

La langue arabe et le TAL : étude de cas

Pr. BOUHADIBA Farouk
Université d'Oran

Résumé :

Cet article traite du problème de la description et l'analyse de l'aspect non-concaténatif (non-enchaînement) qui caractérise la langue arabe sur la base de certains modèles de concaténation qui ne reflètent pas nécessairement les processus de dérivation de cette langue. Par conséquent, des algorithmes basés sur la segmentation et la modélisation de la langue arabe selon ces modèles risqueraient de perturber le traitement automatique de cette langue.

La langue arabe appartient à la famille des langues dites sémitiques. Les historiens, tout comme – beaucoup plus récemment – les linguistes s'accordent à retracer ses origines à partir de langues très anciennes telles que l'Akkadien (3000 avant JC) qui était parlée dans la vieille Mésopotamie. D'autres langues telles que le Phénicien ou l'Ethiopien (l'Amharique) présentent des caractéristiques linguistiques plus ou moins similaires à celles de l'arabe. Néanmoins, l'Hébreu demeure la langue avec laquelle la langue arabe partage le plus de similitudes linguistiques, plus particulièrement sur les plans morpho-syntaxique et lexical. Parmi les particularités (spécificités) de la langue arabe, nous retiendrons dans le cadre de cette présentation son système graphique (écriture de droite à gauche), ses

sons (les gutturales et les emphatiques) et surtout sa morphologie (processus de formation d'un mot, affixation, reduplication et autres spécificités morphosyntaxiques telles que les temps, les modes et voix).

Pour l'histoire, nous noterons que les premières traces d'écriture de cette langue remontent au IV^{ème} siècle de notre ère. Celles-ci ont été relevées dans la péninsule arabique, mais l'on situe l'utilisation de cette langue dans les échanges tribaux dès le V^{ème} siècle. Beaucoup plus récemment, les linguistes partagent l'idée que la langue arabe (classique) que nous connaissons actuellement dérive en fait de deux parlers locaux d'Arabie, le Najd et le Hijaz.

L'expansion arabo-musulmane des VII^{ème}-XI^{ème} siècles a fait que cette langue, qui symbolisait alors un facteur d'intégration à la religion qu'elle véhicule, mais aussi un symbole d'unification des peuples «arabisés» se soit propagée bien au-delà des limites de sa sphère d'origine. Elle représente de nos jours la langue liturgique des musulmans en Turquie, en Iran, en Afghanistan, au Pakistan, en Indonésie, des peuples d'Afrique subsaharienne, et des peuples de l'Eurasie tels que ceux du Kazakhstan, du Kirghizistan, du Tadjikistan, du Turkménistan et de l'Ouzbékistan. Ses traces linguistiques beaucoup plus que religieuses sont apparentes entre autres en Espagnol et dans le Maltais.

Sur le plan sociolinguistique et dialectologique, la langue arabe nous apparaît comme constituant un continuum où l'on peut distinguer plusieurs variétés telles que d'un côté l'arabe classique, l'arabe moderne standard, l'arabe parlé éduqué et à l'autre extrémité du

continuum, les divers parlers et dialectes locaux qui rassemblent à plus ou moins fort dosage des traits de similitude avec ce que l'on appelle communément l'arabe classique, ou langue d'origine.

De par ses caractéristiques morpho phonologiques, morphosyntaxiques et parfois même lexicales, la langue arabe, tout comme les langues de cette même famille sémitique, est basée essentiellement sur une morphosyntaxe dite de non concaténation (non enchaînement) de ses morphèmes. Elle diffère dans ce sens des langues dites Indo-européennes (de morphologie essentiellement de concaténation) pour lesquelles plusieurs modèles théoriques de description sont disponibles (structuralisme, fonctionnalisme, générativisme, et plus récemment le modèle auto segmental). Elle demeure dans ce sens une langue assez complexe à gérer dans le domaine du Traitement Automatique des Langues. Parmi ces difficultés, nous citerons les problèmes de la représentation graphique de l'arabe sur ordinateur qui a été soulevé dès les années 70.

Certaines de ces difficultés ont trouvé solution puisqu'il est tout à fait possible de nos jours de procéder au traitement de texte en arabe, de faire appel à des dictionnaires et vérificateurs d'orthographe, voire même d'envoyer des messages SMS et autres sur le Web en arabe. Tout ceci est révélateur de grands pas franchis depuis les balbutiements des années 70 face à la complexité, voire même à l'impossibilité, de procéder au traitement de l'écrit de cette langue par ordinateur.

