

***NOUVELLES RESSOURCES ET NOUVELLES PRATIQUES PEDAGOGIQUES  
AVEC LES OUTILS TAL***

---

**Mourad Mars**

*Laboratoire LIDILEM, université Stendhal de Grenoble, France  
& laboratoire UTIC, Université de Monastir, Tunisie*  
[mourad.mars@e.u-grenoble3.fr](mailto:mourad.mars@e.u-grenoble3.fr)

**Georges Antoniadis**

*Laboratoire LIDILEM, université Stendhal de Grenoble, France*  
[georges.antoniadis@u-grenoble3.fr](mailto:georges.antoniadis@u-grenoble3.fr)

**Mounir Zrigui**

*Laboratoire UTIC, université de Monastir, Tunisie*  
[mounir.zrigui@fsm.rnu.tn](mailto:mounir.zrigui@fsm.rnu.tn)

**Résumé** : Nous proposons dans cette communication de présenter un analyseur morphologique de la langue arabe réalisé au sein du laboratoire LIDILEM. Nous abordons par la suite le problème de la désambiguïsation grammaticale en présentant quelques méthodes utilisées. La dernière partie est consacrée à la réponse à la question suivante : Comment on peut réellement intégrer un tel outil dans une plateforme d'apprentissage des langues pour permettre aux enseignants la création automatique de plusieurs types d'activités pédagogiques ?

**Summary**: We propose in this communication to present a morphological analyzer for Arabic language realized within the LIDILEM laboratory, we present thereafter the problem of the grammatical classification by presenting some methods used. The last part will be devoted to answer the following question: How can we integrate such a tool in a platform of training of the languages to allow to the teachers the automatic creation of several types of pedagogic activities?

**Mots clés** : TAL, TIC, langue arabe, analyse morphologique, étiquetage, désambiguïsation.

# Nouvelles ressources et nouvelles pratiques pédagogiques

## avec les outils TAL

L'introduction des TIC (Technologies de l'Information et de la Communication) dans les dispositifs d'apprentissage des langues semble soumise à la nécessité de suivre au plus près les progrès techniques. Bien qu'une transformation et une amélioration du rendement aient toujours affleuré, l'observation minutieuse des multiples dispositifs et logiciels que nous avons pu effectuer révèle des failles, que seule l'utilisation des outils issus du TAL (Traitement Automatique des Langues) est capable de corriger, tout en permettant de nourrir et d'élargir la problématique des pédagogues, enseignants des langues. Dans ce contexte, une question se pose : comment les systèmes d'ALAO (Apprentissage des Langues Assisté par Ordinateur), et d'une manière générale les EIAH (Environnements Informatiques pour l'Apprentissage Humain), peuvent-ils tirer profit des procédures, solutions et outils du traitement automatique de la langue ? Certaines plateformes dédiées à l'apprentissage / enseignement des langues, comme la plateforme MIRTO, apportent des réponses à cette question. Développé au sein du laboratoire LIDILEM de Grenoble, MIRTO permet la création d'activités pédagogiques quasi-instantanément, en exploitant les possibilités des procédures et outils du TAL. MIRTO peut générer des activités destinées aux apprenants pour l'apprentissage de plusieurs langues.

Nous présentons au cours de cette communication un analyseur morphologique de l'arabe et nous montrerons comment on peut réellement intégrer un tel outil dans une plateforme d'apprentissage des langues afin de permettre aux enseignants la création automatique de plusieurs types d'activités pédagogiques.

### 1 – PRESENTATION DE L'ANALYSE MORPHOLOGIQUE DE L'ARABE ET DE L'ANALYSEUR REALISE

Si l'apport du traitement automatique de la langue (TAL) pour l'élaboration de systèmes d'apprentissage des langues n'est plus à démontrer, cette élaboration demande des outils TAL performants, robustes et à grande couverture langagière. Pour l'arabe, la grande

majorité de ces outils reste encore à créer. Le travail que nous présentons concerne la création d'un analyseur morphologique pour la langue arabe. Il doit être intégré dans la plateforme MIRTO développé par le laboratoire LIDILEM à l'Université Stendhal de Grenoble. L'analyseur servira à la création automatique et semi-automatique d'activités pédagogiques pour l'apprentissage de l'arabe.

