

De l'apport pertinent du TAL pour les systèmes d'ALAO : L'exemple du projet MIRTO

ANTONIADIS, Georges

Laboratoire LIDILEM, Université Stendhal
BP 25
38040 Grenoble cedex 9, France
Georges.Antoniadis@u-grenoble3.fr

1. Introduction

C'est aux USA, en 1809, que l'on trouve trace de la première tentative d'apprentissage/enseignement « mécanisé » des langues. A cette date, H. Chard dépose le brevet d'une machine à enseigner la lecture, machine dénommée *Mode of Teaching Reading*. Il faut néanmoins attendre plus de cent ans (1912) pour voir naître les premières théories et les plans ou les prototypes des premières machines « théorisées » (Edward Thorndike ; [Bruillard, 1997]).

La première machine « moderne » à enseigner est celle élaborée par Sidney Pressey [Pressey, 1927], en 1924. Il s'agit d'une machine, appelée *Drum Tutor*, capable de corriger les Questions à Choix Multiples (QCM), avec feedback immédiat (vrai/faux), et sauvegarde des traces des actions de l'apprenant. Certains, dont Burrhus Frederic Skinner [Skinner, 1968], lui reprochent d'avoir fondé sa machine sur des connaissances insuffisantes du phénomène d'apprentissage. Ils vont se focaliser sur les phénomènes d'apprentissage et créer l'enseignement programmé.

L'enseignement programmé s'appuie, en particulier pour Skinner, sur les résultats des travaux en psychologie du comportement. Le conditionnement opérant en tant que théorie du contrôle des mécanismes d'apprentissage, Skinner envisage la création d'une technologie scientifique de l'enseignement qui utilise l'enseignement programmé linéaire susceptible d'être dispensé par une machine à enseigner. Ainsi, dans sa machine construite en 1953, les exercices se trouvent sur un rouleau que l'apprenant fait défiler grâce à une molette. Les questions apparaissent dans une fenêtre ; l'apprenant inscrit sa réponse sur un espace blanc réservé à cet effet, puis il compare la réponse à la correction et actionne le levier pour passer à la question suivante. Selon Skinner, l'apprenant doit commettre moins de 10% d'erreurs lors d'une session d'apprentissage. Norman Crowder [Crowder, 1963] conteste la machine de Skinner et propose un système alternatif concernant le cheminement et le type d'exercices. Contrairement à Skinner qui cherche à limiter l'erreur, Crowder lui attribue une fonction importante et formatrice. Une fois la réponse donnée, si elle est bonne, l'apprenant passe à l'information suivante, dans le cas contraire, l'apprenant est dirigé vers des exercices de rattrapage pour ensuite revenir à l'exercice auquel il a échoué. L'enseignement programmé a ouvert de nouvelles pistes de recherche sur les méthodes et théories d'enseignement et d'apprentissage et a marqué le début de l'enseignement assisté par ordinateur.

L'enseignement assisté par ordinateur (EAO) voit le jour au début des années 60. Si au départ l'utilisation de l'ordinateur se limitait à automatiser ce qui était fait mécaniquement par les machines à enseigner, dans les années 70, avec les travaux sur les systèmes experts, apparaissent les premières tentatives de le rendre « intelligent ». Le terme lui-même (EAO) « évolue » au fil du temps. Ainsi, les termes EIAO (Enseignement Intelligemment Assisté par Ordinateur), tuteurs intelligents, systèmes coopératifs, environnements interactifs d'apprentissage sont successivement utilisés. Depuis quelques années le terme EIAH (Environnements Informatiques pour l'Apprentissage Humain) a pris la relève et il est actuellement le plus souvent utilisé par la communauté scientifique. Dans cette évolution terminologique, les années 70 sont marquées par la modélisation et l'utilisation des premiers micro-mondes, les années 80 par les tuteurs intelligents, les années 90 par les systèmes coopératifs et les environnements interactifs d'apprentissage, enfin, le début du 21^e siècle par le souci d'élaboration d'EIAH faciles d'utilisation par les apprenants et/ou les enseignants, incluant toute sorte de connaissances.

Presque depuis le début de ces recherches, les objectifs à atteindre sont clairement définis. Ils sont principalement centrés autour de l'apprenant¹. En fonction de ces objectifs, un bon didacticiel ou environnement d'apprentissage doit permettre de :

- dialoguer avec l'apprenant en langage naturel ;
- sélectionner la suite de ce qui doit être enseigné ;
- anticiper, diagnostiquer et comprendre les erreurs de l'apprenant ;
- améliorer les stratégies d'enseignement et être capable de le modifier en fonction de l'apprenant.

Ces buts, fort ambitieux, sont toujours d'actualité et constituent le thème qui sous-tend une bonne partie des recherches actuelles du domaine.

Depuis le début, l'enseignement/apprentissage des langues fait partie des préoccupations et visées de l'EAO. Néanmoins, assez vite, le sigle d' ALAO (Apprentissage des Langues Assisté par Ordinateur) voit le jour. Ce nouveau domaine définit, petit à petit, sa propre problématique et construit ses outils spécifiques. A notre connaissance, cette situation est unique² ; elle a permis l'émergence de bon nombre de travaux et produits spécifiques, ainsi que de revues et manifestations scientifiques où les travaux du domaine sont présentés. Comme François Mangenot [Mangenot, 1997] le signale, les premiers produits d'ALAO (mais aussi une grande partie d'entre eux par la suite) s'intéressent principalement à des exercices sous différentes formes et avec des objectifs variés. L'essor de la microinformatique, du multimédia et des réseaux a accentué cette tendance et il a permis le partage des produits et ressources élaborés. En même temps, diverses tentatives d'utilisation des résultats d'autres disciplines ou domaines voient le jour. Elles ont permis de faire plusieurs pas vers la réalisation des objectifs (ci-dessus) assignés aux produits informatiques d'apprentissage ou d'enseignement.