L'analyse par ordinateur de l'arabe oral a également connue des résultats probants et ce grâce aux différentes techniques de traitement du signal. Elle reste

cependant moins débattue de nos jours par rapport au débat actuel sur la modélisation du morphème de l'arabe dans le cadre du traitement automatique et de la traduction.

Nous essayerons de relever ci-dessous les différences linguistiques majeures de l'arabe par rapport aux langues qui ont pris un essor déterminant dans le traitement automatique et la traduction automatique.

Sur le plan de la représentation graphique d'un texte, la langue arabe dispose de vingt-huit segments, qui des labiales aux vélaires (b.... k, g) pourraient être latinisées, mais le problème demeure, dans une certaine mesure, dans la représentation des consonnes gutturales et emphatiques, dans la représentation de la gémation qui n'apparaît que rarement dans le texte (signe diacritique reconnu tacitement par le natif mais qui ne risque pas d'être reconnu, du moins aussi facilement, par la machine) et de l'allongement vocalique qui se fait dans le texte arabe en ajoutant des consonnes à la syllabe allongée. Là aussi, la machine devrait reconnaître dans quel contexte ces consonnes (alif, waw, ya) jouent-elles leur rôle de consonne dans la chaîne parlée ou écrite et dans quel autre contexte elles représentent des signes d'allongement vocalique.

Sur le plan de la morphosyntaxe, le mot en arabe ne représente pas nécessairement la même charpente, ni la même formation que le mot dans les langues Indo-européennes. En partant du principe général que cette langue est basée sur un système de racines (trilitères et quadrilitères essentiellement), celles-ci reposent sur un

sens général (en anglais kernel meaning) qui est construit en fonction de l'agencement des radicales, représentées par des consonnes seulement alors que la racine dans les langues indo-européennes est généralement à base syllabique, c'est-à-dire qu'il existe un agencement de consonnes et de voyelles qui constituent la racine dans les langues indo-européennes. Ainsi, pour la langue arabe, nous dirons que la racine <ʕʕl> réfère à la notion générale de « faire quelque chose, produire, entreprendre,... » alors que si l'on actionnait le processus d'agencement de cette racine en déplaçant ces radicales au sein de la racine, nous pourrions avoir <ʕlʕ> qui dénote par contre l'idée « de soigner, soulager, traiter... ». De même, la suite des radicales /ʕ, l, m/ dans cet ordre donne la racine <ʕlm> 'savoir, savant, ... » alors que la série /l, m, ʕ/ donne la racine <lmʕ> pour le sens « briller, illuminer, brillance, illumination, ... ». Il en est de même pour les racines quadrilitères telles que <Θwrt> « révolution, ... » au lieu de <Θrwt> « richesse,... ». Certes, ce ne sont pas n'importe quelles combinaisons de radicales qui vont automatiquement générer une racine nouvelle, donc un sens général (ou champs sémantique) nouveau. La langue dispose d'un certain nombre de contraintes imposées par la hiérarchie des sons en arabe. Parmi ces contraintes, nous citerons à titre d'exemple le fait qu'une labial ne peut être suivie d'une chuintante en début de mot (contraintes syllabiques).

La morphologie du mot devient encore plus complexe lorsque des processus de flexion et de

dérivation tels que l'affixation, l'infixation, la reduplication et autres sont déclenchés. Ainsi, alors que le natif ou néo-natif dans cette langue reconnaît relativement aisément un mot tel que « mustasmaḥun » comme étant composé du préfixe (instrumental) {mV} (où V représente une voyelle) suivi du préfixe {istV} qui devient infixe dans ce cas particulier, puis de la racine <smḥ> « pardonner, accepter, ... », le tout pour rendre le sens équivalent de 'permis' en français ou « allowed » en Anglais, la question demeure à savoir si la machine serait capable d'agir de la même façon que le locuteur. Cet exemple très simple pose de prime abord un problème non seulement de reconnaissance du mot par la machine pour lui trouver grâce à ses systèmes experts et ses filtres un équivalent en français ou en anglais, mais surtout cela nécessite une expérimentation soutenue pour arriver à une segmentation de l'arabe qui est loin d'être similaire à la segmentation des langues indo-européennes. Des propositions sont faites pour modéliser la segmentation du français ou de l'anglais et de l'adapter à l'arabe. C'est le cas du système de compilation morphologique MORPHE¹ et autres mais qui demeurent encore très récents et sujets à débat.