L'analyse morphologique de l'arabe s'intéresse, comme pour les autres langues, aux formes des mots. Pour l'arabe, vu la richesse du mot graphique et la présence d'agglutinations, l'opération est délicate. Dans ce qui suit, nous présentons quelques travaux dans ce domaine et nous décrivons le principe de l'analyseur morphologique réalisé en exposant les différentes étapes d'analyse et les ressources utilisées.

#### 1.1 - Des travaux du domaine

- L'analyseur morphologique de Shaalan:

(Shaalan, 1989) Il s'agit d'un système fondé sur des règles écrites en SICStus Prolog et a besoin d'une certaine expérience en Prolog qui est difficile à réaliser par des linguistes. Ce système utilise un ancien système de translittération.

- Les travaux de Ahmed (Caire, 00):

(Ahmed, 2000) il décrit un modèle hybride pour l'analyse morphologique de l'arabe: Morph3 peut être considéré comme une démonstration de la manière dont une base de connaissance des règles et une base de connaissances statistiques peuvent être combinées pour la résolution des problèmes en morphologie"

- L'analyseur Morphologique de Sakhr: [www.sakhr.com](http://www.sakhr.com)

Le Multi-Mode Morphological Processor (MMMP) de Sakhr est un analyseur synthétiseur morphologique, qui fournit l'analyse de base pour tout mot arabe. Cet analyseur couvre toute la langue arabe moderne et classique. L'analyseur identifie toutes les bases possibles d'un mot, c-à-d, trouver sa forme de base après l'extraction des suffixes et préfixes. Malheu-

reusement, pas de version d'essai pour cet analyseur.

Pour chaque mot en entrée, cet analyseur donne toutes les solutions possibles. Il ne s'intéresse pas à la phase de désambiguïsation.

- L'analyseur Morphologique de Tim BUCKWALTER "ARAMORPH" :

ARAMORPH, un analyseur morphologique réalisé par Tim buckwalter, est téléchargeable à partir du site de LDC à l'adresse <http://www.nongnu.org/aramorph/french/>. Le texte en entrée doit être translittéré en ASCII avant tout traitement, et le résultat doit être reconverti en Arabe pour que ça soit compréhensible. Ce système ne permet pas l'analyse des textes contenant des chiffres 0..9.

- L'étiqueteur APT de Khoja:

APT, 'Arabic Part-of-speech Tagger' (Khoja 2001, Khoja et al 2003), utilise une combinaison de données statistiques et des règles techniques. Selon eux, une telle combinaison donne des résultats encourageants. Les étiquettes de l'APT sont essentiellement dérivées du système d'étiquetage BNC de l'anglais, elles ont subies quelques modifications pour répondre aux contraintes de la grammaire arabe. Le nombre d'étiquettes utilisées est 131. Un corpus de 50000 mots du journal saoudien Al-Jaziira a été utilisé pour entraîner le système.

- Freeman: Adaptation de la méthode d'Eric Brill sur la langue arabe:

(Freeman 2001 et 2002) décrit un autre étiqueteur développé pour l'arabe. Il est disponible sur [Arabic-l@byu.edu](mailto:Arabic-l@byu.edu), basé sur l'étiqueteur d'Eric Brill (Brill, 1993). Un système d'apprentissage qui peut être entraîné sur un corpus déjà étiqueté. Il utilise 146 étiquettes pour étiqueter des lexèmes.

- L'analyseur morphologique de XEROX par Beesley:

Pour l'analyseur morphologique de Kenneth Beesley, (Beesley 96) les bases de données (manuellement) en collaboration avec Lauri Karttunen. Cet analyseur utilise le « finite state technologie : FST » développée a XEROX. Une version d'essai se trouve sur le site de XEROX

[www.xrce.xerox.com/research/mltt/arabic](http://www.xrce.xerox.com/research/mltt/arabic).

