

MODELISATION DE L'INTEGRATION DE RESSOURCES TAL POUR L'APPRENTISSAGE DES LANGUES : LA PLATEFORME MIRTO

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

Laboratoire LIDILEM

Université Stendhal

BP 25

38040 Grenoble cedex 9, France

[/{Antoniadis, Echinard, Kraif, Lebarbe, Ponton}@u-grenoble3.fr](mailto:{Antoniadis, Echinard, Kraif, Lebarbe, Ponton}@u-grenoble3.fr)

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 *scénarii*.

Mots-clés : TAL, ALAO, TAL pour l'ALAO

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

S'il 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 dénué de toute interprétation, alors que pour la didactique des langues, la forme n'existe que par ses propriétés et des concepts qu'elle est sensée 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 des langues sont pensés et construits comme des produits informatiques, capables de ne prendre en considération que la forme langagière dénuée de toute sémantique, ou munie d'une sémantique extrêmement pauvre. Des règles simples, comme celles liées à l'interprétation de l'espace, sont ignorées, ce qui conduit à des situations d'apprentissage « cocasses ». Ainsi, par exemple, la réponse de l'apprenant « la maison » (séquence comportant deux espaces entre les deux mots la constituant) sera refusée, puisque seule la réponse « la maison » (séquence avec un seul espace) est considérée comme exacte par le logiciel. Les conséquences pédagogiques de ce « traitement » de l'espace sont manifestes ; le logiciel « enseigne » que la séquence de deux espaces ne fait pas partie de la langue, mais aussi, que tout mot précédé ou suivi d'un espace n'a rien de commun avec le même mot sans espace ! Le « syndrome du blanc » est, à notre avis, caractéristique des insuffisances des didacticiels actuels pour les langues.

Comme le signalent (Chanier, 1998) ou (Brun & al., 2002), mais aussi comme nous avons pu le montrer dans (Antoniadis & Ponton, 2004) et (Antoniadis, 2004), 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...) utilisés doivent être prédéfinies et elles ne peuvent être, à de rares exceptions près, ni modifiées ni enrichies. Les processus de traitements d'affichage, de traitement des réponses... sont intimement liés à ces données. Ils ne peuvent donc en aucun cas 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, 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). Les activités de dialogue, grammaticales, ou autres sont prédéfinies mais l'apprenant les réalisera dans un ordre qui dépendra de son cheminement dans sa mission. Ce dernier type de parcours, s'il permet une plus grande liberté d'action pour l'apprenant (ordre de la mission, choix des activités...), n'offre donc pas pour autant une réelle personnalisation

et une adaptation des activités à l'apprenant. En effet, le cheminement est indépendant de ses réponses à chaque étape, faute de pouvoir les évaluer. Signalons enfin, que si l'ordre des activités peut varier en fonction de la mission de l'apprenant, le contenu de chaque activité reste, lui, invariable et sera le même, chaque fois que l'apprenant l'inclura dans son parcours.

Un dernier problème, dérivant en partie de deux premiers, caractérise les logiciels actuels de l'ALAO. Produits didactiques, ces logiciels devraient être élaborés, a priori, sur la base, uniquement, de solutions didactiques exprimées, sans contraintes, à l'aide des concepts pédagogiques. Or, force est de constater que les didacticiels actuels, produits informatiques, contraignent leurs utilisateurs (des enseignants des langues, avec des connaissances informatiques limités 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 abandonner des solutions pédagogiques, parce qu'ils ne savent pas les exprimer informatiquement ou parce que l'informatique n'est pas capable de les prendre en charge. Par exemple, la didactique des langues, à notre connaissance, est capable d'imaginer actuellement des scénarios pédagogiques ouverts, des exercices qui s'adaptent à l'apprenant, de faire varier les exemples lorsqu'on répète la même séance, de choisir des textes appropriés pour illustrer les contextes pédagogiques, de proposer des situations d'apprentissage ouvertes et variables... l'informatique est (et sera) incapable de prendre en considération ces aspects avec sa propre problématique. Le recours à d'autres connaissances (linguistiques et didactiques linguistiques) 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 les connaissances didactiques linguistiques.

Les problèmes que nous venons de présenter expliquent, à notre avis, la nature des didacticiels de langue jusqu'ici. Ils ont été pensés et élaborés comme des problèmes et produits informatiques pour lesquels on décline (on fait varier) les aspects de la didactique des langues que l'informatique est apte à considérer. Assez souvent, les solutions pédagogiques sont altérées ou tronquées pour qu'elles puissent être informatisées. La vue réductrice que l'informatique ne peut que porter sur la langue (simple suite de codes) est à l'origine de cette approche, mais aussi la cause principale des insuffisances des logiciels d'ALAO.