Si l'on était appelé à relever certaines difficultés dans la traduction automatique de l'arabe vers le français ou l'anglais, et en considérant les parties du discours les plus élémentaires telles que l'article, la préposition, la

¹. Compileur morphologique conçu pour traiter un type de morphologie de concaténation puis adapté à la langue arabe (cf. John R.R. Leavitt, 1994).

conjonction ou le pronom (les particules) ou même les catégories telles que l'adjectif, l'adverbe, le nom, ou le verbe, nous remarquerons d'emblée que la segmentation ne pourrait se faire de façon quasi identique comme ce serait le cas pour le français et l'espagnol et dans une moindre mesure pour le français et l'anglais. Nous discuterons ceux qui nous semblent les plus problématiques dans la reconnaissance des équivalents par rapport au français et à l'anglais, en l'occurrence la détermination / indétermination, la particule de coordination {waw} et la catégorie du nom, plus précisément les doublets sémantiques en arabe.

En règle générale, la langue arabe diffère du français et de l'anglais principalement sur le plan de sa structure syntaxique, de ses temps et modes. C'est une langue du type beaucoup plus VSO (Verbe-Sujet-Objet) que SVO. Elle n'a pas de forme passive, ou du moins elle a une construction du passif où l'agent n'est pas signalé. Elle fonctionne avec trois temps, le perfectif, l'imperfectif et le participe et cinq modes, l'indicatif, l'impératif, le subjonctif, le jussif et l'énergétique. En voici quelques exemples :

a) Structure VSO : kataba al kaatib kitaab

lit. Ecrivit l'écrivain un livre.

b) Forme passive : kutiba al kitaab

lit. Ecrivit (participe passé) le livre.

c) Le perfectif: katab(a)
qui fait fonction en même temps d'infinitif (Il a écrit / écrire)

L'imperfectif:	jaktub(u)	Il écrit
Le participe :	kutib(a)	Ecrit
L'indicatif :	taktub(u)	Tu écris
L'impératif:	aktub	Ecris !
Le subjonctif :	utilisé avec des particules	

comme dans : {lazam, ʕalajka an, ...}

lazam taktub, ʕalajka an taktub lit. Il faut tu écris
(Il faut que tu écrives)

Le jussif: utilisé dans des constructions avec
la particule de négation {lam} lam aktub lit. pas j'écris

L'énergétique: utilisé pour exprimer la fermeté
dans l'intention : saktubu / taktubanna (rendus souvent
par un impératif en français).

La langue arabe ne dispose pas d'une forme
d'infinitif telle que conçue en français ou en anglais. La
troisième personne du masculin singulier du perfectif sert
de forme verbale de base pour la conjugaison, comme
dans {kataba} 'Ecrire' ; {tarʕama} 'Traduire'

Pour revenir aux particules simples et leur
utilisation dans la langue par rapport au français et à
l'anglais, il faudrait signaler que l'article défini en arabe
est unique et invariable {al}. Placé en début de mot
auquel il est annexé, il dénote en fait la détermination par
opposition à l'indétermination qui n'est pas marquée
comme cela est le cas pour l'anglais. Par contre, le
nombre est déterminé par des alternances vocaliques au

sein du schème alors que le genre pour ainsi dire est marqué par la ‘ta marbuTa’² ou marque du féminin singulier en arabe. Il est possible, néanmoins, de trouver en arabe des noms masculins avec une ‘ta marbuTa’ tels que Hamzatun (lire /ħamzatun/ (pour Hamza, le nom d’un chevalier arabe ou chevalier du désert). Le féminin pluriel étant constitué par un allongement vocalique suivi d’une {ta}. Cela ne poserait en apparence pas de problème d’équivalence entre l’arabe et le français. Cela réduirait même les difficultés puisqu’il serait possible de regrouper les articles (le, la, les) du français et leur attribuer l’article {al} moyennant l’utilisation des mécanismes d’alternance pour le genre et le nombre. Ainsi, nous aurons :

Le livre {al kitaab}, la bibliothèque {al maktabatu}
(où l’allongement vocalique est représenté par V₁V₁)

Les livres {al kutub(u)}, les bibliothèques {al maktabaat(u)} (où la voyelle finale (v) représente la forme pausale en arabe)

A l’indéfini, nous aurons :

Un livre {kitaab(un)}, une bibliothèque {maktabat(un)} (où t n’est pas prononcé [maktaba])

Des livres {kutub(un)}, des bibliothèques {maktabaat(un)} (où t est prononcé [maktabaat])