Les premières tentatives d'utilisation du TAL (Traitement Automatique des Langues) pour l'apprentissage des langues datent des années 80, avec un pic d'activité durant la période 1985-1995. Ce constat est partagé par Jung [Jung, 2005], et il est attesté par bon nombre de symposiums et de travaux européens [Swartz & Yazdani, 1992], nord américains [Holland & al., 1995] ou français [Chanier & al., 1993]. Cela correspond au plein développement des ordinateurs personnels et à la première tentative d'insertion des TIC dans les dispositifs pédagogiques. Pendant cette période, les modèles et prototypes d'analyse et de génération des langues se diversifient ; ils serviront de point de départ pour des tentatives de création de systèmes d'apprentissage des langues. On trouve trace de ces travaux dans toutes les revues traitant de l'apprentissage des langues : CALICO, Language Learning and Technology, ReCALL, CALL, ALSIC, System.

Le plus souvent, ces tentatives, tâtonnantes et maladroites au départ, trouvent assez vite leurs limites ; des gros investissements en terme de ressources humaines sont nécessaires pour dépasser le stade de prototype de laboratoire très restreint. D'autre part, la problématique apprentissage / acquisition des langues secondes (non maternelles) est souvent absente, ce qui empêche tout travail en collaboration entre informaticiens / spécialistes du TAL et didacticiens des langues / pédagogues.

Durant la fin des années 1990, cette tendance commence à être corrigée ; des rencontres interdisciplinaires vont pouvoir se dérouler et des travaux en collaboration voient le jour. Le paradigme du système d'information et communication, avec lequel l'apprenant doit interagir, s'enrichit de quatre composantes : une gérant les connaissances linguistiques de référence, une autre le dialogue homme/machine, une troisième les connaissances de l'apprenant et/ou ses erreurs, enfin, une quatrième relative aux interventions didactiques du système. Le paradigme est toujours d'actualité, l'ALAO devient attrayante pour les didacticiens / pédagogues des langues. Le développement de l'Internet contribue à la mise œuvre du paradigme, les apprenants peuvent facilement utiliser à distance des matériaux riches et diversifiés, mais aussi être en contact quasi-continu avec des locuteurs, natifs ou non, de la langue objet de l'apprentissage.

Dans ce contexte, les réalisations sont mesurées et les avancées se font à petits pas. Tous les aspects des systèmes d'ALAO sont « revisités », redéfinis, confrontés aux exigences et aux réalités de l'apprentissage des langues. Les travaux actuels (les nôtres compris) essaient d'utiliser toutes les possibilités « réalistes » des résultats du TAL en les adaptant aux objectifs visés. Souvent, des programmes de traitement de la langue spécifiques pour les systèmes d'ALAO sont élaborés ; ainsi, le TAL n'est plus qu'un outil pour la

construction de systèmes d'ALAO, il s'enrichit du regard du didacticien sur la langue, de sa problématique, de ses besoins et de ses contraintes.

Les deux formes de la langue, écrite et orale, sont concernées ; néanmoins, l'état d'avancement du traitement de la langue écrite comparativement à celui de la parole influence directement le nombre, et, souvent, la qualité des solutions et des systèmes proposés. Ainsi, les travaux et systèmes concernant la forme écrite sont nettement majoritaires et touchent un nombre plus important de facettes et situations de l'apprentissage des langues.

Implicitement ou explicitement, une idée sous-tend tous ces travaux : l'apprentissage des langues assisté par ordinateur demande des travaux pluridisciplinaires en partenariat, sur un pied d'égalité entre les différentes disciplines. Seule l'association intime des problématiques de chaque discipline ou domaine concernés (Didactique des langues, Informatique, Linguistique, TAL) permet de proposer des solutions et des systèmes opérationnels, dignes d'intérêt pour les apprenants et capables de leur offrir une plus-value didactique par rapport aux méthodes et systèmes classiques. Dans ce travail en association, la tâche essentielle pour chaque discipline ou domaine est la définition et la mise en œuvre de la partie du terme « apprentissage des langues assisté par ordinateur » qui la (le) concerne ; Ainsi, c'est la Didactique qui est la plus apte à déterminer et rendre opérationnel le terme « apprentissage », la Linguistique le terme « langue », l'Informatique celui « assisté par ordinateur ». Chacune de ces définitions doit être confrontée aux autres, adaptée aux contraintes des définitions partenaires, apporter sa part à l'élaboration de la solution globale.

L'utilisation des corpus pour l'apprentissage des langues est un champ d'application majeur, à notre avis, du TAL et, bien sûr, de l'informatique. L'informatique a permis aux enseignants de langues d'utiliser la richesse des corpus (la diversité des situations langagières qu'ils contiennent), de disposer d'une source pratiquement intarissable d'exemples de la langue « réelle », celle qu'il faut apprendre, celle que les apprenants auront à utiliser lors des situations communicatives. Le TAL, parce qu'il permet de rechercher non seulement des chaînes de caractères, mais aussi des formes linguistiques se rapportant à des lemmes, des morphèmes, des traits morphosyntaxiques, des relations fonctionnelles, des constructions complexes, etc. décuple les possibilités d'exploitation des corpus. Par la construction et le traitement des corpus d'apprenants, il permet également d'automatiser l'analyse des productions des apprenants eux-mêmes. L'approche de l'apprentissage basé sur l'exemple (*data driven learning*), a donné lieu à un grand nombre de travaux, des ressources et des systèmes [Tribble & Barlow, 2001 ; Granger, 2002 ; Granger & al., 2007 ; Antoniadis & al., 2009]. Le développement de l'Internet a permis le partage et la diffusion de ces ressources et systèmes ce qui, sans doute, a fortement contribué à l'extension de l'utilisation des corpus pour l'apprentissage des langues. Plusieurs systèmes d'ALAO utilisent et exploitent des corpus, bruts ou annotés ; le soin apporté à leur constitution et annotation comme la pertinence de leur exploitation détermine souvent la qualité du système qui les utilise.

