Compte-rendu de la journée d'étude de l'ATALA à Grenoble le 22 octobre 2004

Lionel Lamadon

▶ **To cite this version:**

Lionel Lamadon. Compte-rendu de la journée d'étude de l'ATALA à Grenoble le 22 octobre 2004. ALSIC - Apprentissage des Langues et Systèmes d'Information et de Communication, 2004, 07 (1), pp.203-210. edutice-00001471

HAL Id: edutice-00001471

https://edutice.hal.science/edutice-00001471

Submitted on 19 Apr 2006

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

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

Apprentissage des Langues et Systèmes d'Information et de Communication

alsic.org ou alsic.u-strasbg.fr
Volume 7, 2004
article mis en ligne en décembre 2004
pp 203 - 210
Points de vue

Compte-rendu de la journée d'étude de l'ATALA à Grenoble le 22 octobre 2004

Lionel LAMADON
Université Blaise Pascal, Clermont-Ferrand, France

- 1. Introduction
- 2. Utilisation des outils TAL à des fins diagnostiques ou correctives
- 3. Modélisation de l'apprenant
- 4. Comment intégrer des outils TAL dans un environnement d'ALAO ?
- 5. Utilisation d'outils TAL pour mettre des aides à la disposition de l'apprenant
- 6. Conclusion
- Références

1. Introduction

La journée d'étude de l'ATALA (Association pour le Traitement Automatique de la Langue) s'est déroulée le 22 octobre 2004 à l'université Stendhal Grenoble 3 et avait pour thème : TAL et apprentissage des langues[1]. Cette journée était organisée par le LIDILEM (Laboratoire de Linguistique et Didactique des Langues Étrangères et Maternelles) de l'université Stendhal, sous l'égide de l'ATALA et avec le soutien de XRCE (*Xerox Research Centre Europe*) et du projet européen *Thetis*.

Un nombre important de soumissions a permis d'emblée de mesurer l'intérêt suscité par le thème de cette journée. Au final, dix articles ont été sélectionnés pour une présentation orale ainsi que quatre articles pour la session *posters*. Autour de la thématique générale des liens entre le TAL (Traitement Automatique de la Langue) et l'ALAO (Apprentissage des Langues Assisté par Ordinateur), les intervenants ont présenté différents axes de recherche permettant de mieux cerner les enjeux de l'intégration d'outils de TAL dans des environnements informatiques d'aide à l'apprentissage. Ces interventions ont également permis de mettre à jour le caractère fondamentalement interdisciplinaire d'un tel type de recherche. En effet, au lieu de se concentrer sur un seul des domaines en jeu (TAL, didactique, linguistique), l'accent a été mis sur le besoin d'articuler une problématique particulière - liée à la résolution de tâches spécifiques - autour d'une réflexion plus générale intégrant chacune des différentes composantes (informatique, didactique et linguistique). Cette volonté de décloisonner des domaines qui, jusqu'à présent, avaient plutôt tendance à s'ignorer est le signe d'un changement profond dans la manière

d'aborder la question que se proposait de traiter cette journée.

En raison de la multitude et de la variété des sous-thèmes abordés durant cette journée, ce compte-rendu ne vise pas à l'exhaustivité. Il se présente plutôt comme une synthèse des différents points de vue théoriques adoptés et des différentes réalisations pratiques proposées. De la même façon, nous tenterons d'illustrer la richesse actuelle du domaine de l'ALAO en montrant la variété des utilisations d'outils de TAL en lien avec des préoccupations d'ordre didactique et linguistique.

Les outils TAL sont souvent mis au service de l'apprentissage à différents niveaux (analyse, génération, indexation, etc.). A ce titre, leur utilisation dans des environnements informatiques d'apprentissage pose des problèmes de nature différente. De même, leur évaluation - tant au niveau informatique que didactique - se doit d'être effectuée de manière spécifique en fonction de la tâche à réaliser. Nous aborderons ces questions en organisant ce compte-rendu autour de la présentation de différentes utilisations d'outils TAL dans des environnements d'ALAO, en soulignant des questions relatives à leur intégration et aux objectifs visés.

