



HAL
open science

Modélisation de l'intégration de ressources TAL pour l'apprentissage des langues : la plateforme MIRTO

Georges Antoniadis, Sandra Echinard, Olivier Kraif, Thomas Lebarbé, Claude
Ponton

► **To cite this version:**

Georges Antoniadis, Sandra Echinard, Olivier Kraif, Thomas Lebarbé, Claude Ponton. Modélisation de l'intégration de ressources TAL pour l'apprentissage des langues : la plateforme MIRTO. ALSIC - Apprentissage des Langues et Systèmes d'Information et de Communication, 2006, 08 (1), pp.65-79. edutice-00109523

HAL Id: edutice-00109523

<https://edutice.hal.science/edutice-00109523>

Submitted on 24 Oct 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Modélisation de l'intégration de ressources TAL pour l'apprentissage des langues : la plateforme MIRTO

Georges ANTONIADIS, Sandra ECHINARD, Olivier KRAIF, Thomas LEBARBÉ,
Claude PONTON
Université Stendhal, France

Résumé : *Le présent article se focalise sur le développement d'outils de traitement automatique des langues (TAL) pour l'apprentissage des langues assisté par ordinateur (ALAO). Après avoir identifié les limitations inhérentes aux outils d'ALAO dépourvus de composantes TAL, nous décrivons le cadre général du projet MIRTO, une plateforme de création d'activités pédagogiques fondé sur des outils TAL en développement au sein de notre laboratoire. Cette plateforme est organisée en quatre couches distinctes et successives : fonctions, scripts, activités et scénarios. À travers plusieurs exemples, nous expliquons en quoi l'architecture de MIRTO permet l'implantation de fonctions TAL classiques au sein de scripts, lesquels facilitent la conception, sans compétence informatique préalable, d'activités didactiques, elles-mêmes éventuellement intégrées au sein de séquences plus complexes, ou scénarios.*

- 1. ALAO : conjuguer le TAL avec la didactique des langues
- 2. La structure générale de MIRTO
- 3. Scripts TAL pour l'ALAO
- 4. Fonctionnalités implémentées de MIRTO et perspectives
- Références

1. ALAO : conjuguer le TAL avec la didactique des langues

Sil est généralement admis que l'informatique peut être d'un grand secours pour l'apprentissage des langues, force est de constater que, le plus souvent, informaticiens et didacticiens des langues n'admettent pas la même acception du terme "langue". Pour les uns, il s'agit d'une séquence de codes, pour les autres d'un système de formes et de concepts. Cette différence peut s'expliquer aisément lorsque l'on constate que l'informatique, par définition, ne peut considérer et ne peut traiter que la forme de la langue dénuée de toute interprétation, alors que pour la didactique des langues, la forme n'existe que par ses propriétés et des concepts qu'elle est censée représenter.

Les conséquences de cette double approche sont "visibles" dans la grande majorité des logiciels d'apprentissage des langues, et à l'origine de bon nombre de leurs imperfections. La plupart des

didacticiels de langues sont pensés et construits comme des produits informatiques, capables de ne prendre en considération que des chaînes de caractères dénuées de toute sémantique, ou munie d'une sémantique extrêmement pauvre. Pour prendre un exemple quelque peu caricatural, certains logiciels considèrent comme erronée la réponse d'un apprenant si elle comporte une espace de trop. Des règles simples, comme celles liées à l'interprétation de l'espace et d'autres variations graphiques superficielles (majuscules / minuscules, œ / oe, majuscules accentuées / non accentuées, etc.) sont ignorées, ce qui conduit à des situations d'apprentissage biaisées par de fausses détections d'erreurs. Les énoncés linguistiques connaissent d'autres niveaux de variation bien plus complexes (orthographique, morphosyntaxique, idiomatique, sémantique, etc.) et ce "syndrome du blanc" révèle l'ampleur du chemin à parcourir pour que les didacticiels de langue puissent traiter le matériau linguistique dans toute sa complexité. Si l'on attend **la maison** comme réponse à une question, non seulement on doit pouvoir accepter **LA MAISON** (avec deux espaces) mais également **la cabane**, quitte à enrichir la rétroaction d'une indication sur la nuance sémantique.

Comme le signalent [Chanier98], [Nerbonne98] ou [Brun02], mais aussi comme nous avons pu le montrer dans [Antoniadis04], seule l'utilisation des méthodes et techniques du TAL permet d'envisager et de traiter la langue en tant que système de formes et de concepts. Leur considération peut conduire à des réponses pour deux problèmes des logiciels d'ALAO actuels :

Le premier concerne la rigidité des logiciels : les données (énoncés, exemples, réponses attendues, etc.) utilisés doivent être prédéfinies et elles ne peuvent être, à de rares exceptions près, ni modifiées ni enrichies. Les processus de traitement d'affichage, de traitement des réponses...sont intimement liés à ces données. Ils peuvent difficilement prendre en considération des entrées nouvelles, non prévues explicitement à l'origine.