Plus de vingt ans après le début des travaux en « TAL et ALAO », force est de constater que si un certain nombre de prototypes ou de systèmes expérimentaux existent ou ont existé [Selva, 2002 ; Brun & al., 2002 ; Antoniadis & Chanier, 2005], les systèmes commercialisés sont encore pratiquement inexistantes. L'avancée insuffisante des recherches du domaine n'est qu'une explication partielle. À notre avis, deux facteurs sont la cause principale de cet état : 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. Non standardisés, ces derniers restent encore difficilement utilisables en l'état et demandent souvent d'importantes adaptations pour être déployés à profit dans le cadre de l'ALAO.

2. Notre démarche : la stratégie « moins-disante »

Notre démarche se positionne dans la problématique énoncée plus haut [Antoniadis, 2004]. Elle est basée sur deux constats :

- La technologie du traitement automatique de la langue propose des méthodes et des traitements susceptibles d'intéresser l'ALAO dans deux directions : l'amélioration des capacités et performances de bon nombre de logiciels issus de ce domaine ; la possibilité pour lui d'étendre sa problématique, en considérant de nouveaux problèmes pour lesquels l'informatique et/ou la didactique des langues ne sont pas capables, par essence, d'apporter des solutions. Ainsi, l'adjonction du TAL permet d'imaginer des solutions automatiques, pédagogiquement pertinentes, pour des problèmes comme : l'indexation pédagogique des ressources ; la recherche de ressources en fonction de critères pédagogiques ; la possibilité d'utiliser des textes variés pour les exercices « informatisés » ; la détection automatique ou semi-automatique des fautes des apprenants et la génération automatique de feedbacks pédagogiques précis en fonction de chaque faute ; la personnalisation du parcours d'apprentissage de chaque apprenant ; etc.
- Sauf dans des contextes particuliers, la technologie du TAL ne produit que rarement, actuellement, des résultats sûrs et/ou complets à 100%. Son utilisation dans le cadre de l'ALAO demande la prise en compte de cette particularité et de ses conséquences. Ainsi, par exemple, l'évaluation automatique de textes libres est un problème qui ne peut pas actuellement trouver de solution satisfaisante. Par contre, la création quasi-automatique de certains types d'exercices (exercices lacunaires, remises en ordre, glisser-déposer, appariements, etc.) ou la détection et la génération automatique de feedback associé pour certaines fautes, font partie des possibilités d'exploitation à bon escient des techniques et résultats du TAL. En ce sens, il est essentiel que les « niches applicatives » (pour le TAL) de l'ALAO soient imaginées préalablement en tenant compte des limites technologiques du TAL.

Ces deux constats nous ont conduits à adopter une démarche « prudente » et une stratégie « moins-disante ». Elles consistent à ne considérer l'apport du TAL que pour des problèmes précis de l'ALAO, en tenant compte, chaque fois, des exigences pédagogiques et en s'assurant que le degré de fiabilité des procédures et résultats du TAL sont capables de les satisfaire. Dans une telle approche, nous privilégierons, le plus souvent, « l'aide à... », et nous exploiterons toutes les résultats intermédiaires utiles des traitements TAL. Assez souvent, ces derniers doivent être partiellement repensés pour mieux s'adapter aux exigences pédagogiques.

Destinés principalement à des non spécialistes (enseignants des langues, apprenants), les logiciels issus de notre démarche doivent, non seulement ne pas demander de compétences spécifiques en informatique ou en TAL pour leur utilisation, mais, surtout, offrir la possibilité à leurs utilisateurs (pédagogues ou apprenants) d'utiliser leur terminologie, de manipuler leurs concepts. Ce dernier point fait l'objet d'une problématique nouvelle ; son approche ne peut s'envisager que par le biais d'un travail en collaboration « réelle » entre pédagogues et informaticiens spécialistes du TAL. Plus que de notre ligne de conduite, il s'agit là de notre « credo scientifique » essentiel.

L'utilisation spécifique des traitements et résultats du TAL en fonction des exigences pédagogiques demande, d'une part, la description et modélisation de chaque objectif pédagogique visé, d'autre part, pour des raisons d'efficacité, la mise en commun de traitements TAL pour la réalisation des objets (objectifs) pédagogiques. En ce sens, notre approche ne peut être que déclarative et la conception du système visé que fortement modulaire. Ces choix sont d'autant plus essentiels que le système visé se veut un cadre d'expérimentation des apports du TAL pour l'ALAO.

Notre démarche s'est concrétisée par l'élaboration d'une plateforme, MIRTO, cadre de nos recherches dans le domaine. Nous présentons par la suite les principes, la structure, les fonctionnalités et l'apport pour l'ALAO de cette plateforme. Plusieurs travaux sont en cours actuellement la concernant. Nous les évoquerons au fur et à mesure de la description de la plateforme.

3. Le projet MIRTO

3.1. Les objectifs

Dans la lignée de la problématique précédemment exposée, le projet MIRTO (Multi-apprentissages Interactifs par des Recherches sur des Textes et l'Oral) 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. Dans un premier temps, la forme écrite de la langue a été privilégiée ; les travaux concernant l'exploitation de sa forme orale viennent de débiter. Plus qu'un produit fini, MIRTO se veut un cadre de recherches pour l'élaboration d'une méthodologie d'élaboration de produits ALAO utiles et performants, mais aussi un outil de mise en œuvre de solutions didactiques pour l'apprentissage des langues.

Le projet MIRTO vise la réalisation d'une plateforme d'enseignement des langues et des matières qui utilisent du matériau langagier (comme, par exemple, la Linguistique). MIRTO ambitionne de mettre au service des enseignants des ressources et des outils issus du TAL afin de leur permettre la conception de scénarios pédagogiques exploitant la diversité et la richesse des corpus textuels. Les scénarios ainsi élaborés seront utilisés par les apprenants à distance ou en présentiel. Le système devrait offrir la possibilité de calculer le « profil » de chaque apprenant (principalement par exploitation de ses réponses et de son parcours), de fournir des explications de ses erreurs et d'adapter les scénarios aux connaissances/compétences dans la matière enseignée pour chaque apprenant.

