

Quelles machines pour enseigner la langue ?

Georges Antoniadis¹, Cédric Fairon², Sylviane Granger³,
Julia Medori²⁻³, Virginie Zampa¹

¹ Université Stendhal de Grenoble, Laboratoire LIDILEM
{Georges.Antoniadis ; Virginie.Zampa}@u-grenoble3.fr

² Université catholique de Louvain, Centre de Traitement Automatique du Langage
{fairon,medori}@tedm.ucl.ac.be

³ Université catholique de Louvain, Centre for English Corpus Linguistics
granger@lige.ucl.ac.be

Résumé

Cet article présente dans un premier temps l'histoire de l'enseignement assisté par ordinateur (EAO) en situant ses origines aux années 1920 avec les premières machines à enseigner mécaniques. L'arrivée de l'ordinateur a par la suite permis de proposer à l'apprenant de langues différents types d'activités : tâches de compréhension, simulations, etc. Cependant, celles-ci ont des limites qui ne peuvent être surmontées sans l'apport du traitement automatique des langues (TAL). Nous présentons ici la problématique de l'intégration du TAL aux systèmes d'ALAO en dressant un bilan des défis que cette intégration doit aujourd'hui relever et nous faisons une synthèse des présentations de l'atelier. Celles-ci proposent des problématiques diverses allant de la détection et la correction d'erreurs à l'enrichissement de dictionnaires électroniques en passant par la mise en œuvre d'outils complets d'aide à l'apprentissage des langues. Nous verrons que la clé de l'intégration du TAL dans l'ALAO réside dans le travail pluridisciplinaire entre informaticiens, didacticiens des langues et spécialistes du TAL.

Mots-clés : TAL, ALAO, systèmes d'apprentissage des langues, intégration du TAL à l'ALAO, correction d'erreurs, dictionnaire électronique, phraséologie.

Abstract

This paper first presents a history of Computer-Assisted Learning (CAL), setting its origins in the 1920s with the invention of mechanical learning machines. The use of the computer then allowed the development of different types of language learning activities: comprehension tasks, simulations, etc. However, without the contribution of natural language processing (NLP), these activities are of limited use. We address the problem of the integration of NLP in CALL systems while summing up the challenges this integration has to overcome today and synthesize the workshop presentations. These presentations deal with a range of issues from error detection and correction to the extension of electronic dictionaries through the implementation of comprehensive language learning tools. We will see that the key to the integration of NLP in CALL is in the pluridisciplinary work between didacticians, IT and NLP specialists.

Keywords: NLP, CALL, language learning systems, integration of NLP in CALL, error correction, electronic dictionary, phraseology.

1. Introduction

L'atelier « Traitement automatique des langues et Apprentissage des langues assisté par ordinateur : quelles perspectives d'intégration ? » organisé conjointement avec TALN 2006 à Leuven (Belgique) est une tentative de réponse à la question du titre : quelles perspectives

peut-on espérer de l'intégration du TAL dans les systèmes d'ALAO ? Cette question sous-entend une autre : l'ALAO peut-il se passer des procédures du TAL ? Si la réponse à cette dernière question est négative, une troisième question est alors à poser : quelle est la plus-value didactique d'une telle intégration ?

La question de la nécessité d'intégration du TAL dans les systèmes d'ALAO est abordée depuis plusieurs années dans plusieurs colloques internationaux, notamment dans des ateliers organisés par le « Special Interest Group in Language Processing » d'EUROCALL et par l'Integrating Speech Technology in (Language) Learning SIG. Un atelier organisé à Grenoble en octobre 2004 dans le cadre d'une journée ATALA¹ a donné lieu à un numéro thématique de la revue ALSIC (Antoniadis et Chanier, 2005). Même si cet atelier ne concernait, en grande partie, que des travaux francophones, la conclusion fut unanime, seul le TAL permet de considérer la langue, objet de l'apprentissage, en tant que système de formes et d'associations forme-sens.

Une telle affirmation n'entraîne pas nécessairement l'utilisation *a priori* du TAL pour les systèmes d'ALAO ; à notre avis, une telle intégration ne peut se justifier que par la plus-value didactique qu'elle induit. En ce sens, l'objet de cet atelier est d'essayer de définir cette plus-value, de la matérialiser, de la quantifier, de la formaliser, tant que faire se peut. Vaste programme pluridisciplinaire !

