche stochastique, cela consiste essentiellement à définir deux (02) sortes d'informations [8] :

- La première est sur le mot a étiqueté (*i.e. l'association entre le mot et l'étiquette*) : Soit « W » une unité lexicale et « T » l'étiquette qui va avec.

$$P(T/W) = \frac{P(T, W)}{P(W)} \cdot \qquad \textbf{(1)}$$

  Où :
  - P(T, W) : le nombre de fois que « W » est étiqueté par « T » ;
  - P(W) : le nombre d'occurrence de « W ».
- La seconde information est contextuelle syntaxique (*i.e. la possibilité de déterminer la probabilité d'avoir une étiquette « T^k » quand elle est précédée de l'étiquette « T^j » dans le texte*).

$$P(T^k/T^j) = \frac{P(T^k, T^j)}{P(T^j)} \cdot \qquad \textbf{(2)}$$

En plus de ces deux (02) hypothèses qui sont considérées comme étant les formules de calcul, une phase d'apprentissage est obligatoire et celà, en entraînant le module de désambiguïsation sur un corpus généralement annoté (*i.e. étiqueter à la main*) au préalable. Parmi les étiqueteurs qui adoptent l'approche stochastique on peut citer : TnT[5] qui utilise le modèle de Markov caché [9].

6 Une approche de désambiguïsation basée sur l'aide multicritère à la décision

Le but de notre étude est de proposer une nouvelle approche de désambiguïsation différente de celles qui existent actuellement, comme l'approche stochastique et l'approche par contrainte. Cette nouvelle approche dédiée à la levée de l'ambiguïté morphologique s'appuie essentiellement sur les principes de la théorie décisionnelle et les méthodes issues de l'aide multicritère à la décision.

6.1 Justification du choix d'une démarche multicritère

L'intérêt d'adopter une approche d'aide multicritère à la décision pour lever l'ambiguïté morphologique dans le T.A.L.A., peut être résumé en deux (02) points, qui sont :

- Réduction de l'ensemble des étiquettes candidates: elle permet de réduire d'amblée le nombre d'étiquettes de correction, en éliminant ceux dominées générant ainsi l'ensemble des étiquette efficaces ;

[5] TnT : Trigrams 'n' Tags.

- Classification des étiquettes efficaces, selon un score global obtenu après traitement suivant un ordre décroissant.

6.2 Différentes étapes d'une démarche d'aide multicritère à la décision

La mise en place d'une démarche basée sur l'aide multicritères à la décision exige le plus souvent de suivre une succession d'étapes afin de générer un résultat. Ces étapes sont [10], [11]:

- Etape 1 : Dresser la liste des scénarios potentiels, cette liste contient les scénarios (*i.e. solution, actions, …etc.*), qui sont provisoirement considérées comme étant candidates ou bien réaliste du point de vue du décideur.
- Etape 2 : Dresser la liste des critères, une bonne application d'une démarche multicritère passe impérativement par un bon choix concernant les critères, sur lesquelles le calcul sera posé. Ces critères seront définis en se basant sur les notions d'indifférence, de préférence stricte ou faible ou de non comparabilité.
- Etape 3 : Définir une fonction d'évaluation pour chaque critère sélectionné une fonction d'évaluation est définie. Cette dernière doit être maximisée ou bien minimisée selon le type de critère utilisé.
- Etape 4 : Etablir un tableau de performance, ce tableau (*i.e. matrice d'évaluation*) contient tous les résultats des évaluations de chaque scénario potentielle suivant tous les critères ; de telle sorte que les lignes de ce tableau correspondent aux scénarios potentielles et les colonnes correspondent aux critères, et les intersections correspondent aux évaluations.
- Etape 5 : Pondération et agrégation des performances, afin de s'assurer à la fois d'une bonne analyse des résultats et d'appliquer le concept de dominance (relation de dominance) une phase de normalisation est obligatoire. Cette phase est appelée la pondération des critères qui consiste à attribuer à chaque critère un poids justifiant ainsi son importance ; ainsi les critères les plus importants suivant les préférences du décideur auront les poids les plus élevés. Après la pondération une phase d'agrégation s'enchaine afin de générer une évaluation en appliquant l'une des trois approches d'agrégation (globale, partielle ou interactive).