Le second problème concerne la non adaptabilité des parcours des apprenants. Deux types de parcours sont généralement proposés par les logiciels d'ALAO. Le premier, le plus classique, propose une suite prédéfinie et linéaire d'activités. Quelles que soient ses réponses et ses motivations l'apprenant passera et repassera (s'il refait un exercice particulier ou l'ensemble du parcours) toujours par les mêmes activités avec les mêmes données. Le second type de parcours proposé est le parcours "libre" dans un univers scénarisé. C'est le cas des logiciels d'exploration proposant à l'apprenant de réaliser une mission dans un univers donné (réalité virtuelle). Dans ce type de simulation, les activités, qu'il s'agisse de communication ou d'exercices structuraux, sont prédéfinies mais l'apprenant les réalisera dans un ordre qui dépendra de son cheminement dans sa mission. Si ce type de parcours permet une plus grande liberté d'action pour l'apprenant (ordre de la mission, choix des activités, etc.), il n'offre pas pour autant une réelle personnalisation et une adaptation des activités aux compétences linguistiques de l'apprenant. En effet, le cheminement est indépendant de ses réponses à chaque étape, faute de pouvoir les évaluer.

Un dernier problème, dérivant en partie de deux premiers, caractérise les logiciels actuels de l'ALAO. Produits pédagogiques, ces logiciels devraient être élaborés, *a priori*, sur la base de solutions exprimées, sans contraintes, à l'aide des concepts didactiques. Or, force est de constater que les logiciels auteurs actuels (comme *Course builder*, *Netquizz*, *Hot Potatoes*), contraignent leurs utilisateurs (des enseignants des langues, avec des connaissances informatiques limitées ou quasi-inexistantes) à manipuler des concepts et des notions qui ne font pas *a priori* partie de leur problématique, relative à l'apprentissage des langues. Ainsi, au lieu d'exprimer des réponses pédagogiques à l'aide d'outils de leur discipline, ils sont contraints de chercher des solutions informatiques qui se rapprochent le plus de leurs modèles et propositions pédagogiques, voire

d'abandonner des solutions pédagogiques, parce qu'ils ne savent pas les exprimer en termes informatiques ou parce que la technologie n'est pas capable de les prendre en charge. La pratique pédagogique fait appel à des scénarios pédagogiques ouverts, des exercices qui peuvent s'adapter à l'apprenant, à des exemples variés malgré la répétition des séances, à des situations d'apprentissage riches sur plan communicationnel... L'informatique, avec ses propres contraintes, peut fournir des outils pour servir cette problématique, ce qui ne signifie pas qu'il faille se contenter de partir des outils existants pour se poser la question de leur pertinence pédagogique : le *moyen* ne doit pas déterminer la *fin*. Le recours à d'autres connaissances (en linguistique et en didactique des langues) ainsi que leur modélisation sont indispensables. L'utilisation des procédures du TAL peut apporter une solution pour les connaissances linguistiques ; le travail en collaboration entre didacticiens des langues et spécialistes du TAL devrait apporter des réponses pour ce champ d'application.

Les problèmes que nous venons de présenter expliquent, à notre avis, les travers de plusieurs logiciels d'ALAO. Ils ont été pensés et élaborés comme des problèmes et produits informatiques pour lesquels on décline (on fait varier) les aspects de la pédagogie des langues que l'informatique est apte à considérer. Assez souvent, les solutions pédagogiques sont altérées ou tronquées pour qu'elles puissent être informatisées. C'est le cas, typique, des évaluations où l'on considère qu'il n'y a qu'un seul énoncé correct, uniquement parce que le logiciel n'est capable de n'en traiter qu'un seul, alors que sur le plan linguistique (et communicationnel) toute une classe d'énoncés pourrait être acceptée. La vue réductrice consistant à considérer la langue comme une simple suite de codes informatiques est à l'origine de cette approche, et des insuffisances de nombreux produits.

Nous considérons qu'un didacticiel de langues est avant tout un produit pédagogique, un programme qui met en œuvre une solution pédagogique pour un problème de didactique des langues sans altérer, ni la solution, ni, *a fortiori*, le problème. La conception de tels logiciels demande que l'on puisse décliner les possibilités de l'informatique afin qu'elle réalise des solutions pédagogiques déterminées en amont. Dans cette démarche, la prise en considération des propriétés de la langue, constamment présentes dans toute solution pédagogique pour les langues, est une condition nécessaire. Ce sont des méthodes, techniques et produits du TAL qui, par définition, pourront satisfaire cette condition. Seule la mise en œuvre de ces techniques, conjuguées à d'autres outils susceptibles de prendre en charge les aspects communicationnels (plateformes, mondes virtuels, etc.) permettra d'offrir aux enseignants et spécialistes de didactique, non seulement des outils non réducteurs pour leur problématique, mais, aussi, des outils à plus-value pédagogique, capables d'élargir la problématique de leur discipline.

L'utilisation du TAL pour la conception de logiciels d'ALAO n'est pas une idée nouvelle ; quelques systèmes comme *ELEONORE* ([Renié95], [ChanierRenié95]), *ALEXIA* [SelvaChanier00], ou la plateforme *EXILLS* [Brun02] font appel à lui et utilisent des produits qui en sont issus. Néanmoins ces exemples restent encore marginaux et concernent des produits, à notre connaissance, restés au stade du prototype. Paradoxalement, l'ALAO et le TAL, deux domaines centrés sur la langue, semblent encore s'ignorer. Assez souvent, la non utilisation du TAL est justifiée par le coût supplémentaire qu'entraîne son utilisation. Le degré de technicité requis, la complexité des formalismes et le coût humain lié à l'adaptation des ressources linguistiques (dictionnaires, grammaires, corpus, etc.) peut s'avérer réhibitoyre. Pourtant un véritable effort de standardisation est en train de s'accomplir dans la communauté, et l'on voit poindre de plus en plus de ressources "génériques" dans la mouvance du logiciel libre (concordanciers, étiqueteurs, lemmatiseurs, etc.). Le plus souvent, ces briques de base ne nécessitent pas d'autre investissement que l'adaptation des formats d'entrée/sortie et la prise en main des paramètres de base. Dans l'état de l'art, la mise en

oeuvre des outils les plus simples est susceptible d'apporter une plus-value qui dépasse de loin ce modeste investissement. Le retard pris dans les applications est aussi à imputer, selon nous, à une certaine méconnaissance du TAL.