- Le système de Maamouri et Cieri :

Le LDC (Language Data Consortium) est en train d'élaborer un POS (Part of speesh) pour étiqueter l'arabe. Cet étiqueteur est basé sur l'étiquetage automatique d'un corpus conte-

nant 734 fichiers de « l'Agence France Presse ». Il utilise l'analyseur morphologique développé par Tim Buckwalter. Ce système a été développé par Maeda Kazuaki et Hubert Jin.

- L'analyseur morphologique Sebawai de Darwish (Darwish, 2002):

Sebawai est un analyseur morphologique pour la langue arabe développé par Darwish en 2003 en un seul jour !! Le système permet de trouver les racines de mots avec un taux de réussite de 84%.

### 1.3 - Des dictionnaires pour l'analyse morphologique de l'arabe

Une des étapes principales de la réalisation d'un analyser morphologique pour l'arabe est la conception et l'organisation d'une base de données linguistiques. La base de données réalisée inclut la majorité des primitives linguistiques d'arabe comme les verbes, les noms, les particules, les pré-bases et les post-bases.

#### 1- Dictionnaire de base.

Dans ce dictionnaire, nous avons trois catégories de mots :

Le verbe : entité exprimant un sens dépendant du temps, c'est un élément fondamental auquel se rattachent directement ou indirectement les divers mots qui constituent l'ensemble.

Le nom : l'élément désignant un être ou un objet qui exprime un sens indépendant du temps.

Les particules : entités qui servent à situer les événements et les objets par rapport au temps et l'espace, et autorisent une séquence cohérente du texte.

Pour chaque élément du dictionnaire, on lui associe son schème, et le sens de ce dernier.

#### 2- Dictionnaires de pré-bases et de post-bases.

Les pré-bases sont le résultat de la concaténation des proclitiques et des préfixes compatibles entre eux. Pour réaliser ce dictionnaire, nous utilisons un ensemble des règles de compatibilité telle que :

« Le proclitique ﻝ /al/ ne se combine jamais avec un préfixe ».

Le nombre total des pré-bases valides est de 88.

Exemple :

ال التعريف + ك التشبيه	كال
kaf attachbih + al attaarif	kal

Les post-bases sont au nombre de 233 ; elles se lient après la base. Elles sont obtenues par la combinaison de suffixe et enclitique(s). Les compatibilités dépendent des pronoms décrits par chacune des particules.

Exemple :

ماض+مؤ+مف	ت
Madhi + Mouannath+ Moufrad	Ta

#### 1.4 - Base des règles et notion de classe morphologique

Cette partie est l'une des étapes principales de la réalisation d'un analyseur morphologique de l'arabe. Il s'agit de la conception et l'organisation d'une base de règles linguistiques pour la langue arabe. Cette base de données doit comporter la plupart des règles linguistiques de l'arabe pour éviter les possibilités d'étiquetage du mot qui sont théoriquement possibles mais inexistantes dans la langue arabe.

##### Méthodologie

Une base des règles grammaticales de l'arabe consiste à regrouper le maximum possible des « formules » grammaticales existantes dans la langue arabe sous la forme :

{Pre-base + Etq Base + Post-base} = Combinaison valide;

Or, cette organisation primaire nous ramène à un nombre d'entrées très grand. Pour y remédier, nous proposons une organisation particulière basée sur le concept de la classe morphologique.

« Chaque classe morphologique représente une famille de données linguistiques de même nature et de même fonctionnement morphologique ».

##### 1- Classes morphologiques

L'idée de cette organisation consiste à regrouper :

Les pré-bases qui ont les mêmes caractéristiques morphologiques et les mêmes propriétés.

Les post-bases qui ont les mêmes caractéristiques morphologiques et les mêmes propriétés.

Ce regroupement nous permet de réaliser 22 classes dans la base de pré-bases, et 30 classes dans la base de post-bases.

Exemples de classes morphologiques de type pré-bases :

La première classe : Pré-base = {vide} Elle a les caractéristiques suivantes :

Conclusion 1.1 : Compatible avec les trois ensembles (verbes, nom et particules).