2. Des machines à enseigner à l'ALAO

Dès 1912, Thorndike rêvait d'un livre manuel mécanisé : « si, par le miracle et l'ingéniosité mécanique, un livre pouvait être agencé de telle façon que seulement pour celui qui aurait fait ce qui est demandé à la première page, la page deux devienne visible, et ainsi de suite, beaucoup de ce qui requiert actuellement de l'instruction personnelle pourrait être assuré par le livre » (Thorndike, 1912, cité par Bruillard, 1997, p. 33-34). L'idée de l'enseignement programmé (EP) mécanisé est ainsi née dès le début du XX^e siècle en réponse à des critiques émises à l'encontre de l'enseignement « classique » et afin essentiellement de permettre un rythme d'apprentissage adapté à l'élève et une activité continue pour ce dernier.

2.1. Les premières machines à enseigner et l'enseignement programmé

2.1.1. *La machine de Pressey*

La première machine à enseigner (ME) est celle élaborée par Sidney Pressey (1927), dans les années 1920. Il s'agit d'une machine automatisée pour corriger les Questions à Choix Multiples (QCM) avec quatre boutons correspondant aux réponses possibles à la question présentée. L'apprenant ne passe à la question suivante que lorsque sa réponse est juste et la machine garde une trace des actions de l'apprenant. Certains, dont Skinner (1968), lui reprochent d'avoir fondé sa machine sur des connaissances insuffisantes du phénomène d'apprentissage. Ils vont ainsi se focaliser sur les phénomènes d'apprentissage et créer l'enseignement programmé.

2.1.2. *L'enseignement programmé*

De Montmollin (1971) définit l'enseignement programmé comme « une méthode pédagogique qui permet de transmettre des connaissances sans l'intermédiaire direct d'un professeur ou d'un moniteur, ceci tout en respectant les caractéristiques de chaque élève pris

¹ Journée ATALA : « TAL et apprentissage des langues », 22 octobre 2004, Grenoble (France), <http://w3.u-grenoble3.fr/lidilem/talal/>.

individuellement ». Cet enseignement repose sur quatre principes : le principe de structuration de la matière à enseigner (il s'agit de découper et de présenter la matière de manière à faciliter la compréhension et la mémorisation), le principe d'adaptation (l'enseignement doit être adapté à l'élève), le principe de stimulation et le principe de contrôle. Le cheminement est soit linéaire (exemple : la machine de Skinner) soit ramifié (exemple : la machine de Crowder (1963)).

Skinner s'appuie sur les résultats de ses travaux en psychologie du comportement et en partant des résultats du conditionnement opérant en tant que théorie du contrôle des mécanismes d'apprentissage, il envisage la création d'une technologie scientifique de l'enseignement qui utilise l'enseignement programmé susceptible d'être dispensé par une machine à enseigner. Pour Skinner, l'efficacité de l'apprentissage repose sur cinq principes :

- le principe de la participation active : le sujet doit construire sa propre réponse et non la choisir (le QCM entraîne des erreurs que l'élève n'aurait jamais commises sans cette suggestion) ;
- le principe des petites étapes : il faut fragmenter la difficulté pour que même les plus faibles puissent répondre ;
- le principe de progression graduée ;
- le principe de l'allure personnelle : chacun doit pouvoir avancer à son rythme ;
- le principe des réponses correctes : trop d'échecs découragent les élèves, il faut les guider.

Ainsi, dans sa machine, les exercices se trouvent sur un rouleau que l'apprenant fait défiler grâce à une molette. Les questions apparaissent dans une fenêtre, l'élève inscrit sa réponse sur un espace blanc réservé à cet effet, puis il compare sa réponse à la correction et actionne le levier pour passer à la question suivante.

Mais l'EP de Skinner est rapidement critiqué et, dès 1959 Crowder propose un système alternatif. Sa machine diffère de celle de Skinner par différents points, notamment le cheminement et le type d'exercices. En effet, la machine présente des informations qui sont suivies par un QCM. Contrairement à Skinner qui cherche à limiter l'erreur, Crowder lui attribue une fonction importante. De plus, il considère qu'apprendre, c'est souvent apprendre à distinguer, à discriminer. Une fois la réponse corrigée, si elle est bonne, l'apprenant passe à l'information suivante, si elle est mauvaise, l'apprenant est dirigé vers des exercices de rattrapage pour ensuite revenir à l'exercice auquel il a échoué. Lorsqu'il répond correctement à plusieurs questions, il passe par des raccourcis. Ce genre de ME permet ainsi une meilleure adaptation à l'apprenant, et prend ainsi une place prépondérante dans l'EP.

L'EP 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.