Notre démarche pour l'élaboration des didacticiels des langues va à l'encontre de celle précédemment citée. Nous considérons qu'un didacticiel des langues est avant tout un produit didactique, un programme qui met en œuvre une solution didactique pour un problème de la 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 a priori. 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. Si l'on considère que les méthodes, techniques et produits du TAL sont les seules capables de satisfaire cette condition, alors un didacticiel des langues devrait être défini comme la déclinaison des possibilités du TAL en vue d'atteindre des buts didactiques prédéfinis concernant l'apprentissage d'une langue. A notre avis, seule une telle approche permet d'offrir aux didacticiens des langues, non seulement des outils non réducteurs ou neutres 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 des didacticiels des langues n'est pas une idée nouvelle ; quelques systèmes comme ELEONORE (René, 1995), ALEXIA (Chanier & Selva, 2000), ou la plateforme EXILLS (Brun & al., 2002) font appel à ses procédures et utilisent des produits qui en sont issus. Néanmoins ces exemples restent encore marginaux et concernent des produits, à notre connaissance, non commercialisés. Paradoxalement, l'ALAO et le TAL, deux domaines dont la langue est au centre de leurs préoccupations, semblent encore s'ignorer. Assez souvent, la non utilisation du TAL est justifiée par le coût supplémentaire qu'entraîne son utilisation. S'il est vrai que ce coût peut être important pour quelques techniques ou produits du TAL, la mise en oeuvre de bon nombre d'entre eux entraîne un rapport gain en efficacité sur coût supplémentaire très favorable. A notre avis, plus que le surcoût, qui est souvent évoqué, c'est le défaut de la culture TAL qui est à l'origine de son absence.

Dans la lignée des systèmes précédemment cités, la plateforme MIRTO (Antoniadis & Ponton, 2004) tente d'apporter une réponse globale aux problèmes des logiciels de l'ALAO par le biais, d'une part, d'une approche TAL, d'autre part, d'un travail en collaboration avec des didacticiens des langues. Plus qu'un produit fini, MIRTO se veut un outil de mise en oeuvre des solutions didactiques pour l'apprentissage des langues. Nous présentons par la suite, les aspects du système qui matérialisent notre démarche et sa mise en oeuvre.

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...) et devraient offrir de nouvelles possibilités (travail sur de longs textes, génération automatique d'aides ou d'exercices, conception de scénarios non-linéaires...). L'approche 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 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.