Con 2.1 : Compatible avec toutes les classes morphologiques de post-base.

$C1Pré = C 1.1 \& C 2.1$  avec C1Pré : La première classe morphologique de type pré-base ;

La deuxième classe : Pré-base = {أن / an, أنت / at, أي / ay, أأ / aa} Elle a les caractéristiques morphologiques suivantes :

Con 1.2 : N'est compatible qu'avec l'ensemble des verbes.

Con 2.3 : Compatible avec les classes morphologiques de post-base qui ne sont pas compatibles qu'avec le verbe.

$C2Pré = C 1.2 \& C 2.3$

La troisième classe : Pré-base = {ال / al, فال / fâl, كال / kal, وال / wal, وبال / wabil}

C'est l'ensemble des pré-bases qui se terminent par ال / al. Elle a les caractéristiques suivantes :

Con 1.3 : N'est compatible qu'avec l'ensemble des noms.

Con 2.4 : N'est compatible qu'avec la classe pronom attaché de type post-bases.

$C3Pré = C 1.3 \& C 2.4$

Classes morphologiques de type post-bases :

La première classe : Post-base = {vide} Elle a les caractéristiques suivantes :

Con 1.1 : Compatible avec les trois ensembles (verbes, nom et particules).

Con 2.2 : Compatible avec toutes les classes morphologiques de pré-base.

C1Post = C 1.1 & C 2.2 avec C1Post est la première classe morphologique de type post-base ;

La deuxième classe : Post-base = {ين / iin, ن / n, ان / an, ون / un} c'est la classe morphologique de flexion verbale exprimant le présent de type post-base.

Cette classe a les caractéristiques morphologiques suivantes :

Con 1.2 : Compatible qu'avec l'ensemble des verbes.

Con 2.5.1 : Compatible avec les classes morphologiques de flexion verbale qu'expriment le présent ou le futur de type pré-base.

C2Post = Con 1.2 & Con 2.5.1 avec  
Con 2.5.1 = Con 2.5 || Con 2.6

Avec:

Con 2.5 : Classe morphologique de flexion verbale qu'exprime le présent de type pré-base;

Con 2.6 : Classe morphologique de flexion verbale qu'exprime le futur de type pré-base ;

On constate que :

La classe morphologique des flexions verbales du présent de type pré-base et la classe morphologique des flexions verbales du passé de type post-base sont compatibles avec les verbes, mais pas entre elles.

La classe morphologique des flexions verbales du présent de type pré-base et la classe morphologique des flexions verbales du présent de type post-base sont compatibles avec les verbes et entre elles.

## 2- Conception de la base des règles.

Comme nous l'avons mentionné auparavant, notre base de données se compose d'un ensemble de classes, chacune représentant une famille de données linguistiques de même nature et de mêmes caractéristiques morphologiques. Ainsi, nous introduisons la notion de classes des règles qui représentent toutes les concaténations possibles entre les différentes composantes morphologiques définies dans les classes morphologiques. L'objectif de l'utilisation de ce type de classes est, d'une part, d'éviter la redondance, et d'autre part, de réduire le nombre des règles de concaténation.

Comme la morphologie arabe se compose essentiellement de trois sous ensembles : le nom,

le verbe, les particules, nous avons organisé l'ensemble de classes en cinq classes de règles qui contiennent toutes les classes morphologiques :

- La classe des règles verbale, nominale et particules.

- La classe des règles purement verbale.

- La classe des règles purement nominale.

- La classe des règles verbales et nominales.

- La classe des règles particules et nominales.