Dans la lignée des systèmes précédemment cités, la plateforme *MIRTO* ([Forestier02], [AntoniadisEchinard04]), dédiée à la création et à la mutualisation de ressources pédagogiques (activités et scénarios), tente d'apporter une réponse à certaines limitations des logiciels de l'ALAO par l'apport des technologies du TAL dans le cadre d'une collaboration étroite avec des spécialistes de la didactique des langues. Plus qu'un produit fini, *MIRTO* se veut un outil d'exploration et de mise en œuvre de solutions didactiques pour l'apprentissage des langues.

2. La structure générale de MIRTO

Le but de *MIRTO* est d'offrir aux enseignants de langue la possibilité de concevoir des scénarios pédagogiques en bénéficiant pleinement d'outils TAL, le tout grâce à une interface orientée utilisateur. Ainsi ces scénarios seront ouverts (base de textes dynamique), permettront une adaptation personnalisée à l'apprenant (génération automatique d'exercices, évaluation qualitative des réponses, etc.) et devraient offrir de nouvelles possibilités (travail sur de longs textes grâce à des outils de concordance, génération automatique d'aides ou d'exercices, conception de scénarios non linéaires, etc.). L'approche qui sous-tend *MIRTO* est résolument orientée utilisateur dans la mesure où la plateforme est destinée à des enseignants de langue, qui, a priori, n'ont que peu ou pas de connaissances en TAL ou en informatique. La nature technique du TAL doit être transparente pour les enseignants de langue et seuls les aspects didactiques et pédagogiques doivent être visibles et disponibles pour l'utilisateur.

Ainsi, quatre niveaux hiérarchiques (fonction, script, activité et scénario), associés à la base de textes, structurent *MIRTO* comme l'illustre la figure 1.

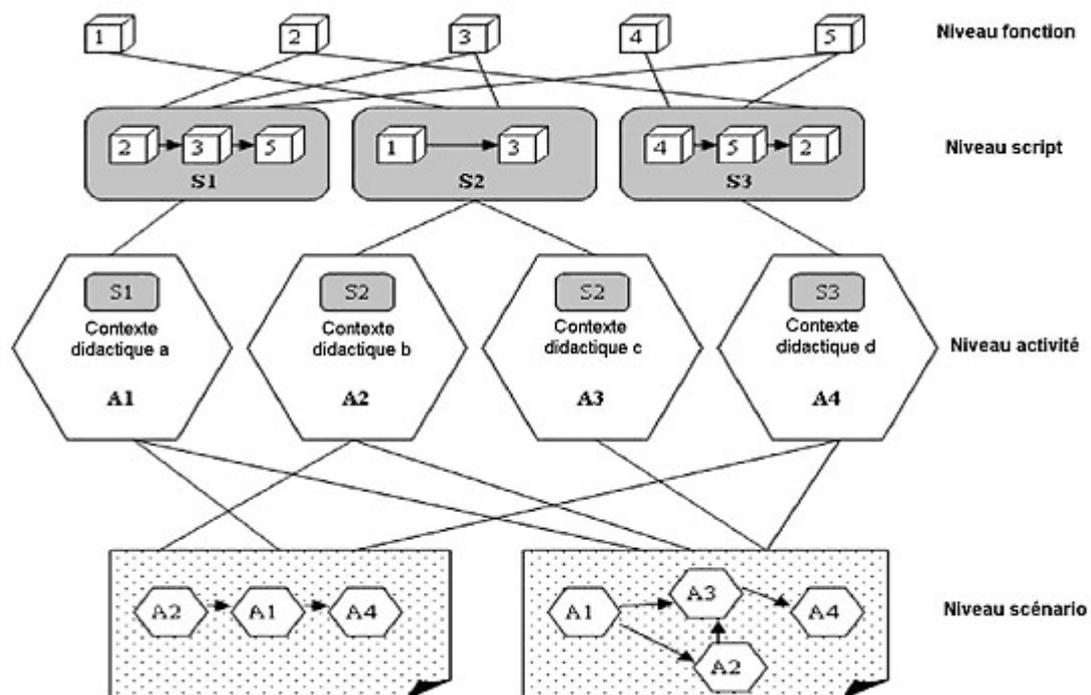


Figure 1 - Structure de MIRTO.

Fonction

Les fonctions représentent le niveau le plus fin de la granularité des modules de *MIRTO*. Elles correspondent à un processus TAL basique comme la "tokenisation" (découpage du texte en unités morphologiques et lexicales) ou l'identification automatique de la langue. Étant donnée la nature technique des fonctions et leur indépendance de toute application pédagogique, ce niveau n'est pas visible par les utilisateurs.

