



HAL
open science

De la nécessité du TAL dans les EIAH en langues Les cas EXXELANT et MIRTO

Georges Antoniadis, Claude Ponton, Virginie Zampa

► To cite this version:

Georges Antoniadis, Claude Ponton, Virginie Zampa. De la nécessité du TAL dans les EIAH en langues Les cas EXXELANT et MIRTO. Jun 2007. <hal-00161556>

HAL Id: hal-00161556

<https://hal.science/hal-00161556v1>

Submitted on 11 Jul 2007

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

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



HAL Authorization

De la nécessité du TAL dans les EIAH en langues

Les cas EXXELANT et MIRTO

Georges Antoniadis, Claude Ponton, Virginie Zampa

LIDILEM

BP 25 - 38040 Grenoble cedex 9

{georges.antoniadis ; claude.ponton ; virginie.zampa}@u-grenoble3.fr

RÉSUMÉ. L'ALAO s'intéresse, entre autres, à l'élaboration d'environnements informatiques pour l'apprentissage des langues. La prise en compte par ces derniers des propriétés de la langue, aussi bien aux niveaux lexical, syntaxique que sémantique, objets de l'apprentissage, a conduit à partir des années 80 à l'utilisation de procédures et d'outils du traitement automatique de la langue (TAL). L'euphorie du départ a laissé progressivement la place à une intégration mesurée du TAL dans les systèmes d'ALAO, en tenant compte de ses possibilités, de ses limites mais également de la plus-value pédagogique possible.

Nous présentons dans cet article cette problématique de l'intégration du TAL dans les systèmes d'ALAO ainsi que l'apport du TAL pour ces systèmes. Les différents niveaux de cet apport sont illustrés à travers deux produits, le logiciel (exemplier) EXXELANT et la plateforme MIRTO. Nous montrons que le TAL peut et doit être au centre de tout système d'ALAO, à condition que son utilisation soit en adéquation avec la pertinence actuelle de ses résultats.

MOTS-CLÉS : EIAH, ALAO, TAL, génération d'activités, concordancier

1. TAL et EIAH en langues

Les premières tentatives d'utilisation du traitement automatique des langues (TAL) pour l'apprentissage des langues datent des années 80, avec un pic d'activités durant la période 1985-1995 (Jung, 2005 ; Chanier et al., 1993). Depuis la fin des années 90, l'ALAO (Apprentissage des Langues Assisté par Ordinateur) est devenu un domaine de recherche pluridisciplinaire. La tâche essentielle pour chaque discipline est la définition de la partie du terme « apprentissage des langues assisté par ordinateur » qui la 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. Mise à part l'interrogation « 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), trois aspects constituent la majeure partie de la problématique du domaine :

Définir et évaluer l'apport du TAL pour l'ALAO. Actuellement, seul le TAL permet d'espérer la création de systèmes et d'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 seule est capable de la considérer. Mais, l'utilisation des procédures TAL ne peut être envisagée que si elle permet, d'une part, d'engendrer des connaissances langagières correctes (non fausses), d'autre part, d'expertiser correctement les productions langagières des apprenants. Concrètement, l'utilisation du TAL n'est possible que lorsque les résultats de ses traitements sont sûrs ou quasi-sûrs ; l'ambiguïté, 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. Les travaux actuels tentent, à petits pas, de cerner les résultats du TAL utilisables pour la construction des systèmes d'ALAO.

Connaître l'apprenant pour mieux définir son apprentissage. En effet, afin de permettre à l'apprenant de travailler en autonomie, il faut être capable de détecter, d'expliquer et de corriger automatiquement ses fautes. Les questions de correction automatique des productions et de génération automatique d'explications "hantent" les travaux en ALAO depuis ses origines (Heift et Schulze, 2003). Certes, les outils apportés actuellement par le TAL, aussi bien pour l'évaluation que pour la création automatique d'activités, ne sont pas la panacée, mais ils constituent la seule piste pour apporter des réponses, partielles le plus souvent, à ces questions. 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). Ainsi, l'analyse d'énoncés erronés 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.