2. Utilisation des outils TAL à des fins diagnostiques ou correctives

Parmi les tâches susceptibles d'être menées à bien par le biais d'outils TAL dans un environnement d'ALAO figure traditionnellement en tête celle de l'analyse automatique de productions (libres ou non, écrites ou orales) d'apprenants. Plusieurs tentatives pour réaliser cette tâche d'analyse ont déjà vu le jour et ont permis de montrer à la fois sa complexité et ses limites. Le problème posé par l'analyse des erreurs est récurrent pour les outils TAL et il est d'autant plus prégnant lorsqu'il s'agit d'envisager une intégration de ces outils dans des environnements informatiques d'apprentissage. En effet, dans ce type d'environnement et notamment pour un apprentissage en autonomie, l'ordinateur est souvent la seule base à partir de laquelle l'apprenant a la possibilité d'améliorer sa production ; et, si l'analyse effectuée et la rétroaction proposée ne sont pas extrêmement précises, le recours aux outils TAL apparaît peu utile pour les apprenants qui ont tendance à ne plus faire confiance à la machine.

Plusieurs présentations ont mis l'accent sur la nécessité de définir précisément le type de techniques de traitement automatique et la nature des ressources linguistiques à utiliser pour assurer une analyse efficace et une rétroaction fiable.

Dans le cadre du projet européen *FreeText* [FreeText02], visant à développer un "prototype de logiciel d'apprentissage" du FLE, différentes techniques de diagnostic et de détection d'erreurs ont été implémentées [LhaireVandeventer03]. En plus d'une correction orthographique, assurée par une version adaptée du correcteur orthographique *Hugo 2000* [Hugo 2000], un diagnostic d'erreurs syntaxiques est effectué grâce à différentes fonctionnalités développées à l'intérieur de l'analyseur *FIPS* [GaudinatWehrli97]. Les techniques du "relâchement de contraintes" et de la "réinterprétation phonologique" offrent ainsi des solutions dynamiques pour marquer le lieu et le type de l'erreur. L'avantage majeur de

ces deux techniques est de donner des alternatives au processus d'analyse automatique en testant des hypothèses sur la nature des erreurs. En revanche, le nombre parfois très élevé de structures potentielles générées ralentit considérablement le système.

L'idée de tester des hypothèses pour mener à bien l'analyse d'un input imparfait produit par un apprenant est également présente dans le système présenté par J. Michel et J. Lehuen. A l'aide d'un "analyseur hypothético-déductif non déterministe" fondé sur une "grammaire sémantique", le système est capable de générer des hypothèses lorsqu'il trouve des éléments inconnus ou erronés dans les énoncés qu'il analyse.

Concernant l'oral, le logiciel *WinPitch LTL* [WinPitch] pour l'enseignement de la prosodie intègre aussi des fonctionnalités permettant d'analyser des erreurs éventuelles en comparant automatiquement la production d'un apprenant à un modèle.

Au niveau de la rétroaction et de l'interface du diagnostic, ces différents systèmes proposent plusieurs fonctionnalités permettant à l'apprenant de mieux saisir la nature des erreurs commises en lui apportant des aides spécifiques : "grammaire en couleur, analyse sous forme d'arbres syntaxiques" [LhaireVandeventer03], "visualisation de facteurs prosodiques, intégration de commentaires" (P. Martin), etc.