Script

Ce niveau correspond à l'application de fonctions TAL à la didactique des langues. Un script est une suite de fonctions liées par un objectif pédagogique. Par exemple, la génération automatique d'exercices lacunaires est considérée comme un script car elle lie les fonctions de diagnostic automatique de la langue, de tokenisation du texte, d'analyse morphologique et de création des lacunes ("trous") en fonction de paramètres entrés par l'utilisateur. Le fonctionnement d'un script est transparent pour l'enseignant concepteur de scénarios puisqu'il lui est présenté comme une boîte à outils.

Activité

Une activité est la mise en œuvre d'un objectif pédagogique minimal comme travailler une notion grammaticale particulière, rédiger un paragraphe sur un sujet, réviser des conjugaisons... Elle se réalise donc par la construction d'un espace de travail pour l'apprenant, lui permettant d'atteindre le but visé. Elle correspond à ce que l'on désigne traditionnellement comme exercice. D'un niveau purement didactique, la conception d'activités est une tâche opérée par les enseignants de langue à l'aide d'un outil spécifique de *MIRTO* : l'éditeur d'activités. Il s'agit d'un environnement de conception (système auteur) permettant de visualiser et de manipuler des objets et des outils pédagogiques tels que des textes (ou corpus de textes), des scripts et des consignes.

Afin d'illustrer cette opération de conception, considérons l'exemple d'un enseignant désirant créer une activité de révision systématique de l'indicatif imparfait au travers d'un exercice lacunaire (cf. la figure 2).

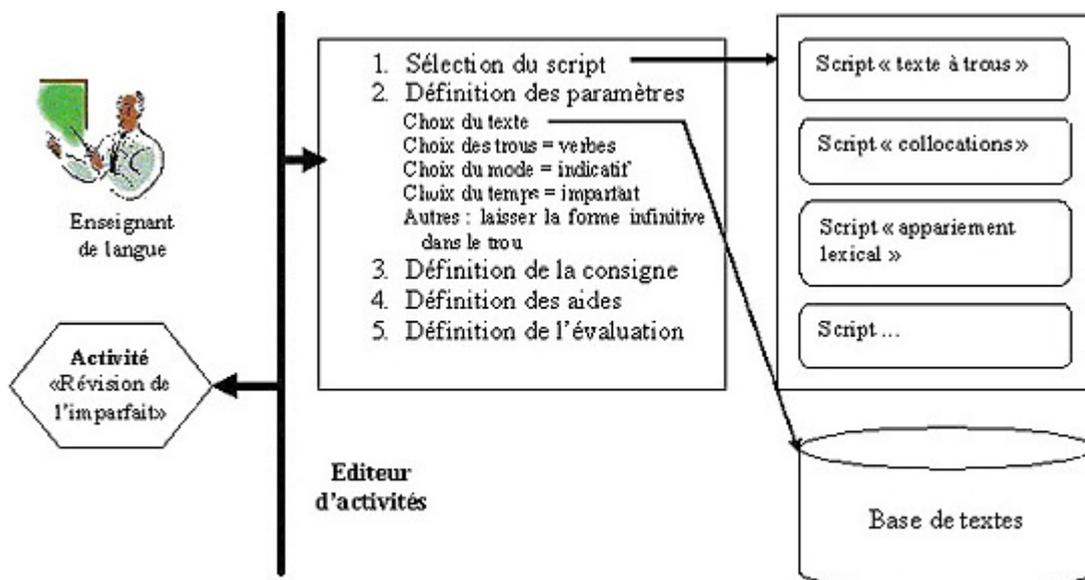


Figure 2 - Exemple de création d'activité.

La première étape de la création d'une activité concerne le choix de l'outil (le script), adapté aux besoins, suivi du paramétrage de son contexte didactique d'utilisation. Dans notre exemple, la sélection du script texte à trous induit un paramétrage, de la part de l'enseignant, au niveau du texte à utiliser et des trous à réaliser sur ce texte (choix des formes à enlever en fonction de critères morphosyntaxiques). Avant la génération effective de l'activité, il reste à définir des paramètres plus généraux comme la consigne, les aides éventuelles (fiches de cours, dictionnaires, etc.) et le type d'évaluation.

Scénario

Ce niveau permet aux enseignants de définir des séquences d'activités. Les scénarios permettent de répondre à leurs objectifs pédagogiques en fonction du parcours de l'apprenant. Ce dernier peut être différent pour chaque apprenant. En effet, chacun d'entre eux aura un processus d'apprentissage personnel lié à différents facteurs. *MIRTO* prend en compte cette réalité en proposant la création de scénarios non linéaires. Le parcours dans le scénario dépendra du processus individuel de chaque apprenant (parcours d'apprentissage, évaluation, échecs ou succès, choix personnels...) et sera stocké dans un historique propre à chaque apprenant. Selon son parcours dans un scénario donné, un apprenant pourra être redirigé vers une activité de remédiation, retenter la même activité sur autre texte ou encore simplement avancer dans le scénario.