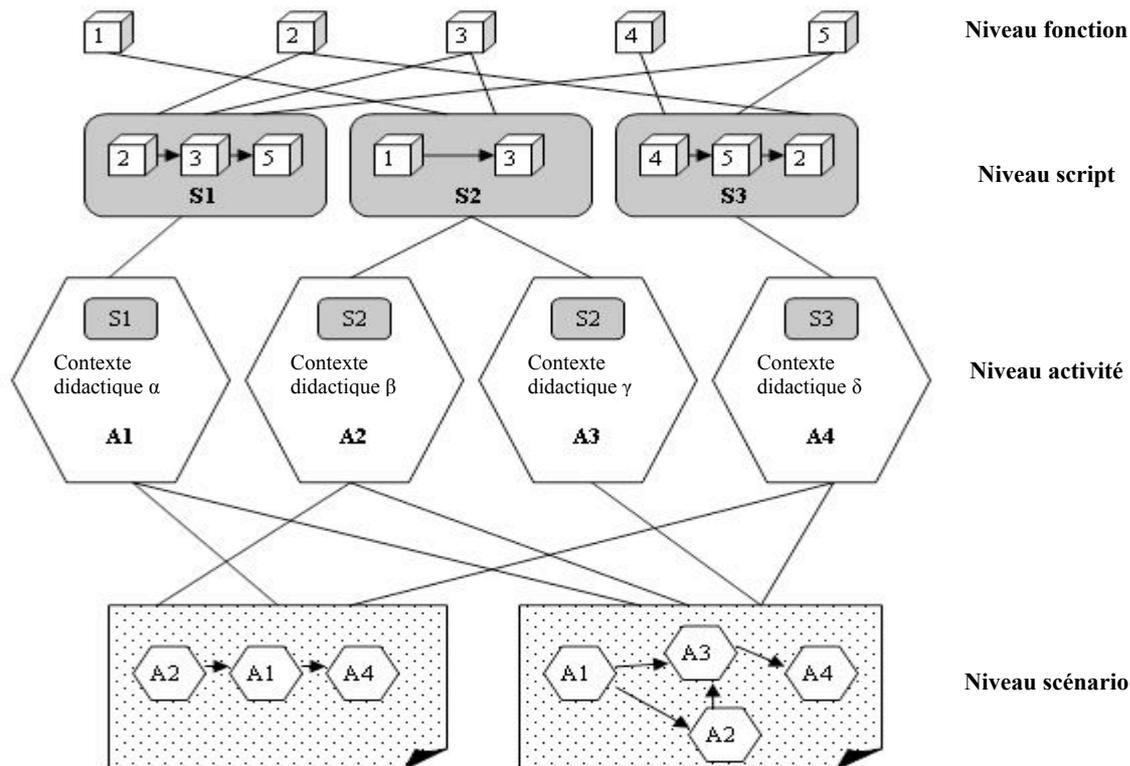


Figure 1 Structure de MIRTO

2.1. Fonction

Les fonctions représentent le niveau le plus bas d'objets MIRTO. Elles correspondent à un processus TAL basique comme la tokenisation (découpage du texte en formes) ou la reconnaissance de langue. Etant donné la nature technique des fonctions et leur indépendance de toute application pédagogique, ce niveau n'est pas visible par les utilisateurs.

2.2. 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 d'identification de la langue, de tokenisation, d'analyse morphologique et de création des "trous" en fonction de paramètres entrés par l'utilisateur.

Le fonctionnement d'un script est transparent pour l'enseignant-concepteur de scénarii puisqu'il lui est présenté comme une boîte à outils.

2.3. 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. L'éditeur d'activités est 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 à trous (Cf. Figure 2).

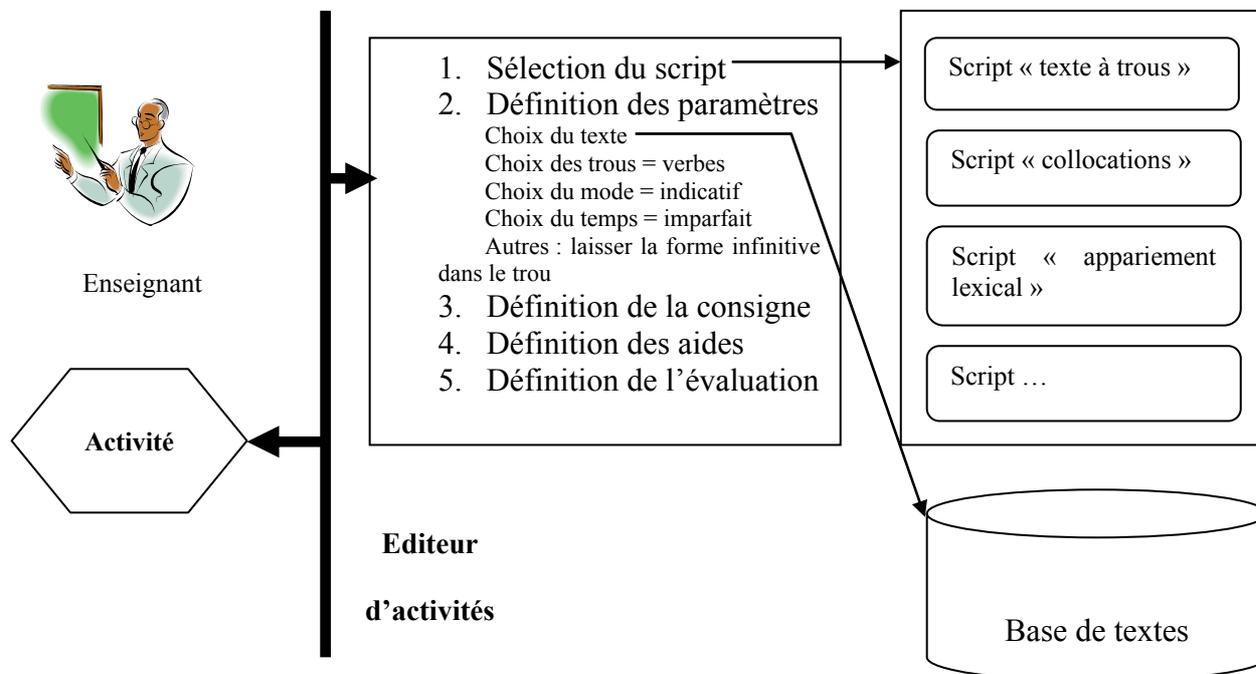


Figure 2 Exemple de création d'activité

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...) et le type d'évaluation.

