

SALA: SYSTÈME D'APPRENTISSAGE DE LANGUE ARABE

MARAOUI Mohsen* and AYADI Rami **

* Laboratoire de linguistique et didactique des langues étrangères et maternelles,
Université Stendhal, 1180 avenue Centrale - BP 25 - 38040 Grenoble cedex 9, France,
Email: maraoui_m7@yahoo.fr

** Institut ISIM-Sfax & laboratoire UTIC, Email: ayadi.rami@planet.tn

Résumé—Après plus que deux décennies depuis le début des travaux, l'avancée de recherches dans le domaine de l'ALAO basé sur le TAL reste insuffisante, à cause de deux facteurs principale : la méconnaissance du TAL de la part des didacticiens des langues, voire des informaticiens, et le coût des ressources et produits issus du traitement automatique de la langue. Pour cela il n'existe que un certain nombre limités de prototypes ou de systèmes expérimentaux pour les langues latines [1], [2], [3], [4]. Les travaux d'ALAO basé sur le TAL pour la langue arabe sont pratiquement inexistant, malgré une riche bibliographie concernant le traitement automatique de l'arabe. En plus de facteurs cités ultérieurement, la carence concernant la langue arabe dans ce domaine est dû au fait qu'elle une langue difficile à traiter automatiquement [5], [6].

L'objectif de cet article est la présentation d'un système d'apprentissage de l'arabe langues étrangères pour des apprenants francophones, développé au sein de laboratoire LIDILEM de Grenoble en utilisant des ressources linguistiques ainsi que d'outils issus du TAL.

Mots-clés—EALAO (Enseignement et Apprentissage de Langues Assistés par Ordinateur) ; TAL (Traitement Automatique de Langue) ; langue arabe.

I. INTRODUCTION

Les environnements informatiques pour l'apprentissage humain (EIAH) sont utilisés dans des nombreux domaines, tels que la conduite automobile ou la formation des pompiers, mais aussi pour l'enseignement et l'apprentissage de langues. Sauf que ce dernier pour qu'il soit efficace, doit nécessairement intégrer des outils TAL.

Il existe pas mal de logiciels d'EALAO sur Internet, Il s'agit le plus souvent d'exercices à trous ou de QCM, conçus à l'aide de systèmes auteurs comme (*Course builder*, *Hot Potatoes* ou *Netquizz*). Ce type d'activités pose plusieurs problèmes [7] comme la rigidité des logiciels (les données utilisés sont prédéfinies et elles ne peuvent être ni modifiées ni enrichies) et la non adaptabilité des parcours aux compétences linguistiques des apprenants (le cheminement est indépendant de ses réponses à chaque étape, faute de pouvoir les évaluer). En effet, afin de permettre à l'apprenant de travailler en autonomie, il faut être capable de détecter, d'expliquer et

de corriger automatiquement ses fautes. Mais les logiciels actuels n'offre pas pour autant une réelle personnalisation et adaptation à l'apprenant car les corrections sont préétablies (comparer la réponse donnée par l'apprenant avec celle en mémoire et dans le meilleur des cas afficher des explications préenregistrées). L'unique avantage de ce genre d'exercices par rapport à ceux rédigés sur papier est l'interactivité de l'ordinateur. L'utilisation du TAL pour la conception de logiciels d'EALAO, est actuellement la seule que puisse résoudre ce problèmes [8].

En fonction de cette situation, et désireux d'enrichir les possibilités de création d'activités pédagogiques pour l'arabe, nous avons créé dans un premier temps un dictionnaire étiqueté de la langue arabe (le plus complet possible), un dérivateur, un conjugueur ainsi qu'un analyseur morphologique de mots arabes [9], [10] et [11]. Dans un deuxième temps, nous avons exploité ces outils pour créer un bon nombre d'applications pédagogiques pour l'apprentissage de l'arabe comme langue étrangère pour des apprenants francophones en utilisant notre système SALA. Dans ce qui suit nous allons décrire en détail les caractéristiques de notre système.