3. Scripts TAL pour l'ALAO

Le niveau des scripts concerne la programmation des outils didactiques disponibles dans l'environnement de *MIRTO*, afin d'élaborer des produits pédagogiques. Les scripts sont des modules qui intègrent des ressources ou des traitements TAL standard, tels que l'étiquetage morphosyntaxique, la lemmatisation, l'analyse syntaxique, ou l'interrogation de dictionnaires. La standardisation de ces fonctions est un aspect central, car *MIRTO* ne vise pas au développement de nouvelles techniques TAL, mais seulement à constituer un cadre permettant de tirer parti de l'état de l'art existant, au sein d'une architecture modulaire.

Ainsi, les scripts représentent le coeur de l'architecture de *MIRTO* : ils constituent le point de rencontre entre les potentialités de la technologie TAL et les besoins de la pratique pédagogique.

3.1. Paramètres

Comme pour tout module informatique, l'exécution d'un script requiert un ensemble de paramètres. Ces paramètres sont accessibles, sous une forme simplifiée, à travers un tableau de bord, dont les commandes doivent être configurées pour un usage didactique. Prenons l'exemple du générateur d'exercices lacunaires. Par le biais d'un formulaire, l'utilisateur peut spécifier.

- a) Les unités à escamoter. Des critères linguistiques doivent pouvoir être précisés : le lemme (par exemple, [conduire](#)), la partie du discours (par exemple, [verbe](#)), des traits morphosyntaxiques (par exemple, [passé simple](#)), voire le contenu sémantique (par exemple, tout lexème sémantiquement lié à [voiture](#) ; cette fonctionnalité est en cours d'étude).
- b) L'information à faire figurer à la place du mot escamoté : aucune, le lemme, les traits morphosyntaxiques, un synonyme, une définition (non encore implémenté), etc.

- c) Si les mots escamotés apparaissent sous la forme d'une liste ordonnée qui accompagne le texte.
- d) Si la réponse de l'apprenant doit donner lieu à une rétroaction immédiate au niveau de chaque champ à remplir, ou au niveau de l'ensemble de l'activité.

Les commandes d'un tel script doivent rester :

- simples – un trop grand nombre de paramètres pourrait décourager l'utilisateur ;
- déclaratives – l'utilisateur n'est pas censé maîtriser un formalisme complexe ; la définition des paramètres doit donc rester intuitive, pour un non informaticien ;
- conviviales – les informations importantes doivent être faciles à appréhender ; les paramétrages avancés et secondaires, doivent être accessibles via un second formulaire facultatif.

Dans certains cas, la définition des critères linguistiques doit donner lieu à un processus de transcription. Par exemple, le générateur d'exercices lacunaires prend comme paramètres les étiquettes morphosyntaxiques utilisées dans l'encodage des textes {base="er\$", ctag="verb", msd="IndP SG P1"}^[1]. L'utilisateur préférera manier la terminologie des grammaires scolaires, dans la langue cible "*Verbo, Prima coniugazione, Indicativo, Presente, Prima persona, Singolare*".

3.2. Une approche incrémentale

Il est difficile de déterminer *a priori* la forme exacte qu'un script doit avoir. Ceci pour deux raisons.

- Les fonctions TAL peuvent demander des paramétrages très complexes pour atteindre le résultat voulu, et leurs entrées et sorties peuvent avoir des formes assez hétérogènes.
- Le champ d'application du TAL pour l'ALAO est encore assez neuf, même si les premiers travaux datent d'une vingtaine d'années. Ces technologies de la langue peuvent déboucher sur de nouveaux types d'activités et de nouvelles habitudes pédagogiques. Les besoins sont donc susceptibles de changer en même temps que l'offre se diversifie.

Ainsi, seule la pratique pédagogique permettra de déterminer, parmi les nouvelles possibilités offertes, quelles sont les solutions les plus intéressantes. Un script peut très bien offrir des fonctionnalités très riches sur le plan technologique, mais sans intérêt pédagogique. Par ailleurs, un script peut se révéler utile pour des applications pour lesquelles il n'avait pas été initialement conçu.

Un double mouvement est donc requis : d'une part il convient d'offrir un large éventail de possibilités, afin que les enseignants puissent sélectionner celles qui leur paraissent les plus intéressantes. D'autre part, les nouvelles pratiques enseignantes pourront déboucher sur de nouveaux besoins, auxquels la technologie s'efforcera de répondre.

Comme nous l'avons suggéré dans [Kraif04], pour amorcer ce va-et-vient entre technique et pratique, il est plus profitable de se consacrer à des activités existantes pouvant être améliorées avec des techniques TAL de base, faciles à mettre en oeuvre et correctement maîtrisées. A partir de

ces activités, nous avons essayé de concevoir des scripts les plus modulaires possibles, réutilisables dans divers contextes et pour un large champ d'applications didactiques.

Enfin, les limites inhérentes aux fonctions TAL employées, en terme de performance et de correction, ont constitué une contrainte majeure de ces développements. Par exemple, quand un traitement TAL produit des erreurs dans 20% des cas, les résultats immédiats du traitement peuvent se révéler inutilisables pour certaines applications pédagogiques, car il n'est pas acceptable de donner à un apprenant des informations erronées.

3.3. Exemples de scripts

La plupart des exemples qui suivent ne sont pas encore totalement implémentés dans la plateforme *MIRTO*. Mais ils sont tous réalistes, dans la mesure où il se basent sur l'état de l'art actuel du TAL et peuvent être mis au point à court terme.

Nous distinguons trois catégories de scripts.