2.4. 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 du scénario dépendra du processus individuel de chaque apprenant (parcours d'apprentissage, évaluation...). Celui-ci 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. Les scripts sont des modules qui intègrent des ressources ou traitements TAL standards, 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 pédagogiques.

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 à trou. Par le biais d'un formulaire, l'utilisateur peut spécifier :

- a) les unités à escamoter. Des critères linguistiques doivent pouvoir être précisés : le lemme (p. ex. *conduire*), la partie du discours (p.ex. *verbe*), des traits morphosyntaxiques (p. ex. *passé simple*), voire le contenu sémantique (p. ex. 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 en entête du texte.
- d) si la réponse de l'apprenant doit donner lieu à un feed-back immédiat 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 : trop 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 facilement appréhendables. Les paramètres 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 à trou prend comme paramètres les étiquettes morphosyntaxiques utilisées dans l'encodage des textes ("*base=er\$*", *ctag="verb"*, *msd="IndP SG P1"*). L'utilisateur préférera manier la terminologie usitée en didactique, 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. Ces nouvelles technologies 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 (Kraif et al., 2004), 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 des fonctions TAL employées, en terme de performance et de correction, ont constitué une contrainte majeure de ces développements. Quand un traitement TAL produit des erreurs dans 20% des cas, les résultats immédiats du traitement peuvent se révéler inutilisables pour certaines applications didactiques, 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.

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
Morpho-syntaxique	Exercice à trou	Remplacer tous les verbes à l'infinitif par leur forme conjuguée au temps approprié	Remplacement de "to wait" par "have been waiting"...	étiquetage morphosyntaxique, lemmatization
Morpho-syntaxique	Quiz lexical	Quel est le contraire de "lentement"?	Entrée de "rapidement"	étiquetage morphosyntaxique, lemmatisation, interrogation d'un réseau sémantique
Morpho-logique	Identification lexicale	Marquez tous les mots de la famille de "traduire"	Sélection de "traducteur", "traduction", "retraduite", etc.	étiquetage morphosyntaxique, lemmatisation, stemming
Morpho-logique	exercice à trou	Remplacer les trous par un mot de la famille de "traduire"	Entrée de "traducteur", "traduction", "retraduite", etc.	étiquetage morphosyntaxique, lemmatisation, stemming

Tableau 1 Exemples de scripts pour la génération d'activités

3.3.1. Générateurs d'activité

Etant 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 à trou*, *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-concordancier qui permettra d'extraire, à partir d'un corpus, toutes les unités (ainsi que leur contexte) qui satisfont 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 tirés de l'usage à l'intérieur d'un corpus.

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. P.ex. "Conjuguer l'expression *tomber en panne sèche* au subjonctif imparfait, première personne du singulier".
- - Quiz lexical. P. ex. "Donner un synonyme du mot *phare*".
- - Exercice de morphologie. "Donner un nom dérivé du verbe *conduire*".

Enfin, les *chatbots*, 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.

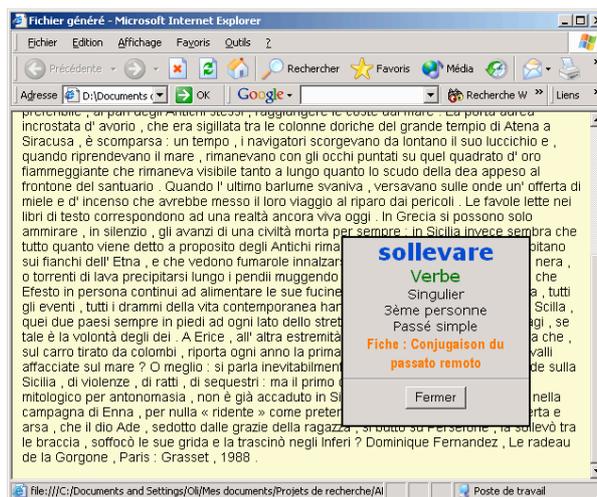


Figure 3 Exemple d'aide contextuelle

3.3.3. Evaluation 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 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 idiomatique, etc.). Nous étudions la mise en place d'un protocole à trois niveaux permettant la comparaison d'une réponse donnée, par rapport à la réponse considérée comme correcte. Quand réponse donnée et réponse attendue diffèrent, on considère trois cas de figure :

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. Si 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, méronymie, 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 à trou, quiz, etc. Selon les objectifs pédagogiques d'une activité, le feed-back requis peut varier du tout au tout. Par exemple, dans un exercice à trou 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 feed-back doivent être détachés. 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 feed-back 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 standards pour chaque situation, l'enseignant doit être en mesure de modifier et d'adapter le feed-back 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 de langue 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'activité 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 provient de plusieurs projets de recherche en cours. Exceptée 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 (Loiseau, 2003)
- 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 à 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

Antoniadis, G. & Ponton, C., 2004, MIRTO : un système au service de l'enseignement des langues, Communication au colloque UNTELE'2004, 17-20 mars 2004, Compiègne.

