

Du bruit, du silence et des ambiguïtés : que faire du TAL pour l'apprentissage des langues ?

Olivier KRAIF, Claude PONTON

LIDILEM - Laboratoire de linguistique et didactique des langues étrangères
et maternelles (<http://www.u-grenoble3.fr/lidilem/labo>)
{Olivier.Kraif,Claude.Ponton}@u-grenoble3.fr

Résumé. Nous proposons une nouvelle approche pour l'intégration du TAL dans les systèmes d'apprentissage des langues assisté par ordinateur (ALAO), la stratégie « moins-disante ». Cette approche tire profit des technologies élémentaires mais fiables du TAL et insiste sur la nécessité de traitements modulaires et déclaratifs afin de faciliter la portabilité et la prise en main didactique des systèmes. Basé sur cette approche, ExoGen est un premier prototype pour la génération automatique d'activités lacunaires ou de lecture d'exemples. Il intègre un module de repérage et de description des réponses des apprenants fondé sur la comparaison entre réponse attendue et réponse donnée. L'analyse des différences graphiques, orthographiques et morphosyntaxiques permet un diagnostic des erreurs de type fautes d'orthographe, confusions, problèmes d'accord, de conjugaison, etc. La première évaluation d'ExoGen sur un extrait du corpus d'apprenants FRIDA produit des résultats prometteurs pour le développement de cette approche « moins-disante », et permet d'envisager un modèle d'analyse performant et généralisable à une grande variété d'activités.

Abstract. This paper presents the so-called "moins-disante" strategy, a new approach for NLP integrating in Computer Assisted Language Learning (CALL) systems. It is based on the implementation of basic but reliable NLP techniques, and put emphasis on declarative and modular processing, for the sake of portability and didactic implementation. Based on this approach, ExoGen is a prototype for generating activities such as gap filling exercises. It integrates a module for error detection and description, which checks learners' answers against expected ones. Through the analysis of graphic, orthographic and morphosyntactic differences, it is able to diagnose problems like spelling errors, lexical mix-up, error prone agreement, bad conjugations, etc. The first evaluation of ExoGen outputs, based on the FRIDA learner corpus, has yielded very promising results, paving the way for the development of an efficient and general model tailored to a wide variety of activities.

Mots-clés : ALAO, apprentissage des langues, diagnostic d'erreur, feed-back d'erreur.

Keywords: CALL, language learning, error diagnosis, error feedback.

1 Introduction

Dans le cadre de l'ALAO (Apprentissage des Langues Assisté par Ordinateur), et plus particulièrement pour les systèmes dits « de structure » (systèmes de répétitions, d'entraînement, tutoriels, etc. par opposition aux systèmes de référence et aux systèmes

d'exploration, cf. Wyatt, 1987; Meunier, 2000), la détection et l'analyse d'erreurs constituent un élément central pour un diagnostic et une production de rétroactions adaptées permettant ainsi un apprentissage interactif, autonome et personnalisé. Or, la plupart de ces systèmes actuels ne disposent que de techniques rudimentaires de tests d'identité de chaînes entre « réponse donnée » et « réponse attendue ». Ces approches ne peuvent conduire qu'à de simples rétroactions du type vrai/faux n'offrant pas ainsi à l'apprenant une possibilité de réflexion sur les stratégies qu'il a mises en œuvre et lui permettre ainsi de modifier sa production langagière. En traitant les différents niveaux de la langue, le TAL semble pouvoir offrir de meilleures perspectives à cette problématique. Toutefois, comme le fait remarquer J. Rézeau (2001), « (...) on constate que la quasi-totalité des didacticiels de langue sur le marché à la fin des années 1990 proposent des exercices du premier type (i.e. qui attendent une seule réponse), et donc une analyse de réponse que nous qualifierions de "minimale" (i.e. vrai/faux) ». Outre le manque de communication entre les chercheurs et praticiens des différents domaines, l'explication de cette désaffection du TAL en ALAO découle de diverses raisons liées aux contextes didactiques des systèmes et donc de l'analyse d'erreurs.