3.3.1. Générateurs d'activités

Étant donné un texte fourni en entrée, les techniques du TAL permettent de sélectionner des lexèmes et des expressions en fonction de critères lexicaux, idiomatiques, grammaticaux ou sémantiques. Sur cette base, on peut créer un large éventail d'activités de type exercice lacunaires, identification de mots, ou quiz (avec des questions sur les synonymes, contraires, définitions, etc. en rapport avec des lexèmes du texte).

Le tableau 1 donne différents exemples d'activités générées sur ce modèle.

En amont, il est possible de faire appel à des scripts pour la constitution du texte d'entrée. Ce sera par exemple le rôle d'un script dédié à l'extraction des concordances qui permettra d'identifier, à partir d'un corpus, toutes les unités (ainsi que leur contexte) qui satisfont à certains critères. Un tel concordancier, monolingue ou multilingue, pourrait également donner lieu à des activités où l'apprenant doit résoudre une difficulté par analogie, en recherchant des exemples attestés à l'intérieur d'un corpus.

Critère de sélection	Script	Exemple de consigne	Réponse attendue	Fonction TAL mise en oeuvre
Sémantique	Identification lexicale	Marquez tous les mots du texte en rapport avec voiture	Sélection de drive, taxi, engine, road , etc.	étiquetage morphosyntaxique, lemmatisation, interrogation d'un réseau sémantique
Sémantique	Quiz lexical	Donner un équivalent italien pour conduire	Entrée de guidare	étiquetage morphosyntaxique, lemmatisation, interrogation d'un dictionnaire bilingue
Morphosyntaxique	Exercice lacunaire	Remplacer tous les verbes à l'infinitif par leur forme conjuguée au temps approprié	Remplacement de to wait par have been waiting...	étiquetage morphosyntaxique, lemmatisation
Morphosyntaxique	Quiz lexical	Quel est le contraire de lentement ?	Entrée de rapidement	étiquetage morphosyntaxique, lemmatisation, interrogation d'un réseau sémantique
Morphologique	Identification lexicale	Marquez tous les mots de la famille de traduire	Sélection de traducteur, traduction, retraduite , etc.	étiquetage morphosyntaxique, lemmatisation, extraction des radicaux
Morphologique	exercice lacunaire	Remplacer les trous par un mot de la famille de traduire	Entrée de traducteur, traduction, retraduite , etc.	étiquetage morphosyntaxique, lemmatisation, extraction des radicaux

Tableau 1 - Exemples de scripts pour la génération d'activités.

Les critères de sélection peuvent être appliqués non sur des textes mais sur des ressources de type dictionnaire, afin de générer des activités hors contexte du type :

- Conjugueur. Par exemple, Conjuguer le verbe **tomber** au subjonctif imparfait, première personne du singulier.
- Quiz lexical. Par exemple, Donner un synonyme du mot **phare**
- Exercice de morphologie. Donner un nom dérivé du verbe **conduire**.

Enfin, les *chatbots* (automates de dialogue), capables de simuler une conversation avec un interlocuteur virtuel, dans un contexte pragmatique déterminé, constituent une autre application

intéressante des techniques du TAL pour la génération d'activité.

3.3.2. Aides à la compréhension

Pour tous les types d'activités (lecture de texte, exercice, etc.) on peut proposer des aides interactives pour l'apprenant. La plupart des outils TAL disponibles sur la plateforme *EXILLS* appartiennent à cette catégorie. Par exemple, à tout moment, l'apprenant peut y consulter un dictionnaire dont les définitions sont désambiguïsées en fonction du contexte, ou bien utiliser un phonétiseur pour trouver l'orthographe correcte d'un mot.

De telles aides peuvent être génériques (dictionnaires, fiches grammaticales, conjugueurs) ou contextuelles, comme dans le cas où un clic sur un mot d'un texte permet d'accéder à son analyse morphosyntaxique, sa définition et/ou sa traduction. Par exemple, nous avons implémenté un système d'aide contextuelle permettant de relier automatiquement certains lexèmes à des fiches de grammaire en rapport avec la catégorie du mot, ses traits morphosyntaxiques, sa fonction syntaxique ou tout autre trait linguistique. Dans l'exemple de la figure 3 tout verbe conjugué au *passato remoto* est automatiquement relié, via un hyperlien, à une fiche explicative concernant cette conjugaison.



Figure 3 - Exemple d'aide contextuelle.

3.3.3. Évaluation automatique

La production de l'apprenant, dans le cadre d'une activité, peut prendre des formes très variées : ce peut être un clic sur une case à cocher, un mot, une phrase voire un texte. La correction de phrases ou de textes est un problème complexe. Les techniques du TAL ne peuvent donner de résultats

fiables sur des aspects qui demandent une interprétation humaine (sens, style, etc.). Même pour la simple détection de fautes de grammaire, les modèles existants sont à la fois silencieux et bruyants : certaines erreurs ne sont pas détectées tandis que des énoncés corrects sont considérés comme erronés. A l'opposé, l'évaluation d'un QCM est un problème trivial qui ne demande aucune technologie linguistique.

Pour le moment, nous pensons que les applications les plus prometteuses et réalistes du TAL concernent l'évaluation de productions lexicales simples (lexèmes, expressions semi figées, expressions idiomatiques, etc.). Nous étudions la mise en place d'un protocole à trois niveaux permettant la comparaison d'une réponse donnée, avec la réponse considérée comme correcte. Quand la réponse donnée diffère de la réponse attendue, on considère trois cas de figures.