² . ta marbuTa dans ses deux formes : isolée : ﺕ ou bien reliée: ﺕﺍ

En faisant abstraction des cas assez fréquents de pluriels irréguliers en arabe auxquels il est fait appel à d'autres mécanismes de la langue, la question des équivalences concernant la détermination / l'indétermination entre l'arabe et le français ne paraît pas soulever de problèmes majeurs dans des cas de formations isolées. Le problème est tout autre lorsque cet article se trouve dans une structure syntaxique, représentée par une phrase telle que //mamnuuʔ al tadχiin// qui serait équivalente à 'Interdit de fumer' ou au classique « No smoking » de l'anglais où l'on constate une absence d'article dans la traduction française ou anglaise. L'article défini de l'arabe ne peut être l'équivalent de 'le, la, les' dans certains cas et l'on est obligé de faire appel à une particule en français ou en anglais. Ainsi, //kayfa al ĥaal // serait traduite par « Comment vas-tu / allez vous ? » ou bien « How are you ? » entre autres que par *« comment le temps » que nous retrouvons dans certaines versions de traducteurs automatiques sur le marché.

La particule de coordination « *wāw* » est également source d'ambiguïté dans la traduction de l'arabe vers d'autres langues. Elle n'est pas forcément traduite par un équivalent tel que le 'et' français ou bien par le 'and' anglais. Elle permet de relier des phrases sans pour autant impliquer une relation logique entre les phrases en question. La fréquence de son usage en arabe ainsi que les différents sens qu'elle recouvre sont source de problème en traduction. Par conséquent, il est parfois difficile de lui attribuer un équivalent (en termes de

coordination) dans la langue française ou la langue anglaise. Les grammairiens (arabes et persans) ont attribué au « *wāw* » de l'arabe plusieurs fonctions telles que '*Wāw al 9aTf*' ou *wāw* de l'adjonction qui permet de relier deux ou plusieurs unités grammaticales au niveau inter-phrastique et extra-phrastique, tout en exprimant l'addition, l'opposition ou bien la conséquence. Elle est généralement utilisée dans l'exposition ou la narration. '*Wāw el Hāl*' ou *wāw* qui exprime un état, une condition ou une circonstance. Elle introduit soit une phrase nominale soit une phrase verbale telles que {*ja'a aHmed wa shshamsu Tāli9atun*} Lit. « Ahmed venait au moment du lever du soleil », ou bien {*ja'a aHmed wa qad Tala9ati shshamsu*}, Lit. « Ahmed venait alors que le soleil était levé ». '*Wāw al 'isti'nāf*' exprime une opposition. Elle est généralement utilisée pour exprimer deux actions contradictoires ou bien des conseils. Citons comme exemple {*la ta'kul wa tatakallam*} Lit. « Ne manges pas (alors que tu) parles ». '*Wāw al ma9iyya*' exprime la juxtaposition, telle que {*masha zaydun waTTarīq*} (Wright, 1981 : 83), Lit. « Zaïd marchait le long de la route ». '*Wāw al qasam*' est utilisée pour jurer telle que {*wallah al 9aDhīm*} Lit. « Par Allah, le Puissant ». '*Wāw rub*', connue sous le nom de *waw rubba* (Wright, 1981 : 216), telle que {*wa ka'sin sharibtu*} Lit. « Plusieurs verres ai-je bus ! ». '*Wāw al faSl*' sert à distinguer des nom tels que {*9Amru / 9Umar*}, Lit. « Amr / Omar ».

Steingrass, F. (1978: 191) donne plusieurs équivalents anglais à *wāw* : and (et), also (aussi), then (puis, ensuite), while (pendant que, alors que), at the

same time (au même moment), together with (avec), but (mais), however (cependant), et dans le sens de *rubba* : frequently (fréquemment), sometimes (parfois), perhaps (peut-être). Il ne s'agit pas seulement de trouver des équivalents, mais il s'agirait beaucoup plus de reconnaître le type de structure dans laquelle *wāw* se trouve et d'identifier le rôle de cette copule au sein de la phrase ou bien entre les phrases.