2.2. L'utilisation de l'ordinateur pour l'enseignement

Au départ, l'utilisation de l'ordinateur se limitait à automatiser ce qui était fait mécaniquement par les ME. L'enseignement assisté par ordinateur (EAO) n'est réellement né qu'au début des années 60 et ce n'est que dans les années 70, avec les travaux sur les systèmes experts qu'apparaissent les premières tentatives de rendre « intelligent » l'EAO. Ces recherches avaient pour finalités de combler les limites existantes, c'est-à-dire :

- de dialoguer avec l'apprenant en langage naturel ;
- de sélectionner la suite de ce qui doit être enseigné ;

- d'anticiper, de diagnostiquer et de comprendre les erreurs de l'apprenant ;
- d'améliorer les stratégies d'enseignement et de le modifier en fonction de l'apprenant.

Puis les années 70 sont marquées par les premiers micro-mondes, les années 80 par les tuteurs intelligents et les années 90 par les systèmes coopératifs et les environnements interactifs d'apprentissage avec ordinateur.

Les exercices proposés en ALAO se divisent, pour Mangenot (1997), en six catégories (cf. ci-dessous). En fonction du support (DVD / Web), du mode d'utilisation (formation avec un enseignant, autoformation, etc.), de l'âge des personnes à qui il est adressé, etc., ces différentes catégories occupent une place plus ou moins importante.

- **Les tâches de compréhension.** L'évolution des ordinateurs a permis rapidement d'introduire des séquences audio, puis des séquences vidéo dans les tâches de compréhension. Mais le plus souvent, les exercices qui s'y rapportent, sont sous la forme de QCM ou d'objets (séquences audio/vidéo, mots, etc.) à cliquer ou glisser et la correction est en termes de vrai/faux avec parfois quelques commentaires oraux ou écrits tels que « bravo », « c'est bien », etc.
- **Les exercices ayant pour but l'acquisition de connaissances discursives.** Il s'agit le plus souvent de puzzle, de repérage de séquence, d'appariement. La correction est en termes de vrai/faux.
- **Les enregistrements d'énoncés et les exercices oraux de transformation d'énoncés.** Ce type d'exercices correspond à une tâche de répétition ou de transformation d'énoncés. L'apprenant a la possibilité de s'enregistrer, de s'écouter. Dans certains logiciels sa voix est utilisée pour doubler un personnage, dans d'autres l'apprenant peut regarder son sonagramme et le comparer à celui de la personne qu'il répète. Dans tous les cas, il est en situation d'autocorrection.
- **Les simulations.** Mangenot distingue trois formes de simulations : celles qui permettent de laisser le choix à l'utilisateur (ce qui correspond au fonctionnement des « livres dont vous êtes le héros ») par exemple, entre aller au restaurant et aller visiter un monument ; celles qui consistent à prendre connaissance de documents des personnages et à faire la même chose qu'eux, par exemple, un CV ; le troisième type de simulation consiste à associer des éléments graphiques pour obtenir une réaction du système. Ce type d'exercice n'entraîne pas de correction mais plutôt des commentaires.
- **Les productions écrites.** La production écrite n'est pas très présente dans les logiciels. Ceci est dû au fait que sa correction est trop difficile. Il existe déjà des problèmes pour les questions à réponses obligatoirement courtes (exemple : les textes à trous), mais le problème est bien plus grand avec les véritables productions pour lesquelles il faut prendre en compte la syntaxe et la sémantique. De ce fait, en ALAO, les corrections apportées lors d'exercices de production écrites sont soit formulées sous forme de correction type que l'apprenant doit comparer à sa production, soit d'indications

Que ce soit sur CD-Rom ou sur Internet, les logiciels d'ALAO ne comportent que très peu d'exercices de lecture et d'écriture. Les exercices les plus présents sont ceux nécessitant le moins de corrections, c'est-à-dire ceux pour lesquels la réponse est donnée grâce à un clic de souris.

3. Le TAL au secours de l'ALAO

Les premières tentatives d'utilisation du TAL pour l'apprentissage des langues datent des années 80. Maladroites au départ (voire scientifiquement contestables lorsque, par exemple, elles essaient d'assimiler le TAL à l'utilisation des programmes tels que ELIZA ou SHRDLU), elles se développent au début des années 90. Le nombre de symposiums et de travaux européens (Swartz et Yazdani, 1992), nord américains (Holland *et al.*, 1995) ou français (Chanier *et al.*, 1993) pendant cette période atteste de cette activité. Ce constat est partagé par Jung (2005), qui situe le pic des travaux dans la période 85-95. Cela correspond au plein développement des ordinateurs personnels et à la première tentative d'insertion des TIC (Technologies de l'Information et de la Communication) dans les dispositifs pédagogiques. Une trace de ces travaux est présente dans toutes les revues traitant de l'apprentissage des langues : *CALICO*, *Language Learning and Technology*, *ReCALL*, *CALL*, *ALSIC*, *System*.