Proposer des outils et des systèmes ne demandant pas de compétences techniques spécifiques. Tout produit d'ALAO, destiné à des praticiens des langues ayant des connaissances limitées en informatique, en TAL voire même en linguistique, doit satisfaire à deux impératifs : son utilisation doit demander un minimum de compétences non didactiques et doit permettre le maniement de ces concepts pour la mise en oeuvre 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 (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, trois facteurs sont la cause principale de cet état : la méconnaissance du TAL de la part des didacticiens des langues, voire des informaticiens ; le coût des ressources et produits issus du traitement automatique de la langue ; le manque de fiabilité des outils TAL. Toutefois, les travaux actuels, en collaboration entre didacticiens des langues et spécialistes du TAL tendent à prouver qu'une langue de travail commune est en train de se forger. En guise d'illustration, nous présentons, deux systèmes d'ALAO utilisant le TAL : EXXELANT et MIRTO.

2. L'exemplier EXXELANT

EXXELANT (EXample eXtractor Engine LANguage Teaching) a été développé dans le cadre du projet IDILL (Integrated DIgital Language Learning ; <http://www.idill.org>) du réseau d'excellence européen Kaléidoscope (<http://www.no-kaleidoscope.org>). Il s'agit d'un « exemplier » (forme de concordancier) destiné à la fois aux enseignants de langue, aux chercheurs en didactique des langues et aux chercheurs en TAL. Il permet l'interrogation du corpus d'apprenants de FLE (Français Langue Etrangère) FRIDA-bis afin d'en extraire des exemples de phrases comportant certaines caractéristiques notamment au niveau des erreurs.

Le corpus FRIDA-bis est un corpus dérivé du corpus FRIDA (French Interlanguage Database) (Granger *et al.*, 2001), recueilli dans le cadre du projet FreeText. Il est composé de 764 textes soit 9466 phrases, 20 474 erreurs et 179 642 mots. Il est au format XML et possède un double système d'annotation : balisage manuel des erreurs d'apprenants et balisage automatique morphosyntaxique.

Ce corpus est indexé dans une base de données, ce qui permet, par le biais d'une interface web, de l'interroger facilement selon différents critères issus des balisages : nature du corpus (origine des sujets, densité d'erreur et nombre de mots), critères sur une expression (que ce soit pour le mot cible, sa correction ou les deux, la recherche peut être faite sur la forme, la forme lemmatisée, la catégorie, les traits morpho-syntaxiques), critères sur les contextes gauche et droit.

Les utilisations d'un tel système sont variées tant sur un plan de didactique appliquée que sur un plan de recherche en didactique ou en TAL. Pour les chercheurs en TAL, l'exploration de tels corpus devrait permettre une amélioration des modèles de détection et d'évaluation des erreurs pour la production de rétroactions adaptées. Les pédagogues et didacticiens peuvent, entre autres, cerner et quantifier les erreurs typiques des allophones, visualiser les fréquences d'emploi de mots et de structures, comparer des productions entre différents groupes d'apprenants (en fonction de leur langue maternelle, leur niveau), etc. Ceci avant de générer des exercices sur le type d'erreurs précises qu'ils ont choisis.

3. La plateforme MIRTO

Le projet MIRTO vise la réalisation d'une plateforme d'enseignement des langues. Il ambitionne de mettre au service des enseignants des ressources et des outils issus du TAL afin de leur permettre la conception d'activités et de scénarios pédagogiques (la notion de « scénario » renvoie ici à un enchaînement d'activités dont l'ordonnancement est fonction des résultats et du profil de l'apprenant) exploitant la diversité et la richesse des corpus textuels. Les activités et scénarios ainsi élaborés seront joués par les apprenants à distance ou en présentiel. Actuellement, seule la partie génération d'activités est développée. D'autres aspects, comme l'évaluation des réponses et la production de feed-back adaptés font l'objet de plusieurs études en cours, tout comme la conception d'une base de textes indexés pédagogiquement (Loiseau *et al.*, 2005).

L'approche MIRTO est 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 eux et seuls les aspects didactiques doivent être visibles et disponibles. Quatre niveaux hiérarchiques (fonction, script, activité et scénario), associés à une base de textes, structurent MIRTO. Les fonctions représentent le niveau le plus bas. Elles correspondent à des processus TAL basiques comme la tokenisation (découpage du texte en formes) ou l'étiquetage morpho-syntaxique. Étant donné la nature technique des fonctions et leur indépendance de toute application pédagogique, ce niveau n'est pas visible par les enseignants. Il permet la création de scripts qui correspondent à l'application de ces fonctions à la didactique des langues. Un script est un objet didactique qui constitue l'élément de base de création d'activités puis, éventuellement, de scénarios par les enseignants. 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 de langue à l'aide d'un outil spécifique de MIRTO : l'éditeur d'activités. Ce dernier 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.