A titre indicatif, dans des structures syndétiques de base de la langue arabe, telles que: {*jalastu wa jalasati ssayidatu warda*} (Gibrān Khalil Gibrān, 1949 : 112), la coordination par *wāw* exprime la simultanéité mais surtout l'indépendance des deux actions régies par deux acteurs séparés (Moi par rapport à Warda). Dans ce cas simple, le *wāw* peut être traduit par la copule de coordination « et » du français pour donner une traduction telle que : « Nous nous sommes assis Madame Warda **et** moi ». Cependant, une structure similaire contenant *wāw* exprimant la simultanéité et l'indépendance telle que {*in tarakūhu halaka wa halakū*} (Mustafa Sādiq al Rafī'i, 1951 : 6, 5) ne peut se traduire en français par « s'ils le laissent (l'abandonnent), il périra **et** ils périront », mais beaucoup plus par « s'ils l'abandonnent ils périront tous » où l'on observe l'absence du « et » dans la traduction vers le français. En anglais, ce serait la copule « so » qui rendrait la traduction vers l'anglais, pour donner : « If they leave him, he will perish and so they will ». *Wāw* exprime la simultanéité en arabe là où le français et l'anglais utilisent d'autres constructions grammaticales sans le « et » ou le « and ». Une phrase arabe telle que {*qāTa9aha wa SāHa*} (MuSTafa LuTfi al-ManfalūTi,

1952 : 236) ne peut se traduire en français par « * Il l'a interrompue **et** il a crié » ou en anglais par « *He interrupted her **and** he shouted » mais par « Il l'a interrompue tout en criant » et « He interrupted her, shouting ... » respectivement. Nous remarquons que l'anglais utilise une virgule pour indiquer la simultanéité dans les actions là où le français a recours à la double préposition « tout en ».

Dans la langue arabe, les structures asyndétiques (absence de coordination) servent à mettre en relief l'indépendance entre les phrases tout en exprimant une complémentarité des phrases en question. Ces constructions asyndétiques sont fréquentes dans la description d'un état de fait. Elles sont parfois source d'ambiguïté lors de leur traduction en français ou en anglais. Elles nécessitent parfois l'utilisation de structures syndétiques dans la langue cible où il est fait appel à la coordination. Elles sont traduites parfois soit en élaborant des constructions plus complexes dans la langue cible, soit en faisant jouer la ponctuation de la langue cible. Ainsi une phrase telle que {*innama huwa bnu akhī hāshim qadimtu bihi min yathrib*} (Muhammed Husain Haikel, 1952 : 98) serait traduite en français par « C'est mon neveu Hashim que j'ai ramené de Yathrib ». La langue anglaise va jouer sur la ponctuation pour traduire cette phrase arabe. Ce qui donnerait : « He is my nephew Hashim, I brought him from Yathrib ». Une telle phrase risque de poser un problème de sens qui est, naturellement, dicté par la structure syntaxique, mais qui ne peut être « fidèlement » représenté dans la langue cible (français / anglais) que si le traducteur maîtrise le sens dans la langue source. Cette phrase prête à confusion si

elle est mal interprétée du fait que le nom {*hāshim*} est source d'ambiguïté lors de la traduction. Si elle est traduite en français par «*C'est le fils de mon frère Hashim, je l'ai ramené de Yathrib » ou même «*C'est le fils de mon frère Hashim que j'ai ramené de Yathrib » l'ambiguïté demeure en ce sens que non seulement il n'est pas clair qui est Hashim ? Le père ou le fils ? Mais aussi qui a-t-on ramené de Yathrib ? Le père ou le fils ? L'ambiguïté demeure en anglais : « He is my brother's son Hashim whom I brought from Yathrib ». Néanmoins, elle paraît moindre par rapport au français dans la mesure où le cas possessif de l'anglais réfère au possesseur et non pas à ce qui est possédé. Par conséquent, à la question « Qui est Hashim ? », la réponse ne peut être que « c'est le père ». Par contre, le relatif « whom » prête à confusion quant à la question « Qui a-t-on ramené de Yathrib ? » Le père ou le fils ? Il n'en demeure pas moins que telle qu'elle est construite en arabe, {*hāshim*} ne peut être que le nom du frère (le père) et non pas celui du neveu (le fils). Il s'agirait de savoir si la machine pourrait déceler de telles ambiguïtés dues non pas à la langue arabe mais à sa traduction vers d'autres langues.