II. NIVEAU D'INTÉGRATION DU TAL DANS SALA

Pour l'apprentissage des langues, la manière avec laquelle on utilise le TAL dépend de but des tâches à réaliser par les apprenants. Dans notre cas, c'est-à-dire l'apprentissage des formes de la langue arabe (niveau morpho-lexical), le TAL peut intervenir en deux niveaux :

- Création des activités par l'enseignant de langue, en se basant sur des scripts décrivant des exercices et associant à ces exercices les fonctions TAL nécessitées pour construire l'activité (étiquetage morphologique par exemple). Les activités créées peuvent s'agencer selon des scénarii créés aussi par l'enseignant comme sur la plate-forme MIRTO (**M**ulti-apprentissages **I**nteractifs par des **R**echerches sur des **T**extes et l'**O**ral), [8].

- Expertiser les réponses des apprenants aux exercices, en délivrant les erreurs de celui-ci d'un point de vue lexical et morphologique. Ce rôle se réalise à travers des outils TAL dédiées à des micro-tâches [12], et qui résolvent des points particuliers de lexique ou de

morphologie, l'outil TAL fournit un cadre d'analyse de productions au contenu relativement prédéfini. L'élément traitement de la langue peut être utilisé pour montrer directement les erreurs grammaticales aux élèves comme dans le système FreeText [13] ou German Tutor [14]. Les problèmes de ces systèmes se trouvent au niveau des retours qu'on doit donner à l'apprenant sur ses erreurs, en particulier dans le cas des erreurs multiples dans une réponse, la question étant de savoir quelles erreurs mettre en avant.

Pour notre système, nous avons adopté l'intégration du TAL au premier niveau car il nous semble plus intéressant, et en conservant la possibilité d'utiliser par la suite les outils TAL au deuxième niveau comme des éléments d'aide partiels mais significatifs sur la forme des mots (en indiquant le schème, la racine etc.) de manière à faciliter l'inférence et à orienter l'apprenant vers la bonne solution.

III. ARCHITECTURE DU SALA

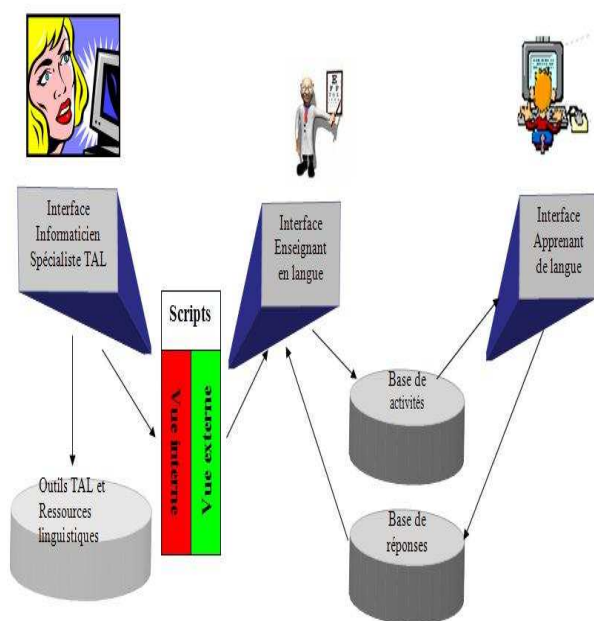


Figure 1. Architecture du SALA

A. Les interfaces

- L'interface Informaticien Spécialiste TAL permet l'accès à l'ensemble des outils TAL et les ressources linguistiques et à la vue interne de scripts. elle est utile pour la mise à jour des outils, des ressources et des scripts. Elle permet l'intégration des méthodes récemment découvertes par des informaticiens (autre que celui qui a conçu le logiciel), pour avoir un outil évolutif et réutilisable.

- L'interface enseignant en langue est une interface simple et facile à maîtriser (aucune

connaissance informatique n'est exigée), qui permet d'accéder à la vue externe de scripts pour créer des activités, et d'accéder à la base de erreurs pour pouvoir améliorer les activités créer. Partant du principe qu'un enseignant de langues n'est, a priori, ni spécialiste en informatique, ni en TAL, les paramètres de chaque activité doivent rester simples (trop de paramètres pourrait décourager l'utilisateur), et le paramétrage déclaratif (l'utilisateur n'est pas censé maîtriser un formalisme complexe ; la définition des paramètres doit rester intuitive) et convivial (les informations importantes doivent être facilement appréhendables).

Mise à part l'élaboration d'interfaces appropriées, une telle approche demande de bannir tout terme spécialisé (informatique ou TAL) du paramétrage des activités et de n'utiliser que des termes qui ont directement un sens pour les enseignants des langues, des termes issus de leur problématique et en accord avec leurs compétences.