MIRTO se veut un environnement et un outil au service des enseignants de langues, capable de leur permettre d'exprimer et de mettre en œuvre des propositions et des solutions pour des problèmes relatifs à l'enseignement des langues. Nous partons du principe que ces propositions et solutions ne peuvent se concevoir qu'à travers (et à l'aide) des concepts du domaine. En ce sens, tout outil utilisé doit, non seulement ne pas imposer des contraintes spécifiques, mais permettre la manipulation aisée de ces concepts et faciliter leur mise en œuvre. Nous concevons MIRTO comme un outil apte à « mettre en avant » la problématique de la didactique des langues et « mettre en arrière » les contraintes informatiques ; un outil capable de manipuler la langue comme un système de signes, propre à prendre en considération la sémantique qui leur est associée. Notre approche est résolument orientée utilisateur, dans la mesure où la plateforme est destinée à des enseignants de langue, qui, à 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, et seuls les aspects didactiques doivent être visibles et disponibles.

L'enseignement comme l'apprentissage d'un système langagier se réalise, le plus souvent, par l'exploitation et l'utilisation des produits de ce système, c'est-à-dire d'exemples (textuels ou oraux) que le système est capable d'engendrer. L'emploi d'un exemple (à but d'illustration ou d'exercice) suppose que celui-ci soit attesté ou qu'il soit « attestable ». Les corpus de textes actuellement disponibles sont fort nombreux et variés ; ils peuvent constituer une source intarissable d'exemples attestés, à condition que l'on puisse avoir accès et qu'il soit possible d'en puiser des exemples en fonction de critères pédagogiques préétablis. Un des objectifs de MIRTO est d'offrir la possibilité de choisir un ou plusieurs textes (ou morceaux de textes) à partir d'objectifs pédagogiques indiqués par l'enseignant, dans un corpus de textes indexés pédagogiquement. L'enseignant peut néanmoins choisir d'utiliser son propre texte. Dans ce cas, et si il le désire, le texte enrichira le corpus constitué jusque là, après une étape d'indexation pédagogique.

3.2. Structure fonctionnelle de MIRTO

Centrée autour de l'utilisateur, la conception de la plateforme MIRTO repose sur quatre principes le concernant directement ou indirectement :

- La plateforme doit offrir la possibilité de manipuler et de mettre en œuvre des concepts pédagogiques d'apprentissage/enseignement des langues.
- Les propriétés intrinsèques de la langue, objet d'apprentissage/enseignement, doivent être considérées et prises en charge par la plateforme ; la langue n'est pas qu'une suite de signes, mais un double système (signes, sens) ; la plateforme doit pouvoir traiter les (des) relations et les (des) propriétés de cette réalité.

- Les traitements informatiques et les procédures TAL mis en œuvre par la plateforme doivent être transparents pour l'utilisateur qui, le plus souvent, n'a pas de compétences dans ces domaines.
- La conception d'activités pédagogiques (les résultats de la plateforme) doit demander un temps réduit ; l'objectif visé étant la conception instantanée ou quasi-instantanée.

La plateforme MIRTO s'adresse principalement aux enseignants de langues. Pour elle, l'enseignant est au centre du dispositif et tout doit être organisé pour faciliter ses tâches (figure 1).

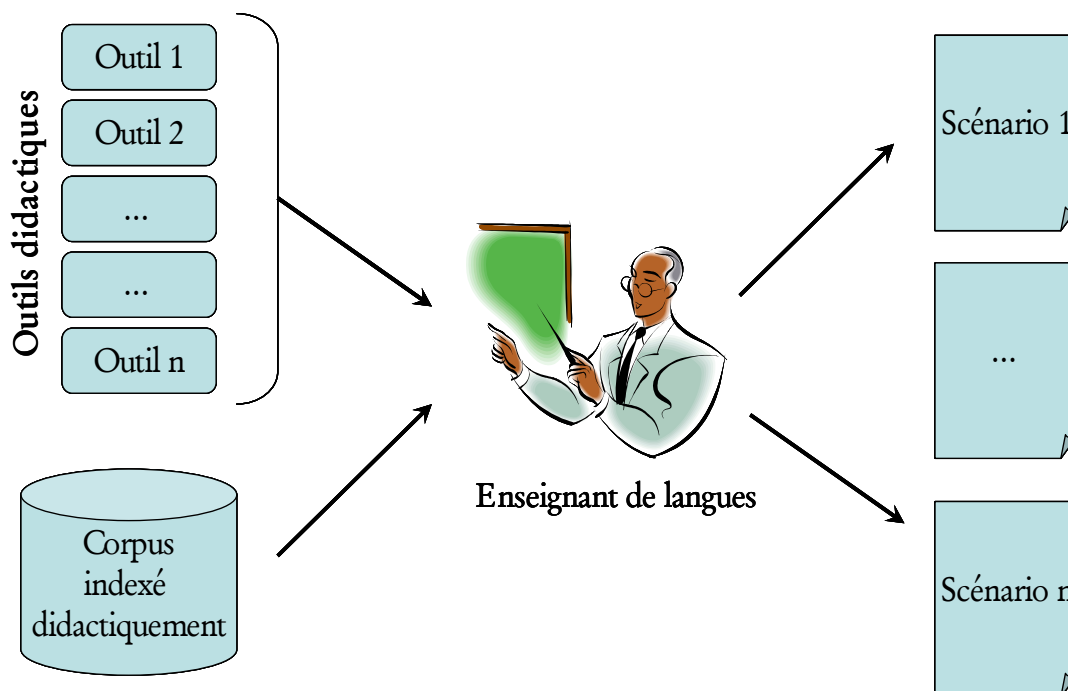


Figure 1 : Structure de MIRTO

La plateforme met à la disposition des enseignants un certain nombre d'outils didactiques, ainsi que des procédures et interfaces simples, pour qu'ils puissent élaborer automatiquement, ou semi-automatiquement, des scénarios pédagogiques. Ces scénarios peuvent utiliser des textes issus d'un corpus de textes indexés pédagogiquement associé à la plateforme. Le cas échéant, l'enseignant peut fournir son propre texte. Une fois traité, le texte peut être indexé pédagogiquement et ajouté au corpus. Très peu de travaux existent concernant l'indexation pédagogique des documents. On trouvera dans [Loiseau M., 2009] la problématique de l'indexation pédagogique des documents, et notre approche, dans le cadre du projet MIRTO, basée sur les concepts de « facette », « prisme » et « vue » didactiques des documents, ainsi que le modèle de la base de données associée.