Exemples des règles :

- La classe des règles verbale, nominale et particules :

{Première classe de type pré-base + base = verbe ou nom ou particule + première classe de type post-base} = valide;

R1 = C1Pré + (Verbe || Nom || Particule) + C1Post ;

R1 = C 1.1 & C 1.2 + (Verbe || Nom || Particule) + C 1.1 & C 2.2;

- La classe des règles purement verbale :

{Deuxième classe de type pré-base + base: verbe + deuxième classe de type post-base} = valide ;

R2 = C2Pré+ Verbe + C2Post ;

R2 = C 1.2 & C 2.3 + Verbe + Con 1.2 & Con 2.5.1 ;

## 1.5 – Autres ressources

La suppression de la pré-base et de la post-base pour récupérer la base ne suffit pas, puisque la concaténation de certaines pré-bases et post-bases avec la base engendre des transformations qui doivent rester réversibles. Lors du découpage de chaque mot à analyser, nous devons être en mesure de restituer la bonne forme des lexèmes, ce qui nous a amené à réaliser une base des règles de transformations. Dans ce qui suit nous présentant quelques règles :

Si la pré-base se termine par ال/al et si la première lettre de la base est une lettre solaire de l'arabe, alors le troisième caractère de la base est forcément la gémination «<» ; dans ce cas, il suffit de le supprimer et on retrouve la base que l'on doit valider.

Remarque : Les lettres solaires de l'arabe sont :  
د / ص / ض / ذ / ط / ن / ت / ث / س / ش / ل / ر / ز.  
Exemple : الشمس /Alchams/ ⇒ ال+شمس ⇒  
ال+شمس.

Si la pré-base contient ل /le/ et la première lettre de la base est la gémination «ّ», il faut la remplacer par la lettre ل /l/.

Remarque : La pré-base لل /le/ est obtenue par la concaténation des particules ل /l/ et ال /al/.

Exemple : للعب /lella''eb/ ⇒ ل + عّب ⇒  
ال+لعب.

## 2 – LA DESAMBIGUISATION GRAMMATICALE

Dans cette partie, nous présentons l'étape qui suit l'étiquetage grammatical, la désambiguïsation, qui consiste à éliminer les solutions morphologiques parasites, en déterminant celle qui est pertinente localement.

La complexité de cette tâche pour la langue arabe n'est plus à démontrer vu le problème d'agglutination et la non voyellation des textes. Quelques statistiques seront présentées dans cette section.

Nous verrons par la suite les différentes méthodes utilisées pour lever l'ambiguïté et attribuer une seule étiquette à chaque mot du texte.

### 2.1 – Problématique de l'ambiguïté et quelques statistiques

En analyse morphologique, le principal problème à résoudre est l'ambiguïté. Il existe différents types d'ambiguïtés. D'abord, les mots peuvent être ambigus aux niveaux lexical et grammatical. Le mot " ذهب " est ambigu lexicalement. Il peut désigner l'or en français ou encore le verbe aller. " كاتب ", quant à lui, est ambigu grammaticalement. Il peut appartenir à plusieurs catégories grammaticales différentes; اسم فاعل, اسم فعل, Le sens de ce mot sera très différent selon sa catégorie, nom : « écrivain » ou verbe : « écrit » ; il peut appartenir à quatre catégories grammaticales différentes. Le nombre de catégories auxquelles un mot peut appartenir dépend du jeu d'étiquettes choisies. En moyenne, le nombre d'étiquettes par unité lexicale (UL) voyellé est de 9, alors que pour les unités lexicales voyellées elle peut atteindre 12 étiquettes par unité lexicale.

Un autre facteur, pour la langue arabe, est le nombre d'unités lexicales ambiguës dans un texte. Contrairement au français et à l'anglais, il peut concerner 66% des unités lexicales qui composent le texte [Debili].

### 2.2 – Méthodes de désambiguïsation

Les textes en langues sémitiques comme l'arabe standard moderne, sont basés sur des systèmes d'écriture permettant la concaténation de différents morphèmes. Ces morphèmes, peuvent appartenir à différentes classes ; leur enchaînement, délimité par des espaces, forme des unités textuelles, qui sont communément appelés *mots*. Par conséquent, la tâche d'étiquetage pour les langues sémitiques se compose d'une tâche de segmentation et d'une tâche de classification. Fondamentalement, les mots peuvent être segmentés en différentes séquences des morphèmes. Pour chaque segmentation, les morphèmes peuvent être ambigus. L'absence de vocalisation augmente l'ambiguïté.