♦ **Analyse de textes libres** : Des systèmes comme FreeText (Granger et al., 2001) ou « le correcteur 101 didactique » de la société Machina Sapiens visent l'analyse d'erreurs sur des textes libres d'apprenants. Or, comme le reconnaît lui-même S. L'Haire, l'un des participants au projet FreeText, le système pêche par « une trop grande surdétection d'erreurs » (L'Haire, 2004). La même critique s'applique au correcteur 101 Didactique. Par exemple, l'analyse par ce logiciel d'un texte d'apprenant de niveau intermédiaire issu du corpus FRIDA (Granger et al., 2001) produit :

« Dans tout le monde, il y a plus que plusieurs langues étrangères [étranger]. Ce pour sa [ça], qui parler [parle] ou connaître plusieurs de langues est [sont] nécessaire. Mais [mai ?], je crois qui parler trois ou quatre langues sont suffisents, parce qui n'est pas possible étudier [étudie?] tout le [toutes les] langues. » (détection en souligné, correction proposée entre crochets)

Si la détection des mots inconnus (*quatre, suffisents...*) est relativement bien traitée, le traitement de l'homophonie indique fréquemment des erreurs inexistantes. Seuls les accords de courtes portées sont bien analysés et des erreurs grossières d'analyse syntaxique (* *plusieurs langues étranger*) compromettent une utilisation en autonomie complète.

♦ **Analyse de textes sous contrôle** : Pour éviter les travers d'une analyse de textes libres, certaines approches ont tenté de contrôler divers paramètres de la production. Par exemple, le système ALEXIA (Selva, Chanier, 2000) travaille sur un domaine ciblé (i.e. celui de l'emploi et du chômage) alors que le système ELEONORE (Rénié, Chanier, 1993) s'intéresse uniquement à la construction des interrogatives. En termes d'analyse et de feed-back, des critiques subsistent certainement mais les résultats restent bien plus précis et pertinents que pour l'analyse de textes libres. Le problème se situe davantage au niveau du ratio coût de développement/apport, d'autant plus que ces démarches semblent peu généralisables et exportables à d'autres types d'activités.

Ce tour d'horizon des diverses tentatives d'utilisation du TAL dans la détection/correction des fautes ne serait pas complet si nous ne citons pas les correcteurs orthographiques et grammaticaux. Contrairement aux systèmes évoqués, ces correcteurs n'ont pas une visée didactique, mais ils ont cependant donné lieu à différentes tentatives d'utilisation en classe de langue. Le constat semble relativement partagé (Cordier-Gauthier, Dion, 2003 ; Charnet, Panckhurst, 1998; Désilets, 1998); la relative qualité des analyses tant par le bruit et le silence générés que par les rétroactions inadéquates de ce type de logiciels permettent à

Que faire du TAL pour l'apprentissage des langues ?

l'enseignant d'amorcer une réflexion avec les élèves autour des fautes détectées mais n'autorisent en aucun cas un travail en autonomie. En résumé, l'analyse automatique d'erreurs d'apprenants se heurte à deux problèmes majeurs. D'une part, des analyses TAL peu fiables car trop bruitées ou trop silencieuses et, d'autre part, comme le souligne également J. Rézeau (2001), du fait des coûts de recherche et de développement, ces projets restent souvent à l'état de prototypes voire de simples spécifications. Face à ces blocages, nous proposons une nouvelle approche du problème, la stratégie « moins-disante ». Les parties 3 et 4 de cet article seront consacrées à l'étude des premiers résultats d'ExoGen, un système fondé sur cette stratégie.

2 Stratégie « moins-disante »

La partie précédente de cet article met en évidence les faiblesses et les blocages actuels de l'analyse d'erreurs. Toutefois, nous tirons deux constats principaux de ces diverses expériences. Le premier concerne la bonne fiabilité des analyses TAL sous certaines conditions : cadre didactique contrôlé, lemmatisation, traitements des accords de courte portée, etc. Le deuxième constat prend appui sur les retours d'enseignants à l'usage de ces systèmes. L'exemple suivant en est relativement significatif :

« Le Correcteur 101 est intéressant du fait qu'il amène l'élève à se questionner sur sa phrase et ses erreurs. Le logiciel n'offre que très rarement la réponse. Deux outils sont aussi offerts: un dictionnaire et une grammaire. Ils peuvent s'avérer très utiles pour dépanner. Le langage utilisé n'est pas conforme à la grammaire nouvelle, ce serait un atout pour permettre un lien à la grammaire étudiée en classe. Parfois, le logiciel ne mentionne pas certaines fautes. »¹