6.3 La méthode d'agrégation « TOPSIS[6] »

Créée par Hwang et Yoon en 1981 [12], cette méthode se base sur la relation de dominance qui résulte de la distance par rapport à la solution idéale. Son fondement consiste à choisir une solution qui se rapproche le plus de la solution idéale (*i.e. la meilleure sur tous les critères*) et de s'éloigner le plus possible de la pire solution qui dégrade tous les critères. Cette méthode est composée de six (06) phases:

- Phase 1: Calcul de la matrice de décision normalisée, Les valeurs normalisées « e_{ij} » sont calculées comme suit :

[6] TOPSIS: Technique for Order by Similarity to Ideal Solution.

$$e'_{ij} = \frac{g_j(a_i)}{\sqrt{\sum_i^m [g_j(a_i)]^2}} \qquad \text{Avec:} \quad \begin{array}{l} i = 1, \ldots, m. \\ j = 1, \ldots, n. \end{array} \qquad \textbf{(3)}$$

Où : les « $g_j(a_i)$ » correspondent aux valeurs déterministes des scénarios « i » pour le critère « j ».

- Phase 2 : Calcul de la matrice de décision normalisée pondérée (*i.e.* Calculer le produit des performances normalisées par les coefficients d'importance relative des attributs). Les éléments de la matrice sont calculés comme suit :

$$e''_{ij} = \pi_j \,.\, e'_{ij} \qquad \text{Avec:} \quad \begin{array}{l} i = 1, \ldots, m. \\ j = 1, \ldots, n. \end{array} \qquad \textbf{(4)}$$

Où : «πj» est le poids du $j^{\text{ième}}$ critère et: $\sum_j^n \pi_j = 1$.

- Phase 3 : Détermination des solutions (profils) idéale (a^+) et des solutions anti-idéale (a _) par rapport à chaque critère.

$$\begin{array}{ll} a^+ = \{\underset{i}{\text{Max}}\, e''_{ij},\ i = 1, \ldots, m \text{ ; et } j = 1, \ldots, n\} \text{ ;} & \\ a^+ = \{e_j^+,\ j = 1, \ldots, n\} = \{e_1^+, e_2^+, \ldots, e_n^+\} \text{ ;} & e_j^+ = \underset{i}{\text{Max}}\, \{e''_{ij}\}. \\ a_- = \{\underset{i}{\text{Min}}\, e''_{ij},\ i = 1, \ldots, m \text{ ; et } j = 1, \ldots, n\} \text{ ;} & \\ a_- = \{e_{j-},\ j = 1, \ldots, n\} = \{e_{1-}, e_{2-}, \ldots, e_{n-}\} \text{ ;} & e_{j-} = \underset{i}{\text{Min}}\, \{e''_{ij}\}. \end{array} \qquad \textbf{(5)}$$

- Phase 4 : Calcul des mesures d'éloignements (*i.e. Calculer la distance euclidienne par rapport aux profils a^+ et a_*) ; L'éloignement entre les alternatives est mesuré par une distance euclidienne de dimension « n ». L'éloignement de l'alternative « i » par rapport à la solution idéale (a^+) et anti idéale (a_-), qui peut être assimilé à la mesure d'exposition aux risque et donné par :

$$D_i^+ = \sqrt{\sum_i^n (e''_{ij} - e_j^+)^2}. \qquad \text{Avec :} \quad i = 1,2, \ldots, m. \qquad \textbf{(6.1)}$$

$$D_{i_} = \sqrt{\sum_i^n (e''_{ij} - e_{j-})^2}. \qquad \text{Avec :} \quad i = 1,2, \ldots, m. \qquad \textbf{(6.2)}$$

- Phase 5 : Calculer un coefficient de mesure du rapprochement au profil idéal :

$$C_i^+ = \frac{D_{i-}}{D_i^+ + D_{i-}}. \qquad \text{Avec:} \quad \begin{array}{l} i = 1, \ldots, m \text{ ;} \\ 0 \le C_i^+ \le 1. \end{array} \qquad \textbf{(7)}$$

- Phase 6 : Rangement des scénarios suivant leur ordre de préférences (*i.e.* en fonction des valeurs décroissantes de C_i^+ ; *« i » est meilleur que «j » si $C_i^+ > C_j^+$*).