Antoniadis, G., 2004, Les logiciels d'apprentissage des langues peuvent-ils ignorer le TAL ?, *Les cahiers de l'APLIUT*, n° XXIII vol. 2.

Antoniadis, G., Echinard, S., Kraif, O., Lebarbé, T., Loiseau, M., Ponton, C., 2004, NLP-based scripting for CALL activities, COLING 04, Genève, Suisse, août 2004.

Antoniadis, G., Echinard, S., Kraif, O., Lebarbé, T., Loiseau, M., Ponton, C., 2004, CALL :from current problems to NLP solutions, EUROCALL 04, Vienne, Autriche, septembre 2004.

Bourda, Y., 2001. Objets pédagogiques, vous avez dit objets pédagogiques ? , in *Cahier de Gutenberg*, GUT2001, 39-40, Metz, France.

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*, Hermès Science Publications (sous la dir. de Segond F.), Paris, 2002, pp. 223-250

Chanier, T. 1998. Relations entre le TAL et l'ALAO ou l'ALAO un "simple" domaine d'application du TAL ?, Communication: International conference on natural language processing and industrial application (NLP+IA'98). août 1998, Moncton, Canada.

Chanier, T., 2000, Hypermédia, interaction et apprentissage dans des systèmes d'information et de communication : résultats et agenda de recherche, in *Apprendre une langue dans un environnement multimédia*, (sous la direction de L. Duquette et M. Laurier), Les éditions Logiques, Outremont (Québec), 2000, pp.53-89

Chanier, T. & Renié, D., 1995, Collaboration and computer-assisted acquisition of a second language, *Computer-Assisted Language Learning* 8, 1: 3-30.

Chanier, T. & Selva, T., 2000. Génération automatique d'activités lexicales dans le système ALEXIA, *Sciences et Techniques Educatives* 7, 2 : 385-412. Paris : Hermes

Demaizière, F., Former les formateurs : nouveaux outils, nouvelles problématiques ? in *Apprendre une langue dans un environnement multimédia*, (sous la direction de L. Duquette et M. Laurier), Les éditions Logiques, Outremont (Québec), 2000, pp.305-322

Forestier, G., 2002, Plats-Formes pour l'enseignement des langues : le cas de MIRTO, Mémoire pour l'examen probatoire en Informatique, CNAM, Grenoble.

Kraif, O., Antoniadis, G., Echinard, S., Lebarbé, T., Loiseau, M., Ponton, C., 2004, NLP Tools for CALL : the Simpler, the Better, INSTILL 04, Venise, Italie, Juin 2004.

Loiseau, M., 2003, Vers la création d'une base de données de ressources textuelles indexée pédagogiquement pour l'enseignement des langues, Mémoire de DEA Sciences du Langage, Université Stendhal, Grenoble.

Meunier, L. E., 2000, La typologie des intelligences humaine et artificielle : complexité pédagogique de l'enseignement des langues étrangères dans un environnement multimédia » in *Apprendre une langue dans un environnement multimédia*, (sous la direction de L. Duquette et M. Laurier), Les éditions Logiques, Outremont (Québec), 2000, pp.211-253

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

Pothier, M. & Chanier, T. (éditeurs), 2000. [Numéro spécial Eurocall'99. Revue Apprentissage des Langues et Systèmes d'Information et de Communication \(ALSIC\).](#) vol.3, 1, juin.

Power, M., 2002, Générations d'enseignement à distance, technologies éducatives et médiatisation de l'enseignement supérieur, in *Journal of distance education (Revue de l'éducation à distance)*, ACED, Vol. 17, N°2, Ottawa (Canada), 2002

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, 1995

Rézeau, J., 1997, Concordances, Cédérom et Internet au service de l'enseignement du français aux adultes, *The Dong-Eui International Journal*, Corée, Juin 1997

Wyatt, D. H., 1987, Applying pedagogical principles to CALL courseware development, in *Modern Media in Foreign Language Education*, Wm. Flint Smith (dir.), Lincolnwood, IL : National Textbook, 1987, pp. 85-98