La détection des erreurs sans correction constitue un atout pédagogique en soi car il amène les apprenants à réfléchir sur leurs erreurs. La production de rétroactions (aides, explications, etc.) adaptées et didactiquement pertinentes facilitent cette réflexion et la recherche de solutions. Or, la qualité de ces rétroactions étant étroitement liée à la détection et à l'analyse des erreurs, ces deux phases doivent être contrôlées pour assurer une fiabilité et une précision maximales. La stratégie « moins-disante » que nous développons est un corollaire de ces constats. C'est une approche empirique qui se base sur les hypothèses suivantes :

- Il existe un ensemble de technologies TAL fiables (tokenisation, lemmatisation, étiquetage morpho-syntaxique...) pouvant fournir des éléments de détection et d'analyse des erreurs pertinents.
- L'analyse hors-contrôle du contexte de production n'est actuellement pas envisageable. Il est donc nécessaire de maîtriser, d'un point de vue final (i.e. didactique), les insuffisances du TAL. Dans notre cas, la connaissance des réponses attendues, qui permet des comparaisons, de lever des ambiguïtés, de cibler les analyses, etc., constitue une donnée du système.
- La fiabilité à 100% des analyses reste, et à notre avis restera, un objectif inaccessible. Pour éviter ce travers, nous préférons privilégier « l'aide à... » à l'automatisation des procédés. Effectivement, dans ce contexte, le TAL nous semble pouvoir améliorer sensiblement les différents niveaux de la chaîne d'ALAO : aide au choix d'exemples

¹ Critique d'une enseignante sur son utilisation du Correcteur 101 didactique en classe (http://c-rdi.qc.ca/products/aff_fiche.asp?fiche=426, consulté le 31/01/2007). A noter que les autres critiques postées sur ce site vont globalement toutes dans ce sens.

(Antoniadis et al., 2007), aide à la génération d'activités (Antoniadis et al., 2005) ou encore l'aide à l'autocorrection (évaluation formative) qui constitue le sujet de cet article.

- En cas d'ambiguïté, conserver la multiplicité des résultats pour une prise de décision au niveau didactique
- Développer une approche déclarative et modulaire des traitements pour permettre une évolution du système et de ses ressources.

La mise en œuvre de notre approche (cf. figure 1) est comparable à la reprise, par D. Anctil (2005), de « la stratégie de résolution de problèmes » développée par T. Andre (1986).

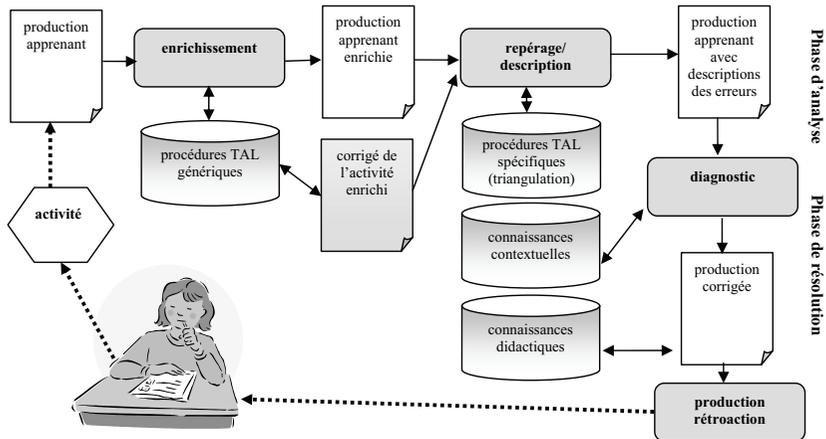


Figure 1 : Mise en œuvre de la stratégie « moins-disante »