3. Modélisation de l'apprenant

L'utilisation d'outils TAL pour définir le profil d'un apprenant en phase d'acquisition d'une langue étrangère constitue un type d'approche particulier de la question des liens entre TAL et apprentissage des langues. L'emploi d'outils d'analyse stylistique comme le *Littératron* [Ganascia01] et l'analyseur de textes linéaires de J. Vergne (1999) pour "extraire automatiquement des motifs syntaxiques récurrents" dans des productions d'apprenants permet de "dégager les caractéristiques des apprenants", selon leur niveau, et de les distinguer des locuteurs natifs (I. Audras & J.-G. Ganascia). Les outils de traitement automatique ainsi utilisés facilitent le repérage d'erreurs caractéristiques dans les productions écrites d'un apprenant ; erreurs à partir desquelles un *diagnostic cognitif* et une *modélisation* de l'apprenant pourront être établis. Sur la base de ce diagnostic, des exercices spécifiques de remédiation seront ensuite proposés à l'apprenant.

Une autre démarche pour tenter de modéliser le profil d'un apprenant et ainsi optimiser l'acquisition d'une langue étrangère consiste à utiliser l'"analyse sémantique latente" (V. Zampa). Le prototype *RAFALES* (Recueil Automatique Favorisant l'Acquisition d'une Langue Étrangère de Spécialité), qui utilise ce type d'analyse, a pour but de fournir à l'apprenant des lectures en rapport avec un domaine de spécialité, en fonction de ses connaissances du domaine. L'idée centrale est que l'acquisition peut être favorisée si l'on sélectionne pour l'apprenant des textes qui sont à la "proximité optimale d'acquisition".

Ces deux démarches originales se caractérisent par le souci de cibler au mieux les besoins des apprenants selon le type de compétence visée (expression écrite ou lecture). Dresser ainsi un profil de l'apprenant conduit à une meilleure individualisation de l'apprentissage par la prise en compte des connaissances propres à chacun.

4. Comment intégrer des outils TAL dans un environnement d'ALAO ?

Cette question revient sans cesse dès lors qu'on aborde la problématique des liens entre TAL et ALAO. Car, il est toujours utile de rappeler que les outils de traitement automatique intégrés dans des environnements informatiques d'aide à l'apprentissage sont le plus souvent utilisés par des non spécialistes de TAL. A cet égard, la présentation du projet *MIRTO* (Antoniadis & al.) - projet de plateforme de création d'activités pédagogiques fondé sur des outils TAL, développé au LIDILEM - a permis de mettre l'accent sur l'importance d'une réflexion sur les modalités d'intégration des outils TAL. Partant de l'idée que les **fonctions** TAL utilisées (*tokenisation*, lemmatisation, analyse syntaxique, etc.) se doivent d'être invisibles pour les utilisateurs, ce projet utilise la notion de **script** pour établir un lien entre des fonctions TAL et des objectifs pédagogiques. Grâce à la définition de certains **paramètres**, il est ainsi possible pour l'utilisateur de configurer des outils pour un usage didactique sans avoir au préalable de connaissances particulières des outils et de leur fonctionnement.

Le même souci de simplicité d'utilisation est présent dans *NooJ*, un "outil TAL de corpus pour l'enseignement des langues et de la linguistique" (M. Silberztein & A.Tutin). En améliorant certaines fonctionnalités du logiciel *INTEX* [Intexnd], les concepteurs ont mis au point un outil capable d'effectuer des recherches complexes sur des informations linguistiques fines à l'intérieur de vastes corpus. L'intégration de nouvelles fonctionnalités plus particulièrement destinées à l'enseignement des langues (constructions de *mini-applications pédagogiques*, définition de *projets*, etc.) a également eu pour conséquence de définir des moyens simples pour constituer et paramétrer des ressources linguistiques.

Dans le cadre du projet *MIRTO*, un axe de recherche particulier consiste à réfléchir à la manière d'organiser et "d'indexer une base de données textuelles destinée à un usage pédagogique" (M. Loiseau). A l'heure où des quantités de plus en plus importantes de données sont disponibles et, parfois, déjà mutualisées, il convient en effet de définir une méthodologie permettant d'accéder de manière rapide et pertinente à ces données. Faisant le constat que la plupart des "standards de métadonnées" disponibles sont inappropriés et trop généralistes pour décrire des textes destinés à être utilisés dans l'enseignement des langues, M. Loiseau propose de définir de nouveaux standards en fonction des besoins propres à ce domaine particulier et à ses acteurs. A terme, une telle base de données devrait être intégrée à la plateforme et alimenter celle-ci en ressources textuelles indexées pédagogiquement.