L'enseignant de langues perçoit de MIRTO (pour l'essentiel, via une interface spécifique) un ensemble de « briques pédagogiques », appelés *scripts*. Un script est une chaîne de traitements opérés par des logiciels, issus essentiellement du TAL, généralement sur un texte qui peut être interchangeable. L'enseignant

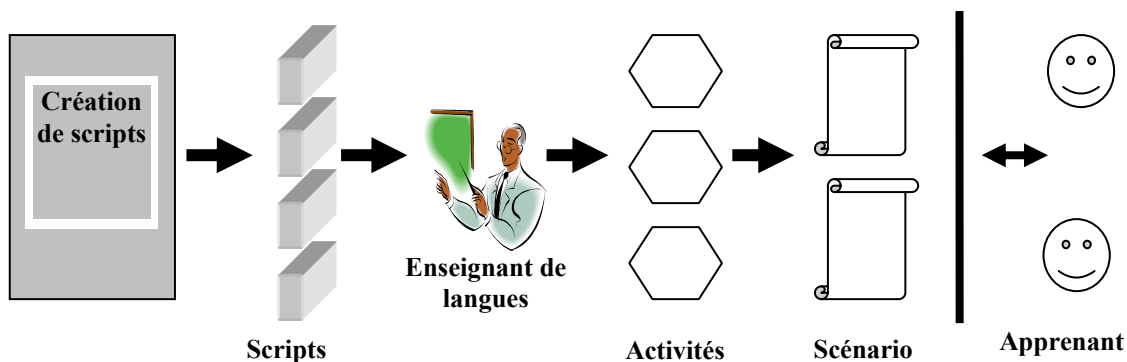


Figure 2 : Schéma fonctionnel de MIRTO

utilise les scripts pour élaborer des objets pédagogiques plus complexes, les *activités*. Une activité est l'application d'un script pour un contexte pédagogique défini par l'enseignant. Une activité peut être créée à partir d'une idée « originale » de l'enseignant, ou élaborée par modification d'une activité existante. Les activités sont des objets didactiques autonomes, et peuvent, donc, être utilisées par les apprenants comme tels, au sein de la plateforme ou en dehors de celle-ci, si elles sont exportées par l'enseignant. Elles constituent les grains des objets pédagogiques finaux, les *scénarios*.

Un scénario est une succession d'activités dans un but pédagogique précis, défini par l'enseignant. La succession d'activités peut être linéaire ou conditionnée par les résultats obtenus à l'activité précédente, lorsque ceux-ci peuvent être évalués automatiquement. Dans ce dernier cas, l'enseignant doit également indiquer les critères conditionnels de passage d'une activité à la suivante. Comme pour les activités, un scénario peut être original, ou issu de modifications d'un scénario existant.

L'apprenant ne perçoit de MIRTO que les scénarios qu'il sera amené à utiliser en présentiel ou à distance), en fonction des indications de l'enseignant, auteur de chaque scénario. La figure 2 montre le fonctionnement de MIRTO ainsi que la place de ses deux types d'utilisateurs.

3.3. Architecture de MIRTO

Quatre niveaux hiérarchiques, associés à la base de textes indexés pédagogiquement, structurent MIRTO : fonctions, scripts, activités et scénarios. La figure 4 illustre cette structure. La manipulation et la création d'objets à chaque niveau demande des compétences spécifiques, variables d'un niveau à l'autre. Le tableau de la figure 3 précise, pour chaque niveau, le type d'utilisateurs ainsi que les opérations possibles qui sont associées.

Niveau	Opération	Utilisateur
Fonctions	Conception	Spécialiste TAL
Scripts	Conception	Spécialiste TAL + Enseignant des langues
Activités	Conception	Enseignant des langues
Scénarios	Conception	Enseignant des langues
	Utilisation	Apprenant

Figure 3 : Type d'utilisateur et opérations associées

3.3.1. Les fonctions

Le niveau le plus bas du système est celui des *fonctions*. Une fonction correspond, dans la plupart des cas, à un résultat d'un processus TAL basique comme, par exemple, la liste des phrases ou des lemmes d'un texte, la reconnaissance de la langue d'un texte, etc. Les programmes TAL utilisés à ce niveau sont issus du commerce ou des laboratoires de recherche. Etant donné la nature technique des fonctions et leur indépendance de toute application pédagogique, ce niveau n'est pas visible par les enseignants de langues, utilisateurs de la plateforme.

3.3.2. Les scripts

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 des traitements TAL standard, tels que l'identification de la langue du texte, l'étiquetage morphosyntaxique, la lemmatisation, l'interrogation de dictionnaires, etc. 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 cœur de l'architecture de MIRTO et ils constituent le point de rencontre entre les potentialités de la technologie TAL et les besoins de la pratique pédagogiques.

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 ; elle lie les fonctions d'identification de la langue, de tokenisation et d'analyse morphologique.