6.4 La méthode de pondération « Entropie »

Comme la méthode TOPSIS ne permet pas de générer de manière automatique les poids des critères, on n'était obligé d'intégrer une autre méthode pour pondérer les critères à l'intérieur de TOPSIS. Cette méthode de pondération est l'Entropie, dont le fondement et l'algorithme sont donnés comme suit :

La méthode Entropie [11] est une technique objective de pondération des critères, l'idée est qu'un critère « j » est d'autant plus important que la dispersion des évaluations des scénarios est importante. Ainsi les critères les plus importants sont ceux qui discriminent le plus entre les scénarios (dans notre cas ce sont les étiquettes). Les étapes de cette méthode sont données comme suite :

- Etape 1 : L'entropie d'un critère « j » est calculée par la formule :

$$E_j = -K \sum_{j}^{n} X_{ij} \log(X_{ij}). \qquad \textbf{(8)}$$

  Où :
  - K : est constante choisie de telle sorte que, pour tous « j », on a $0 \leq E_j \leq 1$, pour notre cas « K » est calculé comme suit :

$$K = \frac{1}{\text{Log}(n)}. \qquad \text{Avec: } n \text{ : le nombre de scénarios de désambiguïsation.} \qquad \textbf{(9)}$$

  - X_{ij} : l'évaluation du scénario « i » suivant le critère « j ».
- Etape 2 : L'entropie E_j est d'autant plus grande que les valeurs de « e_j » sont proches. Ainsi, les poids seront calculés en fonction de la mesure de dispersion (opposée de l'entropie) :

$$D_j = 1 - E_j. \qquad \text{Avec: } j = 1, \ldots, n. \qquad \textbf{(10)}$$

- Etape 3 : Les poids seront ensuite normalisés par :

$$W_j = \frac{D_j}{\sum_{j}^{n} D_j} \qquad \text{Avec: } j = 1, \ldots, n. \qquad \textbf{(11)}$$

7 Exemple illustrative

Afin de mieux cerner la solution proposée, nous allons garder la même démarche multicritère citée auparavant, tout en incluant les modifications et explications nécessaires :

Soit la phrase P= « **رَجَعَ المُغْتَرِبُ إلى الوَطَن.** », qui se trouve à l'entrée de notre analyseur. Après segmentation de la phrase en mots, l'analyse se fait sans problème pour les unités 2, 3 et 4, mais l'unité 1 « **رَجَـعَ** » présente un cas typique d'ambiguïté morphologique. Pour lever cette ambiguïté nous allons appliquer notre approche de désambiguïsation multicritère, selon la démarche suivante :

- ***Etape 1 :*** **Construire la liste des scénarios d'analyse :** Cette liste est obtenue directement après le processus d'analyse morphologique. Ce qui va générer l'ensemble « A ».

Verbe	Scénario (Schème)	longueur
رجـــع	فَـــعِـــلَ	6
	فَـــعِـــلَ	6
	فَـــعُـــلَ	6
	فُـــعِـــلَ	6

Tableau 1 : Exemple d'ambiguïté générée lors de l'analyse du verbe « رجَـــعَ ».