5. Utilisation d'outils TAL pour mettre des aides à la disposition de l'apprenant

L'utilisation d'outils TAL pour mettre des aides à la disposition de l'apprenant sous forme de ressources a constitué un autre thème majeur de cette journée. Plusieurs communications et *posters* ont abordé cette question en présentant différents systèmes.

L'outil *NEDERLEX* a ainsi été présenté comme un "outil en ligne d'aide à la lecture de textes en néerlandais destiné à des apprenants francophones" [DevilleDumortier03]. Les textes, correspondant à des documents écrits authentiques de différents types et de niveaux de difficulté différents, sont entièrement "glosés" et, par le biais d'un simple clic de souris sur un mot, l'apprenant a accès à la "traduction en contexte" de ce mot. Cet outil, adaptable à différentes langues, se différencie ainsi de la plupart des dictionnaires de traduction électroniques existants qui proposent souvent des traductions hors contexte. De plus, de nombreuses fonctionnalités permettant par exemple d'afficher un ensemble de concordances, de créer et d'éditer une leçon, de contrôler le lexique associé à une leçon sont intégrées à cet outil.

La présentation d'*ITOLDU*, site pédagogique pour l'enseignement de l'anglais technique, a permis de faire le point sur la question des ressources linguistiques (V. Bellynck, C. Boitet & J. Kenwright). Ce type de ressources est particulièrement long et complexe à élaborer. Pour cette raison, l'idée centrale du projet est de constituer une "base lexicale multilingue", accessible gratuitement, de manière collaborative. Enseignants et apprenants sont ainsi invités à constituer ensemble des ressources linguistiques pour un domaine particulier. Cette vision collaborative du travail doit permettre à terme d'éviter le coût important d'un recours à des spécialistes en lexicographie et en terminologie pour la construction de ce type de ressources.

D'autres présentations ont également montré comment l'utilisation d'outils spécifiques pouvait contribuer au développement d'environnements informatiques d'aide à la compréhension de l'écrit. Dans le cas de l'environnement *LyText*, on donne à l'apprenant la possibilité de visualiser des informations qui permettent "d'associer automatiquement à certains mots d'un texte une ou plusieurs définitions extraites du Trésor de la Langue Française informatisé" (C. Pélissier & E. Jacquey). Le système présenté par S. Labat-Jaqmin est destiné au "perfectionnement d'étudiants en langue italienne". Il se présente comme un post-traitement de texte qui, tout en marquant les erreurs de l'apprenant, lui propose différents types d'aides (fiche de vocabulaire, fiche de grammaire, exercices, etc.). Ce type d'outil prend ainsi en compte les spécificités de chaque apprenant en proposant des aides adaptées à la nature des erreurs commises.

6. Conclusion

Comme nous l'avons déjà signalé en introduction, l'intérêt majeur de cette journée fut de réunir en un même lieu didacticiens, informaticiens et linguistes autour d'un thème relevant des compétences des trois domaines. Ceci a permis à chacun de pouvoir définir clairement les besoins et les objectifs propres à sa spécialité avec la préoccupation centrale de mener à bien un travail qui se doit d'être collaboratif.