- L'interface Apprenant de langue permet de choisir le niveau de difficulté et de répondre aux questions des activités.

B. Les scripts (vue interne)

Les scripts sont des modules qui intègrent des ressources ou des traitements TAL et qui ont chacun un objectif pédagogique générale. Par exemple, la génération automatique d'exercices de conjugaison est considérée comme un script, qui doit être paramétré ensuite par l'enseignant selon le besoin pour arriver à un objectif pédagogique bien précis. La création de scripts demande des compétences aussi bien en informatique qu'en TAL. Les programmes TAL utilisés à ce niveau sont purement de notre réalisation personnel. Etant donné la nature technique des scripts, le niveau interne n'est pas visible par les enseignants de langues. Chaque script ainsi créé est identifié par un nom qui doit suggérer les exploitations pédagogiques potentielles.

C. Les activités (vue externe de scripts)

D'un niveau purement didactique, la conception d'activités est une tâche opérée par les enseignants des langues, à l'aide d'un interface spécifique. Une activité correspond à ce qui est traditionnellement désigné comme exercice donné à l'apprenant pour lui permettre d'atteindre un but visé. La génération d'activités consiste ainsi à l'application d'un contexte didactique à un script. Une activité est un objet didactique qui peut être utilisé, par les apprenants, d'une manière autonome.

La première étape concerne le choix du script à mettre en œuvre, dans une liste de scripts disponibles. Un script pour l'enseignant est une « boîte noire » qu'il doit paramétrer.

Une fois le script choisi, l'enseignant doit le paramétrer, en fonction des paramètres accessibles, sous une forme simplifiée, à travers un tableau de bord, dont

les commandes doivent être configurées pour un usage didactique. Pour l'exemple d'exercices de conjugaison, l'enseignant doit spécifier la personne, le temps, le genre, le nombre, le mode de conjugaison, le type de racine de verbe et/ou le type de schème de verbe. Puis l'enseignant doit saisir la consigne correspondant à l'activité et indiquer, si l'évaluation automatique est possible, les rectifications éventuelles en fonction des résultats de l'évaluation.

Chaque activité créée doit être validée par son auteur qui doit vérifier qu'aucune erreur éventuelle n'y figure. L'apparition d'une erreur dans une activité créée a pour origine la qualité (degré de justesse) des résultats TAL qui ont été utilisés pour son élaboration. Les erreurs sont possibles, car les résultats des traitements du TAL n'étant pas toujours fiables à 100% surtout pour langue arabe.

Le paramétrage et l'évaluation automatique d'une activité exigent une attention particulière, et en cas d'erreur, l'enseignant apporte une correction manuelle.

TABLE I.
OPERATIONS ET UTILISATEURS

Niveau	Opération	Utilisateur
Scripts	Conception	Informaticien Spécialiste TAL
	Utilisation	Enseignant des langues
Activités	Conception	Enseignant des langues
	Utilisation	Apprenant

D. Les bases

- Base des activités contient des suites des applications, réalisés par l'enseignant en langue pour atteindre des buts et peuvent être changés selon les erreurs commises par les apprenants.
- Base de réponses contient l'enregistrement de toutes les traces produites par les apprenants y compris les erreurs commises. Elle peut être consultée par l'enseignant pour qu'il puisse contrôler le niveau des apprenants et améliorer l'apprentissage en proposant d'autres activités plus adéquates.

IV. ACTIVITÉS D'APPRENTISSAGES

L'objet de notre travail est d'étudier les exercices qu'il est possible de générer automatiquement à partir de ressources lexicales (dictionnaire) et des outils TAL d'analyse morphologique (dérivateur, conjugeur et étiqueteur).

Les exercices que nous cherchons à générer sont des exercices non contextuels, c'est-à-dire des activités dans lesquelles le vocable n'est pas entouré de son contexte.

L'intérêt de la génération automatique ou semi-automatique réside dans le fait que les exercices peuvent porter potentiellement sur tout le lexique et peuvent être refaits avec de nouvelles épreuves, sans intervention (manuelle) de la part des enseignants de langue. En outre, le système possède des informations sur les énoncés qui, couplées avec des ressources lexicales, lui permettent de retourner un diagnostic sur la réponse un peu plus élaboré que le classique vrai ou faux. Par ailleurs, en dotant le système d'une « mémoire », en occurrence une base de réponses dans laquelle sont consignées les annotations des apprenants et le détails de leurs résultats, on parvient à une meilleure gestion des énoncés en évitant, par exemple, de reposer une épreuve qui a déjà été résolue avec succès. Ainsi un tel environnement permet une plus grande individualisation de l'apprentissage.