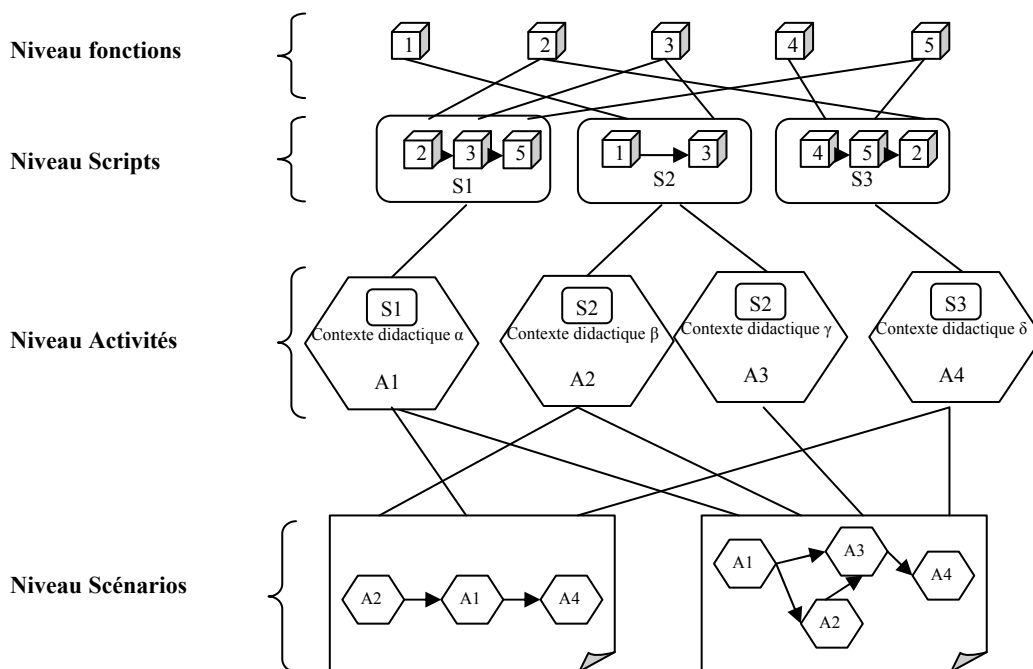


Figure 4 : Architecture de MIRTO

Le plus souvent, un script doit être paramétré. La présence (ou l'absence) de texte sur lequel la séquence de programmes sera appliquée constitue le paramètre le plus habituel. Chaque script ainsi créé est identifié par un nom qui doit suggérer les exploitations pédagogiques potentielles.

La création de scripts demande des compétences aussi bien en informatique qu'en TAL. En général, la création s'effectue en collaboration avec un enseignant de langues. Le rôle de ce dernier est de définir le concept pédagogique que le script doit exprimer. Un script créé est partagé par tous les enseignants

utilisateurs de MIRTO ; il constitue une brique pédagogique pouvant être utilisée, indéfiniment, dans plusieurs situations pédagogiques. A titre d'exemple, le script qui permet de créer automatiquement des exercices à trous, peut être utilisé pour générer de tels exercices pour n'importe quelle catégorie grammaticale à partir de n'importe quel texte.

3.3.3. Les activités

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, etc. Elle se réalise par la construction d'un espace de travail pour l'apprenant lui permettant d'atteindre le but visé. Elle correspond à ce qui est traditionnellement désigné comme exercice.

D'un niveau purement didactique, la conception d'activités est une tâche opérée par les enseignants des langues, à l'aide d'un outil spécifique de MIRTO : l'éditeur d'activités. L'éditeur d'activités est un environnement de conception 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. 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 et, éventuellement, indépendamment de la plateforme MIRTO. Pour ce faire, chaque activité créée peut être exportée par son créateur, sous forme HTML, et proposée à des apprenants via les interpréteurs HTML courants.

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. Comme nous l'avons indiqué, ces erreurs sont possibles, les résultats des traitements du TAL n'étant pas toujours fiables à 100%. En cas d'erreur, l'enseignant apporte une correction manuelle, le plus souvent en neutralisant la partie incriminée du texte. Parmi les étapes nécessaires pour la création d'une activité, deux, à notre avis, demandent une attention particulière : le paramétrage et l'évaluation automatique lorsque celle-ci est possible.

Le paramétrage des activités

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.

Pour MIRTO, l'enseignant de langues étant au centre du système, la « traduction » en termes pédagogiques des concepts et de la terminologie du TAL et de l'informatique constitue un des ses soucis majeurs. MIRTO vise à offrir aux enseignants de langues un environnement de travail « avenant », capable de leur permettre de manipuler des concepts et des objets pédagogiques en rapport avec leur discipline, la didactique des langues.

La traduction pédagogique ci-dessus est décrite dans [Ans & Papadima, 2005]. Elle concerne quatre langues actuellement : le français, l'anglais, l'italien et le grec. Elle est opérationnelle pour toute activité

CREATION D'UNE ACTIVITE : Texte à Trous

Choix de la langue

Langue Français ▾

Choix des paramètres de création de l'exercice

Consigne Conjuguez les verbes indiqués.

Mots à escamoter

- adjectif qualificatif féminin ▾ singulier ▾
- adverbe
- déterminant masculin ▾ pluriel ▾ article indéfini ▾
- négation
- nom --genre-- ▾ --nombre-- ▾ --type-- ▾
- préposition
- pronom masculin ▾ singulier ▾ interrogatif ▾ 2e pers. sg ▾
- verbe --temps-- ▾ --mode-- ▾ --personne-- ▾

Options Affichage du lemme dans les trous ▾

Nom du fichier d'entrée Texte_1

Nom de la sortie Activité_2

Figure 5 : Ecran de paramétrage d'une activité

nécessitant un paramétrage morphosyntaxique. Les paramètres et les valeurs associées sont fonction de la langue. La figure 5 montre un tel paramétrage concernant la création d'exercices lacunaires pour l'apprentissage du français. La mise en œuvre de la traduction pédagogique s'opère à l'aide d'une base de données. Elle concerne actuellement le français et l'anglais. On trouvera la description et la structure de la base dans [Clorio & Hoang, 2006].

L'évaluation automatique des résultats

Dans le cadre d'une activité, la production de l'apprenant peut prendre des formes diverses : une case à cocher, un mot à écrire ou à choisir, une phrase à constituer, voire un texte à rédiger. La correction automatique de phrases ou de textes est un problème complexe. Les techniques actuelles du TAL ne peuvent donner de résultats utilisables que dans des contextes bien contraints. Dans un contexte libre, les résultats du TAL ne sont que partiellement fiables, en particulier 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.