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, de solutions et de 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 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 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 une « conséquence » majeure, à notre avis, de l'utilisation du TAL et, bien sûr, de l'informatique. Le TAL a permis aux enseignants des 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 ; et d'automatiser l'analyse des productions des apprenants eux-mêmes. Cette approche d'apprentissage a donné lieu à un grand nombre de travaux, des ressources et des systèmes (Tribble et Barlow, 2001 ; Granger *et al.*, 2001 ; Granger 2002). 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.

Si, dans les années 80, la problématique de l'utilisation du TAL pour l'ALAO essaie de se construire en tâtonnant, elle se stabilise dans les années 90, et elle a peu varié depuis. Mis à part, la question « peut-on se passer du TAL pour les systèmes d'ALAO ? » qui a préoccupé

et préoccupe toujours des travaux du domaine (Chanier 1998 ; Brun *et al.*, 2002, Antoniadis 2004), quatre problèmes constituent la majeure partie de la problématique du domaine :

- **Il faut définir et évaluer l'apport du TAL pour l'ALAO.** L'apport potentiel du TAL découle de sa problématique et du but qu'il s'est fixé. Lui seul permet de considérer la forme langagière, non pas comme une suite de signes dépourvus d'interprétation, mais comme des éléments d'un système à deux niveaux (forme et sens). Dans le cadre de l'apprentissage des langues, le fonctionnement de chaque niveau de ce système doit être considéré, détaillé, manipulé, mis en pratique d'une manière ciblée ; les liens entre les deux niveaux doivent être mis en évidence concrètement et la polysémie, source de difficultés en apprentissage des langues, doit pouvoir trouver la place et le traitement appropriés. Seule l'utilisation du TAL permet actuellement d'espérer atteindre ce but, de créer des systèmes et des outils permettant aux enseignants des langues (et par extension aux apprenants) de manipuler la langue telle qu'elle est définie dans leur discipline et non telle que l'informatique est capable de la considérer.

Comme les enseignants des langues, un système d'apprentissage des langues ne peut être valide, voire acceptable, que s'il est capable, d'une part d'engendrer uniquement des connaissances langagières correctes (non fausses) et d'autre part d'expertiser correctement les productions langagières des apprenants. L'utilisation des procédures du TAL ne peut être envisageable que si elle permet de satisfaire à cette double condition. Concrètement, l'utilisation du TAL n'est possible que lorsque les résultats de ses traitements sont sûrs ou quasi-sûrs ; l'ambiguïté, problème essentiel en TAL, n'est concevable en ALAO que si elle est maîtrisée par le système. Toute maîtrise partielle peut avoir comme conséquence l'apprentissage de concepts langagiers partiellement ou complètement erronés. C'est en ce sens que l'apport du TAL pour l'ALAO doit être défini et évalué. Les travaux actuels tentent, par petits pas, de cerner les résultats du TAL utilisables pour la construction des systèmes d'ALAO.

- **Il faut connaître l'apprenant pour mieux définir son apprentissage.** L'approche individualisée d'apprentissage n'est une idée ni nouvelle, ni propre à l'ALAO. S'il est généralement admis qu'individualiser l'apprentissage permet de l'optimiser, force est de constater que les systèmes maîtrisant une telle démarche sont à créer. Ce n'est pas tant l'évaluation des connaissances de l'apprenant qui constitue la difficulté majeure, mais l'exploitation de cette évaluation et, surtout, la création automatique d'activités en fonction du niveau mesuré de l'apprenant. Le TAL peut apporter des outils aussi bien pour l'évaluation que la création automatique d'activités. Certes ces outils ne sont pas la panacée ; ils constituent néanmoins actuellement la seule piste pour apporter des réponses, partielles le plus souvent, à ces questions.
- **Il faut être capable de détecter, d'expliquer et de corriger automatiquement les fautes de l'apprenant si l'on souhaite qu'il puisse travailler en autonomie.** La question de la correction automatique des productions d'apprenants comme celle de la génération automatique d'explications pour l'apprenant "hantent" les travaux en ALAO depuis ses origines. Néanmoins, aller au delà d'une correction du type "vrai/faux" à partir d'une liste de "vérités" préétablie, ou de l'affichage d'explications préenregistrées, demande la prise en compte des aspects langagiers des réponses fournies, l'évaluation dans le contexte de l'activité pédagogique et en fonction du profil de l'apprenant, la génération, éventuellement, d'explications en langue, en tenant compte des paramètres précédents. En ce sens, et par définition du domaine, le TAL peut apporter toute une panoplie de procédures et d'outils grâce auxquels les questions d'évaluation et d'explications peuvent se poser sur d'autres bases, dans une démarche