- ***Etape 2 :*** **Application des critères**

  Afin de construire une famille cohérente de critères F, nous proposons deux critères de base pour discriminer entre les scénarios d'analyse, à savoir :

  a) Critère 1 : Concordance des voyelles

  Ce critère va vérifier la concordance entre les voyelles de l'unité lexicale et les voyelles de chaque scénario candidat, de telle sorte que chaque concordance vaut : un (1).

  b) Critère 2 : La Fréquence d'apparition

  Ce critère s'appuie sur un calcul statistique fait sur la base d'un corpus annoté, de telle manière que le scénario qui se manifeste le plus souvent (une plus grande fréquence d'apparition) aura systématiquement le score le plus élevé. Chaque apparition vaut : un (1).

  Remarque :

  Le corpus utilisé est composé de plus de 300 unités réparties sur 10 paragraphes choisis arbitrairement à partir des livres scolaire de l'école algérienne.

- ***Etape 3 :*** **Application de la fonction d'évaluation**

  Pour les deux (02) critères, la fonction d'évaluation est l'addition (+), ainsi il s'agit de deux (02) critères à maximiser.

- ***Etape 4 :*** Générer le tableau (matrice d'évaluation) d'évaluation

➔ Scénario 🡻 Critère	S1«فَـــعَـــلَ »	S2«فَـــعِـــلَ»	S3«فَـــعُـــلَ»	S4«فُـــعِـــلَ»
Concordance des voyelles	3	2	2	1
Fréquence d'apparition	16	5	2	1

Tableau 2 : Tableau (Matrice) d'évaluation.

- ***Etape 5 :*** Agrégation des performances et pondération des critères.

  a) Normalisation du tableau d'évaluation

  En appliquant la formule (3) de la méthode TOPSIS.

➔ Scénario 🡻 Critère	S1«فَـعَـلَ»	S2«فَـعِـلَ»	S3«فَـعُـلَ»	S4«فُـعِـلَ»
Concordance des voyelles	0.71	0.47	0.47	0.24
Fréquence d'apparition	0.95	0.30	0.12	0.06

Tableau 3: Tableau d'évaluation normalisé

b) Pondération des critères.

Pour pondérer les critères nous utilisons la méthode Entropie. Le tableau suivant montre les valeurs de calcul de l'entropie (E_j), l'opposé de l'entropie (D_j) et normalisation des poids (W_j) des deux critères.

➔ Scénario 🡻 Critère	E	D	W
Concordance des voyelles	0.24	0.76	0.47
Fréquence d'apparition	0.15	0.85	0.53

Tableau 4: Pondération des critères

c) Pondération du tableau d'évaluation (normalisé)

Cette pondération est faite en appliquant la formule (4) de la méthode TOPSIS.

➔ Scénario 🡻 Critère	S1«فَـعَـلَ»	S2«فَـعِـلَ»	S3«فَـعُـلَ»	S4«فُـعِـلَ»
Concordance des voyelles	0.33	0.22	0.22	0.11
Fréquence d'apparition	0.50	0.16	0.06	0.03

Tableau 5: Tableau d'évaluation normalisé et pondéré.

d) Détermination de la solution idéale (D^+) et la solution anti idéale (D_-)

La solution idéale correspond à la valeur maximale et la solution anti idéale correspond à la valeur minimale.

	Solution idéale	Solution anti idéale
Concordance des voyelles	0.33	0.11
Fréquence d'apparition	0.50	0.03

Tableau 6: Détermination de la Solution idéale et anti idéale.

e) Calcul des mesures d'éloignements

Après application des formules (6.1) et (6.2) de la méthode TOPSIS on aura les différentes mesures d'éloignement de chaque scénario comme il est illustré dans le tableau suivant :

	S1«فَـعَـلَ»	S2«فَـعِـلَ»	S3«فَـعُـلَ»	S4«فُـعِـلَ»
D^+	0	0.36	0.45	0.52
D_-	0.52	0.17	0.14	0

Tableau 7: Mesures d'éloignements.

f) **Calcul des coefficients de mesure du rapprochement au profil idéal** : Pour calculer ces coefficients C_i^+, nous utilisons la formule (7) de la méthode TOPSIS, nous allons établir un classement de ces coefficients, selon un ordre décroissant et le scénario ayant obtenu le score le plus élevé sera élu. On aura les valeurs suivantes :

$$C_1^+=1 > \quad C_2^+=0,32 > \quad C_3^+=0,24 \quad > \quad C_4^+=0.$$

Selon notre méthode le scénario 1 « فَـعَـلَ » sera sélectionné par le système, donc les informations morphologiques suivantes seront générées.