Dans le cadre de MIRTO, nous pouvons associer une évaluation automatique à certaines activités, celles pour lesquelles les traitements TAL conduisent à des résultats relativement fiables et exploitables pour les aspects qui nous intéressent. C'est le cas, par exemple, pour les exercices lacunaires. Notre but n'est pas la détection « simple » des fautes qui ne peut conduire que à des rétroactions de type « c'est faux ! », « essaie encore une fois », etc. De telles rétroactions, pédagogiquement « vides », n'apportent aucune piste pour la réponse juste, aucun repère ni cadre pour la structuration des connaissances de l'apprenant. Pour MIRTO, en plus de la détection d'une réponse erronée, nous nous efforçons à calculer le différentiel de celle-ci avec la réponse juste. Ainsi, pour une faute, par exemple, concernant un verbe, on pourra discerner que le temps utilisé est correct mais non la personne (ou l'inverse), et générer des rétroactions

en rapport. L'avantage d'une telle approche est de pouvoir prévoir et proposer des remédiations spécifiques et à la carte.

L'évaluation automatique des résultats des apprenants fait actuellement l'objet de travaux avec les hypothèses que nous venons de présenter. On trouvera dans [Blanchard, 2007] une description de l'approche et des premiers résultats obtenus.

3.3.4. Les scénarios

Ce niveau permet aux enseignants de définir des séquences d'activités. Une telle séquence, appelé *scénario*, correspond à un (des) objectif(s) pédagogique(s) défini(s) par l'enseignant. La succession d'activités d'un scénario peut être linéaire ou conditionnée par les résultats obtenus à l'activité précédente, lorsque ceux-ci peuvent être évalués automatiquement. Lorsque l'évaluation automatique est possible, l'enseignant doit également indiquer les critères conditionnels de passage d'une activité à la suivante. Selon son parcours dans un scénario donné, un apprenant pourra être redirigé vers une activité de remédiation, retenter la même activité sur un autre texte ou encore simplement avancer dans le scénario. 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. Comme pour les activités, un scénario peut être original, ou issu de modifications d'un scénario existant. L'enseignant est propriétaire des scénarios qu'il crée. Il peut les rendre publics selon la même procédure que pour les activités.

La création de scénarios s'opère par l'enseignant à l'aide d'un outil visuel spécifique, l'éditeur de scénarios. L'éditeur de scénario est un environnement de conception permettant de visualiser et de manipuler des objets pédagogiques tels que les activités et les scénarios, propres à chaque enseignant ou publics. Tout scénario en cours de création peut être testé par son auteur et modifié en fonction des résultats des tests. Si l'enseignant décide de le rendre public, l'indexation est plus fine et elle est accompagnée d'un descriptif concernant principalement les objectifs pédagogiques visés. La définition précise des caractéristiques des scénarios, en particulier non linéaires, ainsi que du contenu et de l'exploitation de l'historique, est en cours. Elle fait l'objet de travaux qui viennent de débiter. Elle devrait mettre en application les hypothèses ci-dessus et utiliser une première version de l'éditeur de scénarios déjà réalisée pour l'essentiel.

3.4. Les traitements et les outils linguistiques de MIRTO

Comme nous l'avons déjà signalé, une des caractéristiques de MIRTO est d'utiliser, tant que faire se peut, des logiciels de TAL déjà existants, qu'ils proviennent du commerce, du secteur libre ou qu'ils soient le produit de recherche de différents laboratoires.

L'ensemble de logiciels TAL utilisés comporte, actuellement, des logiciels pour le découpage des textes en formes et leur traitement fréquentiel, pour l'identification de la langue des textes, l'étiquetage morphologique et la tokenization, des dictionnaires de traduction, etc. Parmi ces logiciels, ceux concernant l'analyse morphologique sont, à ce jour, à la base de la plus grande partie de scripts mis en place, et, par conséquent, de la plus grande partie des activités mises en oeuvre.

Les langues actuellement prises en charge par MIRTO sont, principalement, le français et l'anglais; des scripts concernant l'italien ont été mis en place à titre expérimental. A part l'anglais, nous visons, dans une première phase, la prise en charge des langues romanes, du moins celles pour lesquelles des logiciels de traitement automatiques sont disponibles avec une licence d'exploitation libre.

Dans une telle démarche, des langues pour lesquelles des produits TAL sont rares ou indisponibles en accès libre ne peuvent pas être raisonnablement prises en charge par MIRTO. Mises à part les langues dites "rares", tel est le cas pour d'autres, bien importantes, comme par exemple l'arabe. Pour ce dernier, si la bibliographie concernant son traitement automatique est relativement riche, les produits opérationnels sont rares et/ou inachevés.

En fonction de cette situation, et désireux d'enrichir MIRTO des possibilités de création d'activités pédagogiques pour l'arabe, nous avons entrepris, en collaboration avec l'université de Monastir en Tunisie (laboratoire UTIC), l'élaboration d'outils TAL pour le traitement automatique de cette langue. Dans un premier temps, notre but est la création d'un dictionnaire étiqueté de la langue arabe (le plus complet possible) ainsi que l'élaboration d'un analyseur morphologique robuste de textes arabes. Ces

outils, pouvant servir pour bon nombre d'applications, doivent s'intégrer dans la plateforme MIRTO, et, en ce sens, tenir compte des applications pédagogiques visées.

L'élaboration du dictionnaire et de l'analyseur morphologiques de textes arabes font actuellement des travaux en collaboration. On trouvera dans [Maraoui, 2009] la description concernant la construction du dictionnaire, sa structure ainsi que des exemples d'activités pédagogiques qui peuvent en être, ou qui en sont, issues. Notre approche pour l'analyseur de textes arabes est décrite dans [Mars & al., 2007] ainsi que la description de la partie réalisée jusqu'ici.