Un problème essentiel lors de la désambiguïsation est la manière d'adapter pour l'arabe les méthodes existantes en tenant compte du fait que les corpus étiquetés pour la langue arabe sont relativement rares. Nous allons présenter dans cette section quelques méthodes de désambiguïsation existantes.

L'utilisation du Modèle de Markov Caché (MMC) pour la segmentation et l'étiquetage de l'arabe (désambiguïsation) était discuté dans des travaux précédents (Shamsi et al, ..). Les MMC peuvent être appliqués sur des corpus non étiquetés en utilisant l'algorithme Baum-Gallois (Baum, 1972). Divers approches ont été suggérées pour l'étiquetage et la segmentation en utilisant le MMC, les plus notable sont : l'utilisation de l'étiquette du mot (unité lexicale), qui déterminent uniquement la segmentation et l'étiquette de chaque morphème; l'utilisation d'un modèle de Markov bidimensionnel avec les étiquettes des morphèmes.

Lee et al (2003) décrivent un système de segmentation de mots pour l'arabe qui emploie un modèle de langue de n-gram. Ce système est basé sur un modèle de langage et un lexique dérivé d'un corpus segmenté manuellement.

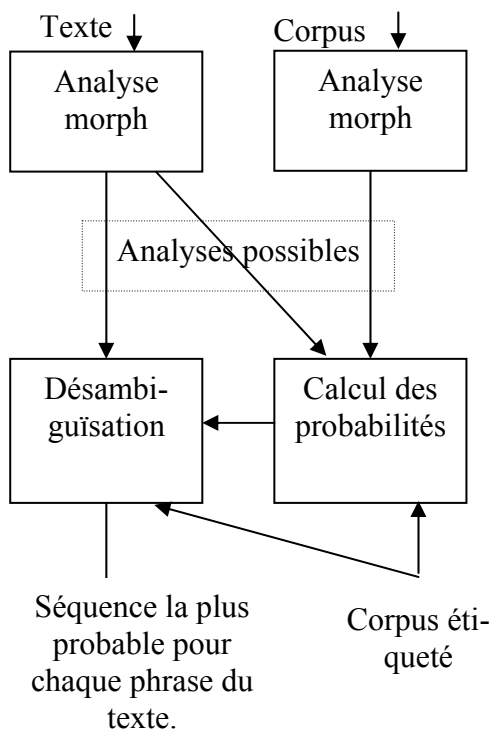
Diab et al, (2004) utilisent SVM (Support Vector Machines) pour les tâches de segmentation des mots et de l'étiquetage. Ils utilisent un jeu

de 24 étiquettes. L'étiquetage a été appliqué aux mots segmentés en utilisant une segmentation à partir du corpus annoté (Mona Diab).

D'autres approches pour l'étiquetage ont été développées dans le passé, y compris l'étiquetage à base des règles (Brill, 95), le modèle entropie maximal (Ratnaparkhi,96) etc.

Toutes ces approches exigent une grande masse de données annotées (pour un étiquetage supervisé) ou un lexique qui inscrit toutes les étiquettes possibles pour chaque mot (pour un étiquetage non supervisé).

Le schéma ci-après donne l'architecture d'un système d'analyse morphologique de textes arabes. Le système comporte trois modules : l'analyseur morphologique permettant d'assigner à chaque unité lexicale une ou plusieurs étiquettes, le calcul des probabilités qui permet d'assigner une probabilité à chaque séquence d'étiquettes potentielle, et le module de désambiguïsation permettant de calculer la séquence d'étiquettes la plus probable pour chacune des séquences du texte à analyser.



### 3 – INTEGRATION DES OUTILS TAL DANS L'APPRENTISSAGE DES LANGUES ET LA GENERATION DES ACTIVITES PEDAGOGIQUES

Dans cette partie, nous abordons le thème de la génération des activités pédagogiques pour l'enseignement de l'arabe langue étrangère, pour lesquels divers développements des travaux en cours sont attendus. Dans ce contexte, notre approche peut s'exprimer par : comment utiliser un système informatique, dans notre cas un analyseur morphologique de l'arabe, qui soit réellement intégré dans des systèmes d'ALAO / EIAH (Apprentissage des Langues Assisté par Ordinateur/ Environnement Informatique d'Apprentissage Humain) et qui permet la création des activités pédagogiques et quel genre d'activités peut-il générer?