Au niveau informatique, nous avons pu voir notamment l'intérêt qu'il y avait, pour les tâches de diagnostic, à travailler avec des outils déclaratifs permettant d'adapter une modélisation linguistique particulière en fonction d'un problème pédagogique particulier. L'apport de systèmes à orientation plus statistique pouvant par ailleurs se justifier en vue d'éventuelles améliorations de certains systèmes d'analyse ; ainsi la définition de

certaines heuristiques basées sur des notions de fréquence ou d'analyse préférentielle permettrait d'éviter l'écueil de la surgénération lié à ces tâches de correction automatique. Dans le domaine de la génération, ces deux types de systèmes peuvent se compléter pour extraire et traiter des énoncés à partir de corpus afin, par exemple, de générer automatiquement différents types d'exercices. Enfin, il est important de noter que l'évaluation systématique des performances des outils TAL utilisés au sein d'environnements d'ALAO apparaît comme un élément capital dans la plupart des projets ou réalisations présentées. L'évaluation, parfois sensible dans le domaine du TAL, est la condition fondamentale pour envisager une intégration fiable et utile de ces outils dans ces environnements. Ce n'est qu'en connaissant parfaitement les imperfections d'un système que l'on peut évaluer la pertinence didactique de cette intégration. Dans cette perspective, tous les intervenants ont montré les avantages mais aussi les limites de ces outils par rapport à certaines attentes de la didactique des langues.

Au niveau linguistique, une attention toute particulière est portée à la nature des ressources utilisées et des modélisations implémentées. Plutôt que de mettre en avant des théories linguistiques générales, l'accent est mis sur des modélisations optimisées en fonction de tâches particulières. Les multiples accès à différents types d'aides lexicales, syntaxiques ou sémantiques, élaborées à partir de ressources linguistiques variées, constituent des fonctionnalités intégrées dans la plupart des environnements présentés. L'intérêt des corpus électroniques pour l'enseignement des langues a une nouvelle fois été démontré et approfondi ; la capacité des outils TAL à travailler sur des volumes textuels importants offrant de multiples possibilités aux enseignants de langue pour utiliser des ressources linguistiques riches dans différents types d'activités pédagogiques. Enfin, il est intéressant de noter que, de plus en plus, les questions d'ordre linguistique ne se limitent pas à l'analyse morphologique et syntaxique ou à la reconnaissance et la synthèse de la parole ; en effet, en prenant conscience, par exemple, que la vérification syntaxique d'une phrase ne suffit pas pour vérifier la cohérence d'une réponse donnée par un apprenant, de nombreux systèmes tentent d'intégrer des traitements de nature sémantique, même si aucun modèle d'analyse présenté n'a pour ambition de répondre à une problématique générale concernant l'analyse sémantique des langues naturelles.

Au niveau didactique, plusieurs questions ont été soulevées. La plupart des intervenants ont défini précisément les critères qui permettent de définir au mieux la pertinence de l'utilisation des outils TAL intégrés dans des environnements d'ALAO ; en cherchant, par exemple, à savoir à quels besoins pédagogiques particuliers peuvent répondre ces outils. De même, les questions concernant le type de compétence visée, la nature des tâches proposées et le type de public concerné apparaissent comme centrales du point de vue de la didactique des langues. Les différentes présentations ont cherché à bien faire la distinction entre les systèmes proposés pour un apprentissage en présentiel et ceux plutôt destinés à être utilisés de manière autonome par l'apprenant. Les problèmes relatifs à la rétroaction notamment ne se posant pas de la même façon dans les deux cas. Il importe également de définir précisément le type d'utilisateur pour chaque environnement : la question étant de savoir si les outils TAL vont être utilisés pour aider l'apprenant en phase d'acquisition ou pour aider l'enseignant dans la mise au point de ses contenus d'enseignement. On a pu également remarquer que des préoccupations d'ordre

ergonomique étaient toujours présentes. La prise en compte des particularités liées aux domaines de compétences des utilisateurs implique souvent, de la part des concepteurs d'environnements d'ALAO, une réelle volonté de simplifier et de paramétrer les outils TAL intégrés dans ces environnements.