Il s'agit de séparer la phase d'analyse du problème (i.e. l'erreur de l'apprenant pour notre système) de sa phase de résolution ; cette dernière passe par la production de rétroactions adaptées qui feront l'objet d'études ultérieures en collaboration avec des enseignants et des didacticiens. La phase de résolution permet, quant à elle, une distinction entre repérage, description et analyse de ce problème. Comme chez D. Anctil, nous avons réuni les étapes de repérage et de description du problème. En se basant sur un prétraitement² générique à la fois de la production de l'apprenant et du corrigé de l'activité, cette phase consiste en une analyse fine des différences entre réponses données (RD) et réponses attendues (RA). La désambiguïsation de ces analyses repose sur une triangulation entre RD, RA et le contexte.³

L'étape d'analyse du problème (appelée diagnostic) consiste, en fonction de connaissances sur le contexte de production (type d'activité, modèle apprenant, etc.) à rechercher les causes

² Il s'agit ici d'un enrichissement de la production de l'apprenant à l'aide de procédures de tokenisation, d'analyse morpho-syntaxique, etc.

³ Par exemple, pour une activité portant sur l'accord du participe passé après l'auxiliaire avoir, si la forme correcte comporte des flexions (féminin ou pluriel) on peut en déduire que le verbe est transitif et qu'il existe un COD précédant le participe passé : ces inférences, étroitement liées à l'activité, permettent de désambiguïser les résultats d'une analyse classique (étiquetage) et d'appliquer ensuite des règles simplifiées pour le diagnostic.

Que faire du TAL pour l'apprentissage des langues ?

potentielles de l'erreur et à sélectionner les informations pertinentes et fiables permettant un calcul de la rétroaction.

3 L'exemple du système ExoGen

Pour valider notre approche, nous avons développé un prototype, nommé ExoGen, qui permet la génération d'activités à partir d'un corpus de textes lemmatisés et étiquetés. Les activités proposées sont de type "lecture d'exemples" et "exercices lacunaires", et sont fondées sur l'extraction aléatoire de phrases contenant des patterns d'expressions régulières portant sur les formes, les lemmes, les catégories et les traits morphosyntaxiques (Kraif, 2006). Ces patterns permettent par exemple d'identifier des constructions telles qu'un participe passé dans un contexte de passé composé avec l'auxiliaire avoir, ou une flexion irrégulière pour le pluriel d'un nom, etc. A chaque génération d'activité, on peut obtenir un exercice lacunaire portant par exemple sur l'accord du participe passé avec l'auxiliaire avoir, les participes passés ayant été escamotés et remplacés par leur forme lemmatisée. Ce modèle de génération d'activité, très simple (analogue à celui du système Alfalex), se prête bien à l'analyse d'erreur précédemment décrite, basée sur la comparaison entre la réponse attendue RA et la réponse donnée RD.

Dans cette première implémentation, nous n'avons pas recouru au principe de triangulation, selon lequel l'analyse de la réponse attendue permettrait de désambiguïser à la fois le contexte et la réponse donnée, en vue d'appliquer une règle déclarative de diagnostic. Nous nous sommes contentés d'appliquer une heuristique simple lors de la phase d'analyse des différences, qui permet de privilégier, en cas d'ambiguïté, les analyses maximisant la similitude entre RA et RD (dans l'idée que cette similitude n'est pas fortuite). Qui plus est, nous n'utilisons pas les données issues de l'étiquetage et de la lemmatisation (obtenus avec Treetagger), afin de montrer jusqu'où cette heuristique permet de désambiguïser sans traitement préalable.

L'analyse et le diagnostic sont donc basés sur la seule comparaison de RA et de RD, indépendamment de tout contexte. Pour comparer ces deux formes nous ne disposons que d'une seule source d'information externe (en dehors des formes elles-mêmes) : les analyses possibles des formes fléchies, données par le dictionnaire de formes fléchies mis en ligne par l'ABU (<http://abu.cnam.fr/>). Chaque entrée de ce dictionnaire est une forme fléchie simple, à laquelle sont associés un lemme et les analyses possibles en terme de combinaison de traits morphosyntaxiques (nombre, genre, personne, temps, mode, etc.). La figure 2 donne un échantillon de quelques enregistrements de ce dictionnaire.

glace	glacer	Ver:IPre+SG+P1:IPre+SG+P3:SPre+SG+P1:SPre+SG+P3:ImPre+SG+P2
glacé	glacer	Ver:PPas+Mas+SG
glacent	glacer	Ver:IPre+PL+P3:SPre+PL+P3
glacera	glacer	Ver:IFut+SG+P3
glaceraient	glacer	Ver:CPre+PL+P3