constructive et cumulative. Néanmoins, cet apport du TAL est à considérer en tenant compte de ses possibilités actuelles : modéliser et « mesurer » l'attendu, en rejetant les erreurs (l'inattendu). En ce sens l'analyse d'énoncés erronés² et la génération d'explications associées ne peut pas être une application directe des procédures TAL; des méthodes et des heuristiques spécifiques restent à inventer.

- **Il faut que les outils et systèmes proposés ne demandent pas de compétences techniques spécifiques.** En tant que domaines de connaissances, l'informatique comme le TAL, la linguistique comme la didactique, manipulent des concepts qui leur sont propres et demandent l'usage d'outils spécifiques, fondés, le plus souvent, sur ces concepts. Pour le profane, l'utilisation de tels outils peut entraîner de longs apprentissages et suppose l'acquisition d'un minimum de concepts du domaine à l'origine de chaque outil. Dans le cadre de l'apprentissage des langues, si les enseignants sont *a priori* des spécialistes de didactique, leurs compétences en informatique ou en TAL, voire en linguistique, sont le plus souvent bien plus limitées, voire minimales ou inexistantes. Cet état peut rendre problématique l'appropriation didactique ou pédagogique de tout outil en dehors du domaine des préoccupations scientifiques des enseignants (et *a fortiori* des apprenants). En ce sens, tout produit de l'ALAO, destiné *a priori* à des didacticiens des langues, doit satisfaire à deux impératifs : son utilisation ne doit demander qu'un minimum de compétences non didactiques et doit permettre le maniement de concepts didactiques en vue de la mise en œuvre de solutions didactiques ou pédagogiques.

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 (Chanier *et al.*, 1995 ; Brun *et al.*, 2002 ; Antoniadis *et al.*, 2005), les systèmes commercialisés sont extrêmement rares. 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 langues³, 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.

Ne soyons pas pessimistes ! Un véritable effort de standardisation est en train de s'accomplir dans la communauté TAL, 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 œuvre des outils les plus simples est susceptible d'apporter une plus-value qui dépasse de loin ce modeste investissement. Quant au travail en collaboration entre didacticiens des langues et spécialistes du TAL, l'existence de cet atelier et les travaux qui y sont présentés tendent à prouver qu'un langage commun de travail est en train de se forger.

4. Synthèse de l'atelier

Dans la première présentation de l'atelier, Tschichold aborde le problème majeur qui se présente à quiconque aborde la place du TAL dans l'ALAO, à savoir le manque d'interaction et de collaboration active entre les partenaires principaux : les entreprises commerciales qui

² Il ne s'agit pas ici de détecter seulement l'existence d'une faute, mais de calculer en plus le différentiel entre la forme erronée et celle qui aurait dû être à sa place.

³ L'inverse étant aussi vrai, le travail en commun n'est souvent que le résultat du hasard.

produisent des didacticiels, les enseignants de langues et les spécialistes du TAL. Elle met en évidence les qualités ainsi que les imperfections des différents acteurs de l'apprentissage des langues assisté par l'ordinateur. D'une part, les systèmes commerciaux savent rendre leurs produits attractifs par des interfaces travaillées et des outils multimédia de pointe mais souvent ne prévoient pas de scénarisation pédagogique. D'autre part, les professeurs de langue ont des connaissances didactiques basées sur leur expérience de l'enseignement et les contenus qu'ils développent sont dès lors ciblés et bien adaptés à leur public. Cependant, les systèmes ALAO qu'ils produisent sont souvent peu sophistiqués du point de vue technologique et linguistique. Finalement, les linguistes disposent des outils de TAL qui leur permettent d'ajouter des fonctionnalités aux logiciels d'ALAO mais leurs recherches aboutissent souvent à des prototypes plutôt qu'à des outils réellement utilisables par les enseignants. Ces différents acteurs bien que complémentaires ne communiquent que très rarement entre eux. Cette constatation a été abordée entre autres par Salaberry (1996) et Borin (2002). Borin évoque ce problème comme étant dû à des différences de culture et une mauvaise compréhension entre les différentes disciplines. Salaberry réfute l'idée que l'ordinateur pourrait se substituer à l'enseignant mais défend plutôt l'idée qu'il servirait comme outil de support à l'utilisateur. Nous verrons que la plupart des articles présentés lors de l'atelier évoquent effectivement des outils d'aide à l'apprenant.