En conclusion, on peut dire que cette journée a permis de souligner les difficultés que rencontre le décloisonnement des deux domaines du TAL et de l'ALAO. En effet, on a pu constater des différences dans les approches méthodologiques et une certaine méconnaissance des outils TAL de la part des didacticiens, d'où une certaine difficulté pour ces derniers à énoncer des besoins précis en rapport avec l'utilisation de ces outils. À l'inverse, du côté des spécialistes de TAL, une méconnaissance de la complexité des situations d'enseignement/apprentissage des langues les amène à ne pas toujours cerner précisément les besoins des didacticiens. Il apparaît cependant que les spécialistes des deux domaines ont parfaitement compris l'intérêt qu'il y avait à travailler de façon collaborative autour de cette problématique de l'intégration d'outils TAL dans des environnements d'ALAO.

Références

Les liens externes étaient valides à la date de publication.

Bibliographie

[Ganascia01]
Ganascia, J.-G. (2001). Extraction automatique de motifs syntaxiques. Actes de la conférence Traitement Automatique du Langage Naturel (TALN'2001). Tours. http://www.up.univ-mrs.fr/veronis/Atala/TALN/pdf/art17_p193_202.pdf

[GaudinatWehrli97]
Gaudinat, A. & Wehrli, E. (1997). "Analyse syntaxique et synthèse de la parole: le projet FIPSVox", *TA-Informations*.

[LhaireVandeventer03]
L'Haire, S. & Vandeventer Faltin, A. (2003). "Diagnostic d'erreurs dans le projet FreeText. Apprentissage des langues et systèmes d'information et de communication", *Alsic*, vol. 6, 1. pp. 21-37. Consulté en décembre 2004. http://alsic.u-strasbg.fr/Num11/lhaire/alsic_n11-rec3.htm

Sites Internet

[Atala04]
Actes de la journée d'étude de l'ATALA – Grenoble 3. Consulté en décembre 2004. http://www.u-grenoble3.fr/lidilem/talal/actes.php?PHPSESSID=8a07d3e52cb7c87987c544aeae81198b

[FreeText02]
FreeText (2002). *FreeText : French in Context: An advanced hypermedia CALL system*

featuring NLP tools for a smart treatment of authentic documents and free production exercises. Genève : Université de Genève. Consulté en décembre 2004. http://www.latl.unige.ch/FreeText/

[Intexnd]
INTEX/NooJ. Consulté en décembre 2004. http://intex.univ-fcomte.fr

[WinPitch]
WinPitch LTL (2004). Consulté en décembre 2004. http://www.winpitch.com

[DevilleDumortier03]
Deville, G. & Dumortier, L. (2003). Site du cours en ligne Tussen de Regels – deel II. Néerlandais : Lecture de textes juridiques. Consulté en décembre 2004. http://obelix.droit.fundp.ac.be/droit2/index.php

Logiciels

[Hugo 2000]
Hugo 2000, Correcteur orthographique et grammatical, Editeur : Softissimo

Notes

[1] On trouvera l'intégralité des communications de cette journée sur le site [Atala04].

À propos de l'auteur du compte-rendu

Lionel LAMADON est actuellement doctorant en linguistique informatique. Après une expérience dans le domaine de l'enseignement du FLE, il s'intéresse désormais à l'adaptation d'outils TAL pour l'apprentissage/enseignement des langues.
Courriel : lionel.lamadon@univ-bpclermont.fr
Adresse : Laboratoire de Recherche sur le Langage (L.R.L.), Université Blaise Pascal, 4, rue Ledru, 63057 Clermont-Ferrand - Cedex 01, France

Référence de l'article :
Lamadon, L. (2004). "Compte-rendu de la journée d'étude de l'ATALA à Grenoble le 22 octobre 2004". Apprentissage des langues et systèmes d'information et de communication (ALSIC), vol. 7, 1, décembre 2004, pp. 203-210. http://alsic.u-strasbg.fr/v07/lamadon/alsic_v07_12-poi1.htm, mis en ligne le 22/12/20004.

© Apprentissage des Langues et Systèmes d'Information et de Communication, décembre 2004