4. Conclusion et perspectives

Les tâches telles que : l'élaboration de plans pédagogiques en adéquation avec les notions à enseigner et le niveau des apprenants ; la recherche de textes et/ou d'exemples appropriés ; l'élaboration et/ou la recherche d'activités relatives aux notions abordées ; l'évaluation des activités effectuées par les apprenants et la remédiation pédagogique en fonction des résultats ; font partie du « quotidien » pédagogique d'un enseignant de langue. De son côté, l'apprenant, en autonomie, semi-autonomie et/ou à distance, est amené à lire, à écouter des documents langagiers, à faire des exercices appropriés, à analyser ses erreurs grâce à des aides et des commentaires pédagogiques, à communiquer avec ses pairs ou avec ses tuteurs, à suivre une progression en fonction d'un parcours préétabli, mais aussi de sa progression réelle.

L'ALAO s'intéresse à l'ensemble des tâches ci-dessus ; plusieurs plateformes, logiciels, démarches, etc., tentent d'apporter une aide, voire de les automatiser complètement ou partiellement. Néanmoins, le plus souvent, ces solutions butent sur les limites intrinsèques de l'informatique. Science du traitement automatique de la forme d'informations, elle peine à considérer la sémantique associée aux séquences langagières. En général, ces séquences ne constituent pour elle qu'une concaténation de codes, avec comme seule sémantique, celle qui est associée aux propriétés mathématiques de la concaténation, bien loin de la sémantique que considèrent et manipulent les enseignants de langues.

Le TAL est actuellement le seul moyen pour tenter de trouver réponse à ces questions. Sans qu'il soit la panacée, il peut apporter une plus-value pédagogique aux systèmes et procédures d'ALAO. Tous les aspects, pratiquement, que l'ALAO traite sont concernés par le TAL. EXXELANT et MIRTO, comme EXILLS (Brun et al., 2002) et ALFALEX (www.kuleuven.be/alfalex/) sont des tentatives de solutions afin de tenir compte des propriétés de la langue pour des tâches telles que la création d'activités ou de scénarios pédagogiques, la détection (Heift & Schulze, 2003) ou l'évaluation des erreurs des apprenants pour l'élaboration d'un feedback à plus-value pédagogique (Antoniadis, Echinard & Ponton, 2005).

5. Bibliographie

- [ANTONIADIS 04] 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 et al. 05] Antoniadis G., Echinard S., Kraif O., Lebarbé T., Loiseau M., Ponton C. (2005) " MIRTO : a new approach to CALL systems " in Toulouse, Edité par J.P. Courtiat, C. Davarakis, T. Villemur Eds *IFIP TC3 Technology Enhanced Learning Workshop (TeL'04), World Computer Congress*.
- [ANTONIADIS, PONTON & ECHINARD 05] Antoniadis G., Echinard S., Ponton C (2005) " Le TAL au service de l'évaluation automatique des fautes des apprenants. Du vrai/faux à l'évaluation pédagogique ", *6ème colloque des Usages des Nouvelles Technologies pour l'Enseignement des Langues Etrangères, 23-26 mars 2005, Compiègne*.
- [BRUN et. al. 02] 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 98] Chanier T. (1998). "Relations entre le TAL et l'ALAO ou l'ALAO un simple domaine d'application du TAL ?". International conference on natural language processing and industrial application (NLP+IA'98). août 1998, Moncton, Canada. <http://lifc.univ-fcomte.fr/RECHERCHE/P7/pub/Moncton/index.htm>
- [CHANIER et al. 93] Chanier T., Renié 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>
- [GRANGER et al. 01] Granger S., Vandeventer A., Hamel M.J. (2001). Analyse des corpus d'apprenants pour l'ELAO basé sur le TAL. *Traitement Automatique des Langues*, Vol. 42, n°2, 609-621.
- [HEIFT & SCHULZE 03] Heift T., Schulze M. 2003. Special issue of CALICO on Error Analysis and Error Correction in Computer-Assisted Language Learning, 20(3).
- [JUNG 05] Jung, U.O.H. (2005). "CALL: past, present and future - a bibliometric approach". *ReCALL*, vol. 17, 1. pp 4-17.
- [LOISEAU et al. 05] Loiseau M., Antoniadis G., Ponton C. (2005) "Pedagogical text indexation and exploitation for language teaching" A. Méndez Vilas and B. Gonzalez Pereira and J. Mesa González and J.A. Mesa González *Recent research developments in learning technologies*, FORMATEX, Badajoz, Espagne, 984-994, III.