L'environnement ALAO pour le basque développé par Aldabe *et al.* illustre bien le type d'approche interdisciplinaire prônée par Tschichold. Leur environnement, qui intègre une **large gamme d'outils de TAL**, est destiné à répondre aux besoins de trois types d'utilisateurs : les apprenants, les enseignants et les linguistes computationnels. Il comprend de nombreuses ressources (conjugueur, concordancier, etc.) destinées à aider les apprenants à résoudre des exercices ou produire des textes et permet, en outre, de collecter des corpus d'apprenants et de traiter leurs erreurs. L'interface a été évaluée par une population d'étudiants, qui l'a jugée très utile pour l'aide à la résolution d'exercices. L'environnement proposé comporte de nombreuses similitudes avec l'environnement d'aide à l'écrit développé par Wible *et al.* (2001) pour l'anglais. Parmi d'autres plateformes intégrant des outils de TAL, citons GLOSSER (Nerbonne *et al.*, 1998), un outil d'aide à la lecture en français par des étudiants de langue maternelle néerlandaise, et MIRTO (Antoniadis *et al.*, 2005), qui intègre des fonctions de TAL basiques que l'enseignant peut combiner de diverses manières pour créer des scénarios pédagogiques pour ses élèves.

L'environnement décrit par Aldabe *et al.* attache une importance particulière à la **détection et la correction des erreurs** des apprenants, un domaine qui est l'objet de nombreux travaux actuellement, comme l'atteste notamment un numéro récent de la revue *CALICO* qui est entièrement consacré à ce sujet (Heift et Schulze, 2003). Deux autres présentations à l'atelier sont centrées sur cette problématique. Audras et Ganascia présentent un outil de diagnostic de l'apprenant basé sur l'utilisation d'un analyseur morphosyntaxique et d'un analyseur stylistique, le *Littératron*. Ce logiciel extrait des motifs syntaxiques récurrents dans un corpus afin de repérer les tournures caractéristiques des étudiants de français de langues maternelles et de niveaux d'apprentissage différents. Hermet *et al.* quant à eux ont développé un environnement virtuel dont le but est d'améliorer la compréhension à la lecture d'étudiants de français langue seconde. Le système, qui comporte un parseur, un dictionnaire de synonymes et un dictionnaire de dérivations, permet de reconnaître et d'évaluer les réponses à des questions ouvertes. En matière de correction automatique, le TAL est plus généralement utilisé pour la résolution automatique d'exercices de type QCM, exercices à trous, questions fermées où les possibilités d'erreur de correction sont limitées. Beaucoup s'accordent pour dire qu'il n'est pas acceptable de donner des informations erronées à l'apprenant au risque de perdre sa confiance en l'outil qu'il utilise. Des systèmes de correction automatique de

questions ouvertes existent basés sur des similarités avec des réponses types ou un vocabulaire attendu dans la réponse pré-entrée dans la machine (cf. par exemple, Perez 2004).