- 1) Niveau orthographique : Si la chaîne n'existe pas dans un dictionnaire de formes fléchies, on suppose qu'il y a une faute d'orthographe. Si la chaîne est très proche de la réponse correcte, un message diagnostique la faute de frappe ou la graphie erronée. Sinon une liste de mots ressemblants est proposée à l'apprenant, afin qu'il y choisisse sa propre solution correctement orthographiée.
- 2) Niveau morphosyntaxique : La réponse donnée est analysée afin d'en déterminer le lemme et les traits morphosyntaxiques potentiels. S'il y a identité de lemme avec la réponse correcte, alors les différences de traits peuvent donner lieu à un diagnostic du type : genre incorrect ou nombre incorrect, ou le temps ne convient pas.
- 3) Niveau sémantique : Dans le cas d'une différence au niveau des lemmes, on utilise un réseau sémantique (*EuroWordNet*) afin de déterminer s'il y a une relation sémantique proche (du type synonymie, hyperonymie, hyponymie, métonymie, antonymie) entre la réponse donnée et la réponse attendue. Si c'est le cas, un diagnostic peut être effectué du type : soyez plus précis, cherchez un sens voisin, etc.

Dans l'architecture globale de *MIRTO*, un tel script pourrait être utile dans l'évaluation de diverses activités : exercice lacunaire, quiz, etc. Selon les objectifs pédagogiques d'une activité, la rétroaction requise peut varier du tout au tout. Par exemple, dans un exercice lacunaire destiné à travailler la conjugaison des verbes au passé simple, une différence au niveau des lemmes n'est pas essentielle du moment que la flexion verbale est correcte.

C'est pourquoi, pour une plus grande modularité, comparaison et rétroaction doivent être détachées. L'évaluation impliquera donc deux scripts :

- Un script de comparaison qui prend en entrée la réponse attendue (comportant des informations morphosyntaxiques issues de traitement précédents) et la réponse donnée, et renvoie un code de différence du genre :

0 : pas de différence

1.1 : faute d'orthographe sur la réponse attendue

1.2 : faute d'orthographe sur un autre lexème

2.1 : traits morphosyntaxiques identiques

2.2 : traits morphosyntaxiques différents

2.2.1, 2.2.2,... : nombre différent, genre différent

3.1 : lemme différent avec lien sémantique

3.1.1, 3.1.2 synonyme, hyperonyme, etc.

3.2 : lemme différent sans lien sémantique

- Un script de rétroaction qui prend en entrée un code de différence, et renvoie un message pour chaque cas de figure, par exemple **oui, mais l'orthographe est erronée** (code 1.1), **soyez plus précis** (code 3.1.2), etc.

Même si on peut proposer des messages standard pour chaque situation, l'enseignant doit être en mesure de modifier et d'adapter la rétroaction en fonction de l'activité et de ses objectifs pédagogiques.

4. Fonctionnalités implémentées de MIRTO et perspectives

Le développement de *MIRTO* a commencé il y a environ un an. Un total de trois ans devrait être nécessaire pour achever le premier prototype. Une période de prise en main par les utilisateurs est à prévoir pour permettre aux enseignants de maîtriser l'utilisation du produit. *MIRTO* devrait être intégré à l'intranet de l'université Stendhal Grenoble 3 pour expérimentation.

Jusqu'ici, le développement de *MIRTO* concernait principalement le module de création de scripts. Réaliser ce module a permis l'intégration de divers logiciels TAL (et non TAL). D'autres logiciels, principalement TAL, devraient être intégrés. Le choix du nombre et de la nature des logiciels intégrés ne peut être fait que grâce à un processus d'échange impliquant aussi bien des enseignants et des didacticiens des langues que des experts du TAL. Nous considérons que les logiciels qui sont déjà intégrés permettent la création de suffisamment de scripts pour une utilisation expérimentale de *MIRTO*.

Afin d'effectuer des tests et de valider l'approche globale, une première version des éditeurs d'activités et de scénarios a été implémentée. Cette version permet la création de quasiment n'importe quel type d'activité (excepté l'évaluation) et la conception de scénarios linéaires qui ne garderont pas de trace du parcours d'apprentissage de l'apprenant.

La définition de l'approche sous-jacente à *MIRTO* ainsi que l'usage qui en est fait proviennent de plusieurs projets de recherche en cours. Excepté l'implantation du prototype, nos efforts se concentrent particulièrement sur les aspects suivants :

- l'annotation et l'indexation pédagogiques de textes dans la perspective de la constitution d'un corpus pédagogique pour enseignants en langues [Loiseau03]
- les analyses automatiques et pédagogiques des réponses des apprenants à l'aide d'outils et de méthodes issus des recherches en TAL ;
- l'interfaçage de la rédaction de scripts pour la génération d'activités.

A la jonction de ces trois orientations scientifiques – didactique des langues, traitement automatique des langues et informatique – *MIRTO* soulève de nouvelles problématiques, non seulement spécifiques à chacune de ces branches, mais aussi des problèmes plus généraux

concernant l'ensemble de ces branches, requérant une quasi-osmose entre les différents experts. L'on peut ainsi mentionner parmi les nombreux exemples, la génération automatique de commentaires pédagogiquement adéquats à la réponse de l'apprenant, la modélisation et l'implantation de fonctions manipulant des concepts didactiques (de manière à apporter des outils simples aux enseignants en langues sans compétence particulière en informatique), la définition et l'exploitation pédagogique de traces des activités des apprenants ou la modélisation de scénarios pédagogiques non linéaires. De tels résultats ne peuvent être obtenus que dans un contexte interdisciplinaire, apportant ainsi une forte valeur ajoutée aux outils pour l'apprentissage des langues assisté par ordinateur.