Les exercices sont essentiellement des tests qui ne font que vérifier la maîtrise par l'apprenant des vocables en jeu et de leurs propriétés morphologiques. Dans ce que suit voici quelques types des exemples d'activités pédagogiques utilisées dans la génération des exercices .

A. Activités de familiarisation

Il s'agit des applications de l'identification de lettres et de mots . Le type d'exercices privilégié ici est le repérage qui permet de percevoir dans toute leur diversité les différentes tailles de caractères, les variantes typographiques et les principales calligraphies et qui sera la clef de l'autonomie de l'apprenant et de son investissement personnel dans l'apprentissage de la langue.

B. Activités de reproduction

Il s'agit des exercices d'entraînement à l'usage de la langue. L' apprenant doit s'entraîner à copier correctement des mots (activité de copie), car en arabe il faut, non seulement se familiariser avec une écriture différente, mais encore, apprendre à former correctement les lettres, en évitant, par exemple, de confondre en un même graphisme les lettres ز « za » et ر « ra ». Les activités en reproduction peuvent également faire appel à la mémorisation. L' apprenant doit pouvoir mémoriser un ou plusieurs mots, puis les réécrire (activité d'autodictée).

C. Activités de compréhension

C'est le type d'activités le plus important. Les relations entre morphologie et lexique font que les mots doivent être appris avec toutes les informations qui en permettent le réemploi: singulier et pluriel pour les noms; pour les verbes de la forme simple: accompli, inaccompli, etc.

Ces mêmes relations, font des dictionnaires arabes traditionnellement organisés par racines, un excellent support pour l'apprentissage des structures lexicales. La capacité de recherche de lien que constituent la racine ou la forme communes entre plusieurs unités lexicales.

VI. CONCLUSION ET PERSPECTIVES

La découverte de ce système doit se faire en liaison étroite avec l'apprentissage des régularités morphologiques, première étape de l'appréhension et de l'élucidation progressive du système des racines et des schèmes. C'est pourquoi l'on exclut, dès ce niveau, le recours à la notation systématique des voyelles brèves, dont la présence occulte la perception d'ensemble de l'unité graphique concernée, et qui ne sont nullement nécessaires à la mise en place d'une capacité en lecture oralisée. En effet, les linguistes arabes confirment que :

« le recours systématique à ces signes, qui pourrait sembler à certains être une aide pédagogique légitime pour des débutants, n'a pour effet que d'occulter la réalité de l'enjeu. Le travail sur la langue ne s'en trouve nullement facilité, il est purement et simplement supprimé. Or, l'effort particulier demandé à l'apprenant, pour l'interprétation de la valeur morphologique d'un mot, n'est rien de plus que la découverte, puis la mise en oeuvre permanente de la connaissance du système de la langue. Loin d'être un handicap, cette démarche est celle-là même qui lui permet d'apprendre, de comprendre et de mémoriser : c'est elle qui forge sa compétence linguistique et sa capacité à progresser, autorise une progression rapide, et assure la qualité de la langue qu'il saura plus tard utiliser. Cet effort supplémentaire consenti en début d'apprentissage est une condition essentielle de son développement ultérieur. L'acte de lecture est tout entier inscrit dans cette capacité, qui ne se confond en aucun cas avec la connaissance de la signification lexicale des mots. C'est elle aussi qui garantira que la production écrite des débutants s'incarne dans une graphie correcte et non pas dans une projection aberrante de lettres associées au hasard ».