Un autre domaine clé est celui des **dictionnaires électroniques** dont l'importance en ALAO est appelée à croître en raison notamment du rôle grandissant joué par le lexique dans l'apprentissage des langues (cf. la « lexical approach » de Lewis 1993 et 1997). Deux articles abordent cette question : Zock ainsi que Buvet et Issac suggèrent une extension des informations stockées dans un dictionnaire classique, à savoir les définitions et les traductions. La démarche de Buvet et Issac est clairement onomasiologique, celle de Zock est à la fois onomasiologique et sémasiologique. Zock propose un outil d'aide à la compréhension et à la production. Il part du principe qu'un apprenant connaît souvent le mot qu'il recherche mais n'en a qu'une idée approximative. La forme entrée par l'utilisateur ressemble soit par la forme soit par le sens au mot recherché. Pour les approximations de forme, le système utilise des algorithmes orthographiques pour retrouver la forme exacte. Pour ce qui est des approximations de sens, le dictionnaire peut être enrichi d'un réseau lexical et présenter à l'utilisateur la liste des mots proches par le sens. Ainsi, un apprenant qui cherche le mot 'infirmière' passera par le mot 'hôpital' pour le retrouver. À partir de ce dictionnaire, Zock envisage la création automatique d'exercices d'aide à la compréhension avec la présentation d'une liste de mots à mémoriser et dès que l'un d'eux est retenu, il est enlevé de la liste; mais aussi des exercices d'aide à la production avec un exercice de construction de phrases à partir d'un modèle et de mots à utiliser. Buvet et Issac eux, proposent un outil d'aide à la rédaction qui génère des phrases simples du domaine de l'affect dans le but de familiariser l'apprenant avec le lexique concerné et son emploi. Pour cela, chaque entrée du dictionnaire est formalisée suivant le modèle des classes d'objet contenant des informations syntaxiques, sémantiques (hyperonymes, synonymes, etc.) et les contraintes sur les arguments (*sujet : humain*, par exemple). La génération de phrases se produit par l'intermédiaire d'une interface qui, à partir d'un concept et des arguments choisis par l'apprenant, génère toutes les phrases possibles exprimant cet affect. Ces deux projets sont dans la même lignée que le projet ALEXIA (Chanier, 1995), un environnement proposant une aide à la compréhension avec la construction d'une base de données lexicale personnalisée, mais aussi un outil d'aide à la rédaction qui analyse le texte produit par l'apprenant sur un thème précis. Parmi d'autres dictionnaires d'aide à la compréhension ou à la production, citons le DAFLES (Selva, 2002), un dictionnaire utilisé pour la génération automatique d'exercices lexicaux, et le lexique bilingue de Pecman (2005), qui a pour but de fournir une aide à la rédaction scientifique en anglais et en français.

L'exploitation par le TAL des corpus de textes électroniques a revitalisé et profondément modifié le domaine de la **phraséologie**, qui touche à l'extraction, la catégorisation et l'intégration des unités polylexicales (cf. Granger et Meunier, à paraître). L'article de Wible *et al.* décrit un outil d'extraction et de représentation de séquences préfabriquées (« lexical chunks ») ainsi qu'un outil, le *Collocator*, servant à mettre en évidence les collocations dans un texte afin d'aider l'apprenant à reconnaître les expressions figées. Cet outil permet à l'apprenant de découvrir les collocations en temps réel à l'intérieur des pages Web qu'il parcourt. Il est basé sur un système de statistiques qui permet de trouver des cooccurrences de suites d'un ou de plusieurs mots même si celles-ci sont entrecoupées d'autres mots.

5. Conclusion

Le champ couvert par l'application du TAL à l'ALAO est vaste. Il recouvre de nombreux sujets dont certains sont totalement absents de cet atelier, en particulier ce qui touche au traitement de l'oral. En ce qui concerne le traitement de l'écrit, cependant, les thématiques

abordées donnent une bonne idée des nombreux 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 et il faut bien avouer que cet aspect est souvent marginal, voire inexistant. La clé réside dans le travail pluridisciplinaire et la mise en commun de pratiques et de techniques. Nous espérons que cet atelier contribuera à fédérer les travaux menés par des chercheurs d'horizons divers et à faciliter l'intégration des résultats de leurs recherches dans des outils d'ALAO conviviaux et performants.