Références

Les liens externes étaient valides à la date de publication.

Bibliographie

[AntoniadisEchinard04]

Antoniadis, G., Echinard, S., Kraif, O., Lebarbé, T., Loiseau, M., Ponton, C. (2004). "NLP-based scripting for CALL activities". *Proceedings of the Workshop eLearning for Computational Linguistics, and Computational Linguistics for eLearning, 20th International Conference on Computational Linguistics*, Genève, Suisse, 28 août 2004. pp. 18-25.

[Antoniadis04]

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.

[Brun02]

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.

[Chanier98]

Chanier, T. (1998). "Relations entre le TAL et l'ALAO ou l'ALAO un simple domaine d'application du TAL ?". *International conference on natural language processing and industrial application (NLP+IA'98)*. août 1998, Moncton, Canada. <http://lifc.univ-fcomte.fr/RECHERCHE/P7/pub/Moncton/index.htm>

[ChanierRenié95]

Chanier, T. & Renié, D. (1995). "Collaboration and computer-assisted acquisition of a second language". *Computer-Assisted Language Learning*, vol. 8, 1. pp. 3-30. <http://archive-edutice.ccsd.cnrs.fr/edutice-00000277>

[Forestier02]

Forestier, G. (2002). *Plates-Formes pour l'enseignement des langues : le cas de MIRTO*. Mémoire pour l'examen probatoire en informatique, CNAM, Grenoble.

[Kraif04]

Kraif O., Antoniadis G., Echinard S., Loiseau M., Lebarbé T., Ponton C. (2004). "NLP Tools for CALL: the Simpler, the Better". in Proceedings of *InSTIL/ICALL 2004 Symposium, NLP And Speech Technologies In Advanced Language Learning Systems*, Venise, Italie, 17-19 Juin 2004. pp. 37-40.

[Loiseau03]

Loiseau, M. (2003). *Vers la création d'une base de données de ressources textuelles indexée pédagogiquement pour l'enseignement des langues*. Mémoire de DEA en sciences du langage, université Stendhal, Grenoble.

[Nerbonne98]

Nerbonne, J., Dokter, D. & Smit, P. (1998). "Morphological Processing and Computer-Assisted Language Learning". *Computer-Assisted Language Learning*, 11, 5. pp. 543-559.

[Renié95]

Renié, D. (1995). *Modélisation informatique de l'acquisition des interrogatives directes en français langue seconde dans leur dimension pragmatique, proposition d'un environnement offrant un apprentissage collaboratif : ELEONORE*. Thèse de doctorat, université Clermont II, Clermont-Ferrand.

[SelvaChanier00]

Selva, T. & Chanier, T. (2000). "Génération automatique d'activités lexicales dans le système ALEXIA". *Sciences et Techniques Éducatives (STE)*, vol. 7, 2. pp. 385-412. <http://archive-edutice.ccsd.cnrs.fr/edutice-00000282>

Notes

[1] Les étiquettes d'encodage des textes sont rarement explicite ; elles indiquent ici qu'il s'agit d'un verbe à l'indicatif présent, première personne du singulier.

À propos des auteurs

Georges ANTONIADIS, **Olivier KRAIF**, **Thomas LEBARBÉ** et **Claude PONTON** sont enseignants chercheurs en informatique et TAL à l'université Stendhal de Grenoble. Ils mènent des recherches en TAL et ALAO au sein du laboratoire LIDILEM. Le projet MIRTO (responsable : G. Antoniadis) est issu de ces travaux de recherche. Il a bénéficié de l'appui de l'université Stendhal, de "Grenoble Université" et de la région Rhône-Alpes.

Sandra ECHINARD est doctorante à l'université Stendhal en TAL et ALAO (génération automatique de commentaires pédagogiquement adéquats à la réponse de l'apprenant). Elle a assuré le développement de la plateforme MIRTO.

Adresse : Laboratoire LIDILEM, université Stendhal, BP 25, 38040 Grenoble cedex 9, France

Courriel : Antoniadis@u-grenoble3.fr, Echinard@u-grenoble3.fr, Kraif@u-grenoble3.fr, Lebarbe@u-grenoble3.fr, Ponton@u-grenoble3.fr

Ce texte fait partie des textes de la journée Atala 2005 qui font l'objet d'un numéro spécial d'Alsic.

Date de réception de l'article 1er novembre 2004 ; date d'acceptation le 15 avril 2005.

Référence de l'article :

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". *Apprentissage des langues et systèmes d'information et de communication (ALSIC)*, vol. 8, n° 1. pp. 65-79. http://alsic.u-strasbg.fr/v08/antoniadis/alsic_v08_04-rec4.htm, mis en ligne le 15/11/2005.



ALSIC | Sommaire | Consignes aux auteurs | Comité de rédaction | Inscription

© *Apprentissage des Langues et Systèmes d'Information et de Communication, novembre 2005*