Au niveau du mot, la langue arabe utilise de façon très productive la synonymie. Des mots et parfois même des propositions entières sont doublées. Ce phénomène est reconnu par les traducteurs sous l'appellation « doublets sémantiques » ou plus communément comme la « répétition sémantique ». Ainsi, des phrases telles que :

بصورة مستمرة متواصلة

1) bi Suratin mustamiratin mutawaaSilatin

... قادتنا الى كارثة لا تعد ولا تحصى

2) qaadatna ila kaarithatin la tuḡadd wa la tuḥSa

ne peuvent être traduites en français par :

1. de façon continuellement continue
2. qui ne peut être ni estimée ni évaluée

Ce problème d'équivalence entre les langues a déjà fait l'objet de plusieurs analyses depuis l'article d'Eugen Nida (1964) "*Towards a Science of Translation*". La notion d'**équivalence dynamique** permettrait alors au traducteur de contourner le sens au détriment de la syntaxe du texte source ou du texte cible, de l'idiome, du contexte et même du genre dans le texte. Si cela paraît faisable pour l'homme, il n'en demeure pas moins que la machine reste incapable, à notre connaissance, de gérer ce type d'équivalence.

La répétition sémantique ne concerne pas seulement le nom. Elle peut se produire dans des énoncés tels que « al istiḡsaa wa ttaḥlil » (investigation et analyse), des adjectives « mustamiratin mutawaaSilatin » (continuellement continue) des verbes « jadhiḡuhu wa yadhiluhu » (il l'étonne et le surprend), des adverbes (wazimatin muktaibatin) (silencieusement et sans

bruit)³... Plusieurs stratégies de la traduction permettent au traducteur de gérer ce type d'équivalence entre l'arabe et d'autres langues. Nous citerons par exemple le cas de la transposition grammaticale comme dans « mustamirratin mutawaaSilatin » où il faudrait remplacer ce doublet arabe par un seul concept en français (la troncation) pour aboutir à « continue », ou bien dans des cas tels que « al quijaam wal aḫlaqijjaat » (les valeurs morales) où un doublet nominal est remplacé par un doublet adjectif-nom en français⁴. De même, le traducteur peut faire appel à la technique de la distanciation sémantique dans des cas de doublets sémantiques. Ainsi, une phrase telle que « yudhiṣuḥu wa yudhiluhu » où les verbes adhaṣa / adhala sont très proches en arabe, sémantiquement parlant, (étonner / surprendre en français) serait traduite par la paire de verbes (étonner / alarmer) en français où la distance sémantique entre (étonner / alarmer) est beaucoup plus grande que celle qui existe entre « adhaṣa / adhala » de l'arabe. La question qui se pose serait de savoir si la machine pourrait procéder à de telles stratégies de la traduction.

Nous avons essayé de relever dans cet article quelques cas de difficultés que pourrait rencontrer la

³. Adapté à partir de Dickins and Watson 1999. Standard Arabic: an advanced Course. Cambridge: Cambridge University Press), pp. : 541-53

⁴. Pour plus de détails voir James Dickins, Sandor Hervey and Ian Higgins. Thinking Arabic Translation : a course in translation method: Arabic to English. Routledge London and New York, 2002, pp.61-62

machine (parfois même le traducteur non averti !) lors du passage d'un sens de l'arabe vers des langues telles que le français ou l'anglais. Certes, notre vision est beaucoup plus celle d'un linguiste que celle d'un expert en TAL ou en TA. De ce fait, il est possible que certains aspects ou cas que nous avons relevés comme problématiques aient déjà trouvé solution sous d'autres cieux et dans d'autres études de cas. Il n'en demeure pas moins que l'apport du linguiste concernant la langue reste toujours un apport non négligeable à l'informaticien qui doit reformuler les solutions du linguiste en algorithmes et autres procédés que la machine accepte avant de produire les résultats escomptés dans le TAL ou la TA.

Références :

- McCARTHY, J. 1979. *Formal Problems in Semitic Phonology and morphology*. Doctoral Dissertation, MIT, Cambridge, Massachusetts.
- McCARTHY, J. 1981. A Prosodic Theory of Nonconcatenative morphology'. In *Linguistic Inquiry*, vol. 12:373.
- DICKINS, J., HERVEY, S., HIGGINS, I. 2002. *Thinking Arabic Translation: a course in translation method: Arabic to English*. Routledge, London and New York.
- LEAVITT, John, R.R. 1994 : MORPHE : A practical Compiler for Reversible Morphology.
Center for Machine translation Carnegie Mellon University, Pittsburg, USA
- LEAVITT, John, R.R. 1994: MORPHE: a Morphological Rule Compiler. Technical report, CMU-CMT-94, Memo.
- SOUDI, A.et CAVALLI-SFORSA, V. 2001 Interfacing an Arabic Morphology and Sentence Generation System with an English-to-Arabic Knowledge-based MachineTranslation System. In CNRST, PROTARS III, USA.
- WRIGHT, R. 1981. *A Grammar of the Arabic Language*, 3rd Edition, CUP. London.