Références

- ANTONIADIS G. (2004). « Les logiciels d'apprentissage des langues peuvent-ils ignorer le TAL ? ». In *Les cahiers de l'APLIUT XXIII* (2), juin 2004 : 81-97.
- ANTONIADIS G., CHANIER, T. (Éds.) (2005). *ALSIC* 8 (2), numéro thématique « TAL et apprentissage des langues ».
- ANTONIADIS G., ECHINARD S., KRAIF O., LEBARBÉ Th., PONTON C. (2005). « Modélisation de l'intégration de ressources TAL pour l'Apprentissage des Langues : la plateforme MIRTO ». In *ALSIC* 8 : 65-79
- BORIN L. (2002). « What have you done for me lately? The fickle alignment of NLP and CALL ». Reports from Uppsala Learning Lab.
- BRUILLARD E. (1997). *Les machines à enseigner*. Hermès, Paris.
- BRUN C., PARMENTIER T., SANDOR A., SEGOND Fr. (2002). « Les outils de TAL au service de la e-formation en langues ». In Fr. Segond (dir.), *Multilinguisme et traitement de l'information*. Hermès, Paris : 223-250.
- CHANIER T. (1998). « Relations entre le TAL et l'ALAO ou l'ALAO un simple domaine d'application du TAL ? » In *International conference on natural language processing and industrial application (NLP+IA'98)*. Moncton.<http://lifc.univ-fcomte.fr/RECHERCHE/P7/pub/Moncton/index.htm>
- CHANIER T., RENIÉ D., FOUQUERÉ C. (dir.) (1993). *Actes du colloque SCIAL'93 (sciences cognitives, informatique et apprentissage des langues)*. Université Blaise Pascal, Clermont-Ferrand. Préface et sommaire en ligne à <http://archive-edutice.ccsd.cnrs.fr/edutice-00001148>
- CHANIER T., FOUQUERÉ C., ISSAC F. (1995). « AlexiA : Un environnement d'aide à l'apprentissage lexical du français langue seconde ». In *Conférence Environnements Interactifs d'Apprentissage avec Ordinateur (EIAO'95)*. Eyrolles, Paris : 79-90.
- CROWDER N. (1963). « On the difference between linear and intrinsic programming ». In *Phi Delta Kappan* 44 : 250-254.
- DE MONTMOLLIN M. (1971). *L'enseignement programmé*. PUF, Paris.
- GRANGER S. (2002). « A bird's eye view of learner corpus research ». In S. Granger, J. Hung et S. Petch-Tyson (eds), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching, Language Learning and Language Teaching* 6. Benjamins, Amsterdam/Philadelphia : 3-33.
- GRANGER S., MEUNIER F. (eds) (à paraître). *Phraseology : An Interdisciplinary Perspective*. Benjamins, Amsterdam/Philadelphia.
- GRANGER S., VANDEVENTER A., HAMEL M.J. (2001). « Analyse des corpus d'apprenants pour l'ELAO basé sur le TAL ». In *TAL* 42 (2) : 609-621.
- HEIFT T., SCHULZE M. (eds) (2003). *Special issue of CALICO on Error Analysis and Error Correction in Computer-Assisted Language Learning* 20 (3).
- HOLLAND V.M., KAPLAN J.D., SAMS M.R. (dir) (1995). *Intelligent Language Tutors*. Mahwah. Lawrence Erlbaum Associates, NJ.

- LEWIS M. (1993). *The Lexical Approach : The State of ELT and a Way Forward*. Language Teaching Publications, Hove.
- LEWIS M. (1997). *Implementing the Lexical Approach : Putting Theory into Practice*. Language Teaching Publications, Hove.
- MANGENOT F. (1997). « Synthèse de trois cours de FLE sur CD-ROM ». In *Les Cahiers de l'asdifle, multimédia et langue étrangère, Actes des 19^e et 20^e rencontres* : 79-88.
- NERBONNE J., DOKTER D., SMIT P. (1998). « Morphological Processing and Computer-Assisted Language Learning ». In *Computer-Assisted Language Learning* 11 (5) : 543-559.
- PECMAN M. (2005). « Compilation, formalisation and presentation of bilingual phraseology ». In C. Cosme, G. Gouverneur, F. Meunier, M. Paquot (éds), *Phraseology 2005. The Many Faces of Phraseology. An Interdisciplinary Conference*. Université catholique de Louvain, Louvain-la-Neuve : 335-338.
- PEREZ D. (2004). *Automatic Evaluation of User's Short Essays by Using Statistical and Shallow Natural Language Processing Techniques*. Advanced Studies Diploma Work. Universidad Autonoma de Madrid.
- PRESSEY S.L. (1927). « A machine for automatic teaching of drill material ». In *School and Society* 25 : 549-552.
- SALABERRY R. (1996). « A Theoretical Foundation for the Development of Pedagogical Tasks in Computer Mediated Communication ». In *CALICO* 14 (1) : 5-34.
- SELVA T. (2002). « Génération Automatique d'Exercices Contextuels de Vocabulaire ». In *Actes de TALN 2002* : 185-194.
- SKINNER F.B. (1968). *La révolution scientifique de l'enseignement*. Charles Dessart, Bruxelles.
- SWARTZ M., YAZDANI M. (dir.) (1992). *The Bridge to International Communication : Intelligent Tutoring Systems for Foreign Language Learning*. Springer-Verlag.
- THORNDIKE E.L. (1913). *Educational Psychology 2. The Psychology of Learning*. Teachers College, New-York.
- TRIBBLE C., BARLOW M. (éds) (2001). *Using Corpora in Language Teaching and Learning*. Special issue of *Language Learning and Technology* 5 (3).
- WIBLE D., KUO C-H., CHIEN F-Y., LIU A., TSAO N-L. (2001). « A Web-based EFL writing environment : integrating information for learners, teachers, and researchers ». In *Computers and Education* 37 : 297-315.