Figure 2 : Un extrait du dictionnaire de formes fléchies

L'analyse repose sur une hiérarchisation des différences observées entre RA et RD : on traite en priorité les différences les plus légères et les plus superficielles, c'est-à-dire celles qui requièrent le moins d'inférence. A priori, ce sont aussi celles qui mènent aux diagnostics les plus sûrs :

1. **Différences graphiques.** Espacements, majuscules, variantes graphiques (p. ex. ligatures telle que *oe* et *œ*), etc. Ces différences, peuvent donner lieu à la validation de RD comme étant correcte, sauf si l'exercice porte explicitement sur ces aspects (usage des majuscules, réforme de l'orthographe, ...).

2. **Différences orthographiques.** Si RD est absente du dictionnaire de formes fléchies, plusieurs cas peuvent être considérés :

- a) RD ressemble à RA
 - i/ RD est identique à RA si l'on néglige les signes diacritiques (accents, cédille)
 - ii/ RD est une forme inconnue du dictionnaire
- b) RD ne ressemble pas à RA
 - i/ RD est une forme inconnue du dictionnaire, voisine de certaines formes connues
 - ii/ RD est une forme inconnue du dictionnaire, et aucune forme voisine n'a été trouvée

La ressemblance peut être calculée par une fonction de Levenshtein, ou de recherche de la plus longue sous-chaîne commune (Kraif, 2001). Les formes voisines peuvent être trouvées grâce à une fonction de hachage dont les clés correspondraient à une écriture simplifiée (sans lettre double, sans accent, sans finales muettes, avec réduction des variantes graphiques, etc.). Chacun de ces cas peut donner lieu à un feed-back spécifique, en fonction du contexte didactique. Par exemple pour 2.a.i, on pourrait avoir "Accentuation incorrecte", pour 2.a.ii "Faute d'orthographe", pour 2.b.ii "Pensiez-vous à une de ces formes : [formes voisines]", etc.

3. **Différences morphosyntaxiques.** Si RD est connue dans le dictionnaire, on peut calculer ses lemmes potentiels, ainsi que les catégories et structures de traits afférentes. La comparaison peut alors porter sur chacun de ces aspects :

- a) RD et RA correspondent à un même lemme
 - i/ RD et RA correspondent à une même catégorie. Dans ce cas, seuls leurs traits les différencient.
 - ii/ RD et RA correspondent à un même lemme, mais avec des catégories différentes. P.ex. RD="êtes" (être-Nom-PL), RA="êtes" (être-Ver-P2-PL).
- b) RD et RA correspondent à des lemmes différents
 - i/ RD et RA correspondent à une même catégorie avec les mêmes traits (p. ex. RA="trouverons" vs RD="penserons").
 - ii/ RD et RA correspondent à une même catégorie avec des traits différents (p.ex. RA="conscience", RD="connaissances")
 - iii/ RD et RA correspondent à des catégories différentes (p. ex. RD="mieux" (Adv) vs RA="préférable" (Adj)).

L'étude des cas est donc guidée par les similarités, le long d'un continuum allant de l'identité à la différence complète. Notre heuristique se fonde sur l'idée sous-jacente que les similarités sont rarement fortuites, tandis que les différences, elles, sont plus difficiles à systématiser. On peut en tirer une méthode de désambiguïsation lors de la comparaison des traits ou des catégories. Par exemple :

RA : <i>si j'avais su</i>	Catégorie : Ver	Traits : IImp+SG+P1	ou	IImp+SG+P2
RD : <i>si j'aurais su</i>	Catégorie : Ver	Traits : CPre+SG+P1	ou	CPre+SG+P2

Ici, on observe des différences au niveau du temps/mode ainsi que pour la personne : P1 ≠ P2, CPre ≠ IImp. On peut donc avoir 4 analyses différentes pour le couple (RA, RD). Grâce à l'heuristique de moindre différence, on ne compare que les analyses les plus proches (entre pointillés), ce qui permet de ne retenir qu'une seule différence de trait, entre le conditionnel

Que faire du TAL pour l'apprentissage des langues ?

présent et l'imparfait : CPre \neq IImp. Le feed-back correspondant pourrait être "Dans ce contexte, utilisez un imparfait plutôt qu'un conditionnel présent".