	Information
Verbe	رَجَـعَ
Schème	فَـعَـلَ
Etiquette	VAA3PMSIA
Désignation en français	Verbe Accompli Actif 3e Personne Masculin Singulier Invariable Accusatif.
Désignation en arabe	فعل ماضي مبني للمعلوم للمفرد المذكر الغائب، مبني على الفتح.
Racine	رجع

Tableau 8: Informations générées de l'étiquetage du verbe « رجع ».

8 Conclusion

L'étiquetage morphosyntaxique est considéré aujourd'hui, comme étant la partie vitale dans n'importe quelle application de traitement automatique du langage naturel. De ce fait, la performance d'une application linguiciel[7] dépend directement de l'efficacité de

[7] Logiciel traitant la langue naturel sur les différents niveaux : morphologique, syntaxique, sémantique et pragmatique, utilisant des bases de connaissances linguistique.

son étiqueteur morphosyntaxique et pour obtenir une fiabilité irréprochable on doit s'intéresser à trois (03) points qui constituent la structure d'un étiqueteur, à savoir : une bonne phase de segmentation, une bonne organisation des unités lexicales de la langue et surtout un module de désambiguïsation fiable. C'est ce dernier point qui est le centre d'intérêt de notre étude. Nous avons essayé à travers cet article de mettre l'accent sur l'impact que l'ambiguïté sur le processus d'automatisation de la langue arabe et plus particulièrement l'ambiguïté morphologique, ainsi que les méthodes dédiées à la lever de cette forme d'ambiguïté, comme l'approche par contraintes et l'approche stochastique.

Mais notre principal objectif reste de présenter notre nouvelle approche de désambiguïsation fondée sur les techniques issues de l'aide multicritère à la décision.

Comme conclusion, qu'on peut tirer de cette étude est que l'approche multicritère peut être considérée comme une alternative de choix pour remplacer les deux (02) approches citées auparavant, afin de remédier au problème de l'ambiguïté morphologique qui constitue un point de passage qui condamne la réussite ou l'échec d'une application en T.A.L.A.

Références

1. Hoceini Y.: Un système d'analyse morphologique de la langue arabe, mémoire magister, école nationale supérieure d'informatique, (2002).
2. Baloul S. : Développement d'un système automatique de synthèse de la parole à partir du texte arabe standard voyellé, thèse de doctorat, université du Mans, (2003).
3. Tuerlinckx L.: La lemmatization de l'arabe non classique, U.C.L. Institut orientaliste, Belgique.
4. Vergne J., Giguet E.: Regards théorique sur le tagging, TALN, pp 22-31, (1998).
5. Lechel M. : Analyse et conception d'un correcteur grammatical libre pour le français, thèse de magister, université de Gronoble 3, (2005).
6. Chanod J. P., Tapanainen P. : Les étiqueteurs statistiques et les étiqueteurs par contraintes, (1995).
7. Al-Sulaiti L.: Deigning and Developing a Corpus of Contemporary Arabic, School of Computing, University of Leeds (UK), March (2004).
8. Merialdo B. : Modèle Probabiliste et Etiquetage Automatique, Institut EUROCOM, Sophia Antipolis.
9. Constant M. : Etiquetage morphosyntaxique probabiliste, Cours pour Master en informatique, Université Paris-Est Marne la Vallée, (2007).
10. Belguith L. H., Hamadou A. B. : Traitement des erreurs d'accord, RSTI-RIA, Volume 18, pp 679-707, (2004).
11. Vincke P. : L'aide multicritère à la décision, SMA, Université de Bruxelles, (1989).
12. Hwang C. R., Yoon K.: Lecture Notes in economics and Mathématical System, Springer Verlag Berlin Heidelberg, New York, (1981).