#### 3.1 – Le cas MIRTO pour la langue française et d'autres langues

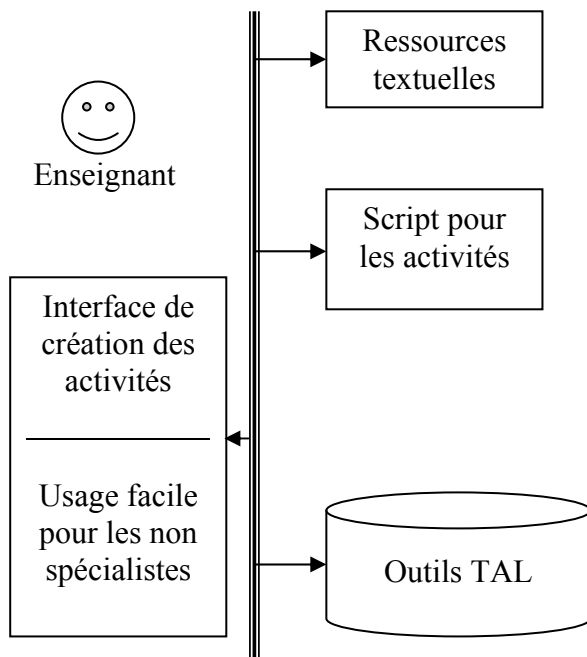
La plateforme MIRTO (Multi-apprentissages interactifs par des recherches sur des textes et l'oral)) utilise des fonctionnalités du TAL qui paraissent prometteuses pour l'enseignement des langues et de la linguistique. Son principal avantage est la simplicité d'utilisation ; il permet à l'enseignant des langues, non spécialiste de TAL, de constituer des ressources linguistiques (à l'aide d'interfaces simples) et de les paramétrer afin de constituer des activités pédagogiques destinés aux apprenants. Pour développer une activité pédagogique à l'aide de MIRTO, il n'est donc pas nécessaire de recourir à une chaîne de traitement complexe, ni de faire appel à des spécialistes d'informatique ou de TAL. Cela garantit une cohérence entre les traitements linguistiques et les exploitations pédagogiques qui en sont faites.

#### 3.2 – Intégration des outils TAL dans des systèmes d'ALAO/EIAH

Les outils TAL qui seront intégrés dans toute plateforme d'apprentissage des langues doivent rester invisibles. Cette transparence permettra aux non spécialistes de bénéficier des résultats de ces outils. Une interface facile à utiliser permet aux enseignants (non informaticiens) de lancer des requêtes pour interroger la base de données de son environnement d'apprentissage et les outils TAL qu'il utilise pour créer des applications et des activités pour les apprenants.

Dans notre cas, l'analyseur morphologique de l'arabe permettra, à court terme, de générer des activités pour l'apprentissage de l'arabe langue

étrangère, un peu à la façon du générateur d'activités de MIRTO [Antoniadis04]. Une activité est donc la mise en œuvre d'un objectif pédagogique minimal, comme travailler sur une notion grammaticale particulière, réviser des conjugaisons etc. Elle se réalise par la création d'un espace de travail pour l'apprenant, lui permettant ainsi d'atteindre le but visé. Elle correspond à ce qu'on désigne traditionnellement par un exercice.



### 3.3 – Résultats attendus

Etant donné un texte arabe fourni en entrée, les techniques et les outils TAL permettront de sélectionner des mots et des expressions en fonction de critères choisis par l'enseignant via des interfaces faciles à utiliser. Sur cette base on peut créer un large éventail d'activités de type exercices lacunaires, textes à trous, conjugaison...