L'analyse pourrait se poursuivre sur le plan sémantique : lorsqu'on trouve deux lemmes distincts, de même catégorie, on peut évaluer leur proximité sémantique, par exemple en se basant sur une ressource du type dictionnaire de synonymes, thésaurus, réseau sémantique, etc. (cette fonctionnalité n'est pas encore implémentée dans notre prototype). Dans le cas d'unités polysémiques, l'heuristique de moindre différence permet encore de désambiguïser. Par exemple, si RA="pomme de pin", RD="pignon", et si le dictionnaire propose plusieurs acceptions pour "pignon" (/fruit/, /engrenage/) la comparaison RD/RA permettra de choisir l'acception la plus proche (i.e. le chemin le plus court à travers le graphe représentant les relations sémantiques).

4 Évaluation

Pour évaluer cette méthode simple d'analyse et de désambiguïation des erreurs, il nous faudrait un corpus de réponses d'apprenants, obtenues dans le cadre d'activités de type QROC (questions à réponse ouverte courte), telles que des exercices lacunaires ou des quiz. Pour chaque réponse donnée, on pourrait ainsi appliquer l'analyse des erreurs en fonction de la réponse attendue. Bien que la constitution d'un corpus de ce genre soit prévue dans les développements ultérieurs d'ExoGen, nous ne disposons pas encore de telles données empiriques. Pour l'évaluation, nous avons donc utilisé une autre ressource, à savoir des exemples issus du corpus FRIDA (*FRench Interlanguage DAtabase*), constitué dans le cadre du projet Freetext (Granger et al., 2001). Il s'agit d'un corpus de rédaction d'apprenants de différents niveaux et de différentes langues maternelles pour lequel les erreurs ont été identifiées manuellement et balisées en fonction d'une typologie indiquant le domaine (morphologie, grammaire, lexicale, etc.), la catégorie (agglutination, graphie, genre, etc.) et la catégorie grammaticale de la forme erronée. Pour chaque erreur identifiée, les annotateurs ont indiqué une correction. Ce corpus permet donc d'extraire des couples forme erronée / forme corrigée comparables aux couples RD / RA issus des QROC. On en tire des exemples du type :

*(...) une seule monnaie (l'ECU) n'adresse pas bien au gouvernement anglais.
Forme erronée : adresse, Forme corrigée : convient*

Cette évaluation peut comporter un biais, car le rapport RD/RA n'est pas identique à celui Réponse erronée/Réponse corrigée : dans le cadre d'un CROQ, la réponse attendue et son contexte préexistent à la réponse donnée, tandis qu'ici, les réponses corrigées sont données *a posteriori*, en fonction d'une erreur et d'un contexte issu d'une production libre. Mais nous pensons que ce biais est limité du point de vue de l'analyse des différences, car le même genre d'écart est observé, et la méthode d'analyse est confrontée au même type d'ambiguïtés (de lemme, de catégorie, de traits, de sens).

Nous avons utilisé un échantillon de 47 productions d'apprenants anglophones, de niveau variable. Ont été retenues toutes les erreurs impliquant deux formes simples (du fait des limitations de notre dictionnaire), hors ponctuation, pour un total de 318 cas d'erreurs. Pour chaque erreur nous avons appliqué l'analyse des différences, et obtenu des descriptions correspondant à 16 cas possibles, avec des précisions concernant les lemmes, catégories ou traits identifiés. On obtient par exemple les sorties suivantes :

Exemple d'erreur	Description (obtenue automatiquement)
(...) avant de retourner (<i>arriver</i>) en Angleterre. et beaucoup d' échafaide (<i>échafaudages</i>).	Forme grammaticalement correcte (verbe infinitif), mais on attendait une autre forme. Orthographe erronée ou mot inconnu du dictionnaire.
Je dois me dépêcher (<i>dépêcher</i>).	Orthographe erronée : problème d'accent.
(...) sommes bien amusées et c'est vrai (<i>juste</i>) de dire que nous avons dansé assez bien	Forme grammaticalement correcte (adjectif ou adverbe ou nom masculin singulier), mais on attendait une autre forme
C'était désespéré (<i>désespérant</i>) mais c'était la seule chance (...)	S'il s'agit du verbe <i>désespérer</i> : cas 1 [masculin singulier] : On attend un participe présent et non un participe passé.
Pour moi l' I' (<i>cette</i>) image crée une ambiance délassante	Forme grammaticalement correcte sur le plan de la catégorie (déterminant), mais on attendait une autre forme avec d'autres traits.
le Premier ministre reste toujours un britannique (<i>Britannique</i>)	Exact, mais il faut une majuscule à l'initiale.