4. Conclusion

Le champ couvert par l'application du TAL à l'ALAO est vaste. Il recouvre de nombreux sujets dont certains demandent encore à être explicitement formalisés, en particulier ce qui touche au traitement de la langue orale. En ce qui concerne le traitement de la langue écrite, les nombreuses thématiques abordées sont autant de défis auxquels les chercheurs sont confrontés à l'heure actuelle. Une chose apparaît de manière sûre : si les chercheurs qui travaillent à l'intersection du TAL et de l'ALAO veulent avoir une chance de voir leurs travaux mis en pratique dans l'enseignement des langues, il est indispensable qu'ils accordent une grande importance à la scénarisation pédagogique, aspect, encore, souvent marginal. La clef réside dans le travail pluridisciplinaire et la mise en commun de pratiques et de techniques. Néanmoins, tout laisse à penser que l'intégration du TAL à l'ALAO qui permet de prendre en considération les caractéristiques de la langue de référence et celle de l'apprenant, se double actuellement de l'intégration de la didactique des langues, seul moyen pour que l'objectif d'apprentissage soit pleinement atteint par le biais de systèmes et d'outils ALAO conviviaux et performants. Nos travaux, dans le cadre du projet MIRTO, s'inscrivent pleinement dans cette perspective.

Références bibliographiques

- ANS M., PAPADIMA O. (2005). Extraction de phrases satisfaisants à des critères morphosyntaxiques définis par un utilisateur. *Mémoire du master 2 « Modèles et traitements pour les industries de la langue : parole, écrit, apprentissage »* de l'université Stendhal de Grenoble.
- 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.
- ANTONIADIS G., CHANIER, T. (eds.), (2005). *Numéro thématique « TAL et apprentissage des langues », ALSIC*, vol. 8, n° 2, novembre 2005
- ANTONIADIS G., GRANGER S., KRAIF O., MEDORI J., PONTON C., ZAMPA V. (2009). Integrated Digital Language Learning. In: *Balacheff, N., Ludvigsen, S., de Jong, T., Lazonder, A. & Montandon, L. (eds.) Technology-Enhanced Learning, Principles and Products, Springer*, pp. 89-103
- BLANCHARD A. (2007). Analyse morphologique des réponses d'apprenants. *Rencontre des Etudiants Chercheurs en Informatique et Traitement Automatique des Langues -RECITAL 2007*, 5-8 juin 2007, Toulouse
- BRUILLARD E. (1997). *Les machines à enseigner*. Paris : Hermès.
- 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.
- CHANIER T., RENIE D., FOUQUERÉ C. (dir.) (1993). *Actes du colloque SCIAL '93 (sciences cognitives, informatique et apprentissage des langues)*. Clermont-Ferrand : Université Blaise Pascal. Préface et sommaire en ligne à <http://archive-edutice.ccsd.cnrs.fr/edutice-00001148>
- CLORIO R., HOANG B. (2005). Conception et réalisation de la base de données traitsmorphologiques et de l'interface qui la manipule pour la plateforme MIRTO. *Mémoire de stage du master 2 « Information, Cognition et Apprentissages, spécialité Informatique, Homme et Société » de l'université Pierre Mendès-France de Grenoble*.
- CROWDER N. (1963). On the difference between linear and intrinsic programming. *Phi Delta Kappan*, 44, pp. 250-254.

- GRANGER S.** (2002). A bird's eye view of learner corpus research. In S. Granger, J. Hung & S. Petch-Tyson (eds.) *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching, Language Learning and Language Teaching 6*. Benjamins: Amsterdam & Philadelphia, pp. 3-33.
- GRANGER S., KRAIF O., PONTON C., ANTONIADIS G., ZAMPA V.** (2007). Integrating learner corpora and natural language processing: a crucial step towards reconciling technological sophistication and pedagogical effectiveness. *ReCALL*, vol. 19, part.3., pp. 252-268.
- HOLLAND V.M., KAPLAN J.D., SAMS M.R.** (dir.) (1995). *Intelligent Language Tutors*. Mahwah, Lawrence Erlbaum Associates, NJ.
- JUNG, U.O.H.** (2005). CALL: past, present and future - a bibliometric approach. *ReCALL*, vol. 17, 1. pp. 4-17.
- LOISEAU M.** (2009). *Elaboration d'un modèle pour une base de textes indexée pédagogiquement pour l'enseignement des langues*. Thèse de Doctorat, université Stendhal de Grenoble, soutenue le 9 décembre 2009.
- MANGENOT F.** (1997). Synthèse de trois cours de FLE sur CD-ROM. *Les cahiers de l'asdifle, multimédia et langue étrangère, actes des 19e et 20e rencontres, 1997*, pp.79-88.
- MARAOUI M.** (2009). *Elaboration d'un dictionnaire multifonction, à large couverture, de la langue arabe. Application aux systèmes d'ALAO*. Thèse de Doctorat, université Stendhal de Grenoble, soutenue le 18 décembre 2009.
- MARS M., ANTONIADIS G., ZRIGUI M.** (2007). A morphological analyser for Arabic language. *The first international conference of information and communication technology and accessibility*, April 12-14 2007, Hammamet, Tunisia
- PRESSEY S. L.** (1927). A machine for automatic teaching of drill material. *School and Society*, 25, pp. 549-552.
- SELVA T.** (2002). Génération Automatique d'Exercices Contextuels de Vocabulaire. *TALN 2002* pp 185-194.
- SKINNER F.B.** (1968). *La révolution scientifique de l'enseignement*. Bruxelles : Charles Dessart.
- SWARTZ M., YAZDANI M.** (dir.) (1992). *The Bridge to International Communication: Intelligent Tutoring Systems for Foreign Language Learning*. Springer-Verlag.
- TRIBBLE C., BARLOW M.** (2001). Using Corpora in Language Teaching and Learning. *Special issue of Language Learning and Technology, Volume 5, Number 3*.

¹ Néanmoins, plusieurs travaux ont pour objet la structuration informatique des ressources pédagogiques, souvent en dehors d'une problématique d'apprentissage.

² Par rapport à l'enseignement/apprentissage d'autres disciplines.