## 4 - CONCLUSION

Enfin, de façon plus originale, les outils ordinaires du TAL peuvent être détournés et manipulés pour proposer une approche réflexive dans l'apprentissage des langues et de la linguistique. Cependant, malgré son intérêt évident, le TAL ne s'adapte pas de façon immédiate à l'apprentissage / enseignement des langues et de la linguistique. La première diffi-

culté est tout d'abord liée aux performances imparfaites des outils de TAL qui doivent de ce fait être utilisés avec circonspection dans les applications pédagogiques. En effet, si quelques applications simples (comme l'étiquetage morphologique) donnent de bons résultats, elles ne sont jamais exemptes d'erreurs, ce qui est gênant dans un contexte d'enseignement. Deux options sont alors possibles : a) proposer un post-traitement des outils TAL, b) proposer une supervision par l'enseignant des outils de TAL.

## BIBLIOGRAPHIE

Andreas Stolcke. (2002). "SRILM-an extensible language modeling toolkit". ICSLP Denver Colorado. Sept 2002.

Antoniadis, G., Echinard, S., Kraif, O., Lebarbé, T., Loiseau, M. & Ponton, C. (2005). "Modélisation de l'intégration de ressources TAL pour l'apprentissage des langues : la plateforme MIRTO". Apprentissage des langues et systèmes d'information et de communication (AL-SIC). vol8.

Antoniadis, G. (2004). "Les logiciels d'apprentissage des langues peuvent-ils ignorer le TAL ?". Les cahiers de l'APLIUT, n° XXIII vol. 2, June 2004. pp 81-97.

Beesley K (1996). "Arabic Finite-State Morphological Analysis and Generation". Proceedings of COLING-1996, pages 89– 94, Copenhagen.

Darwish, K (2002). "Building a Shallow Arabic Morphological Analyzer in One Day", Proceedings of the Workshop on Computational Approaches to Semitic Languages, 2002, pages 47–54.

Eric Atwell, Latifa Al-Sulaiti, Saleh Al-Osaimi, Bayan Abu Shawar. (2004). "Un Examen d'Outils pour l'Analyse de Corpus Arabes", JEP-TALN 04, Arabic Language Processing, Fès, 19-22 April 2004.

Freeman. (2001). « Brill's POS tagger and a morphology parser for Arabic ». ACL/EACL-2001 Toulouse, France.

G. Kiraz. (1994). « Multi-tape two-level morphology: A case study in semitic non-linear morphology ». COLING'94, volume 1, pp 180–186, 1994.



Kraif, O. (2003). "Propositions pour l'intégration d'outils TAL aux dispositifs informatisés d'apprentissage des langues". *Lidil*, n° 28. pp. 53-165.

M. Kay. (1987). « Non-concatenative finite-state morphology ». 3<sup>ème</sup> conférence ECACL, pp 2–10, Copenhagen, Denmark.

Mourad MARS, Mohamed BELGACEM (2006), "Developed of a morphological analyser for arabic language, tool for creation of educational activities of training of Arabic", Workshop "TEL in working context", 13-15 November 2006, Grenoble, France. 2006.

Mourad MARS, Mounir ZRIGUI, Geroges ANTONIADIS, (2007). "Analyse morphologique de l'arabe", ICTA2007, Avril 2007, Hammamet, TUNISIE. 2007.

Mourad MARS, Mohamed BELGACEM, Mounir ZRIGUI, Geroges ANTONIADIS, (2007). "Analyseur morphologique de l'arabe", CITALA2007, 18-19 juin 2007, Rabat, Maroc. 2007.

Mona Diab, Kadri Hacioglu, Daniel Jurafsky (2004). "Automatic Tagging of Arabic Text: From Raw Text to Base Phrase Chunks", HLT-NAACL 2004. PP 149-152.

S. Khoja. (2001). « APT: Arabic part-of-speech tagger ». Workshop NAACL-2001.

Silberztein, M., Tutin, A. (2005). " NooJ, un outil TAL pour l'enseignement des langues. Application pour l'étude de la morphologie lexicale en FLE". *Apprentissage des langues et systèmes d'information et de communication (ALSIC)*. vol. 8.