V. STRUCTURATION DE L'APPRENTISSAGE

La logique de la progression est normalement celle qui va du plus simple au plus complexe, mais la structuration de l'apprentissage linguistique n'implique nullement que la présentation d'activités d'apprentissage soit déterminée seulement par cette logique de niveaux de difficultés. Si tel était le cas, l'enseignement se limiterait à des situations et des contenus stéréotypés, artificiels et pauvres. Une progression programmée ne peut, en effet, coïncider avec l'ensemble des besoins de communication à un moment donné. Dans un apprentissage où entre nécessairement une part de simulation et d'artifice, il faut veiller en permanence à rechercher des situations et des contenus qui suscitent des échanges naturels dans lesquels les apprenants puissent se sentir acteurs de ce qu'ils écrivent. Un équilibre est donc sans cesse à trouver entre le souci de structuration de l'apprentissage et la communication avec l'apprenant, à défaut de quoi la motivation de ce dernier ne saurait être longtemps maintenue. Il faut faire en sorte que la structuration soit effectué en réponse à un besoin d'apprentissage linguistique, suscité par des activités qui permettent de communiquer le plus possible avec l'apprenant.

L'étape suivante consiste à faire en sorte que le système puisse établir un diagnostic « évolué » en discernant des fautes de frappe ou en détectant des réponses approchantes. Le système doit également donner des éléments d'aide partiels mais significatifs sur la forme des mots (schème, racine etc.) de manière à faciliter l'inférence et à aiguiller l'apprenant vers la bonne solution. Dès lors, ces exercices ne seront plus de simples tests mais de véritables activités pédagogiques favorisant l'apprentissage lexical de l'arabe.

REFERENCES

- [1] Selva T. (2002). "Génération Automatique d'Exercices Contextuels de Vocabulaire". TALN 2002 pp 185-194.
- [2] Brun C., Parmentier T., Sandor A., Segond F. (2002). "Les outils de TAL au service de la e-formation en langues". Multilinguisme et traitement de l'information, (dir. Segond F.). Paris : Hermès. pp. 223-250.
- [3] Antoniadis G., Echinard S., Kraif O., Lebarbé T., Ponton C. (2005). "Modélisation de l'intégration de ressources TAL pour l'apprentissage des langues : la plateforme MIRTO". ALSIC, vol. 8, n° 2, novembre 2005, pp. 65-79
- [4] Antoniadis G., Chanier, T. (eds.), (2005). Numéro thématique « TAL et apprentissage des langues », ALSIC, vol. 8, n° 2, novembre 2005
- [5] Aljlal M., Frieder O. (2002). "On Arabic Search: Improving the Retrieval Effectiveness via a Light Stemming Approach", In 11th International Conference on Information and Knowledge Management (CIKM), November 2002, Virginia (USA), pp.340-347.
- [6] Larkey L. S., Ballesteros L. et Connell M. (2002). "Improving Stemming for Arabic Information Retrieval: Light Stemming and Cooccurrence Analysis", In Proceedings of the 25th Annual International Conference on Research and Development in Information Retrieval (SIGIR 2002), Tampere, Finland, August 2002, pp. 275-282.
- [7] Antoniadis G., Loiseau M., Ponton C. (2005). "Pedagogical text indexation and exploitation for language learning". Third International Conference on Multimedia and ICTs in Education, 7-10 juin 2005, Cáceres, Espagne. In Recent research developments in learning technologies, A. M. Vilas & al. (eds.), Bandajoz FORMATEX, pp. 984-994
- [8] Antoniadis G. (2004). "Les logiciels d'apprentissage des langues peuvent-ils ignorer le TAL ?". Les cahiers de l'APLIUT, n° XXIII vol. 2, juin 2004. pp. 81-97.
- [9] Maraoui M., Antoniadis G. et Zrigui M. (2006). "Multi-layer perception system for automatic tagging of the Arab words". IICALL 2006, December 17, The Ohio State University, USA
- [10] Maraoui M., Antoniadis G. et Zrigui M. (2007). " Un système de génération automatique de dictionnaires étiquetés de l'arabe". CITALA 2007, 18-19 juin 2007, Rabat (Maroc)
- [11] Maraoui M., (2008) « Un système de génération et étiquetage automatique de dictionnaires linguistiques de l'arabe » TALN-RECITAL 2009.
- [12] Demaiziere F. et Narcy-Combes J.P. (2005), « le défi de la dénativisation: apport des TIC », Actes de la conférence Untele Compiègne, France.
- [13] L'Haire S. et Vandeventer Faltin A (2003), Diagnostic d'erreurs dans le projet FreeText, Alsic, vol. 3, n° 2, pp. 21-37.
- [14] Heift T. (2003), "Multiple Learners Error and Feedback: A Challenge for ICALL Systems", CALICO, vol. 20, n° 3, pp. 553-548.