Tableau 3 : Exemples d'erreurs (corrigées entre parenthèses) et descriptions correspondantes

On constate que dans certains cas la désambiguïsation est partielle, ce qui n'empêche pas de donner une description pertinente. Pour une évaluation chiffrée des résultats, nous avons évalué manuellement la correction des affirmations liées aux différentes analyses. En outre, nous avons noté, pour tous les cas où les formes (erronées et corrigées) recelaient des ambiguïtés (analyses multiples), si la désambiguïsation est totale, partielle ou nulle.

	Tous les cas	Non ambigus	Complètement désambiguïsés	Partiellement désambiguïsés	Non désambiguïsés
Corrects	312	187	104	14	7
Incorrects	6	1	5	0	0
Précision	0,981	0,995	0,954	1	1

Tableau 4 : Evaluation de la correction des descriptions d'erreur

On constate que la précision est très satisfaisante. L'heuristique de désambiguïsation, opérante dans 1/3 des cas, aboutit très fréquemment à une désambiguïsation complète, avec un peu moins de 5% d'erreurs. Dans de nombreux cas, l'heuristique aboutit à une réduction spectaculaire des ambiguïtés :

une seul monnaie (l'ECU) n' adresse (convient) pas bien au gouvernement anglais.

Ici *adresse* peut correspondre à deux lemmes différents (*adresse* et *adresser*), deux catégories différentes (nom et verbe) et de nombreuses structures de traits (le dictionnaire en donne 6 : Nom:Fem+SG, Ver:IPre+SG+P1, IPre+SG+P3, SPre+SG+P1, SPre+SG+P3, ImPre+SG+P2). La comparaison avec *convient* permet de conserver la seule interprétation commune : verbe à l'indicatif présent, troisième personne du singulier. Quant aux analyses erronées, elles sont dues à deux phénomènes :

- Lacune du dictionnaire (2 cas) : dans l'exemple ci-dessous, le dictionnaire n'enregistre pas *futur* comme nom potentiel, mais seulement comme adjectif.

le futur (avenir) de l'Angleterre -> "On attendait une autre forme, d'une autre catégorie grammaticale (Nom#Adj)."

Que faire du TAL pour l'apprentissage des langues ?

- Mauvaise désambiguïisation (4 cas) : dans l'exemple suivant, la forme corrigée est interprétée comme le déterminant *tous*, et non comme un pronom :

l'heure de se joindre et de parler tout (tous) d'une voix -> "S'il s'agit du Det tout on a : cas 1 [Masculin] : On attend un pluriel et non un singulier."

Notons que même si la désambiguïisation est mauvaise, le feed-back donné à l'apprenant peut présenter une analyse comme hypothétique, ce qui évite d'affirmer une contre-vérité. Par ailleurs certaines ambiguïtés peuvent être réduites en sélectionnant les informations données à l'utilisateur. Prenons l'exemple suivant :

Soudain, nous avons entendus (entendu) un bruit -> "S'il s'agit du verbe entendre [participe passé masculin], on attend un singulier et non un pluriel ; S'il s'agit de l'adjectif entendu [masculin] on attend un singulier et non un pluriel ; s'il s'agit du nom entendu [masculin] on attend un singulier et non un pluriel."

L'analyse donne un résultat ambigu (verbe, adjectif ou nom) mais l'analyse concernant les traits est toujours la même, et peut aboutir au feed-back suivant : "on attend un singulier et non un pluriel". Sur la base de la stratégie moins-disante, on pourra se contenter de cette information, incomplète mais fiable, et centrée sur l'erreur commise par l'apprenant.

5 Conclusion

Nous avons présenté un cadre général pour l'analyse des réponses d'apprenant, basé sur la comparaison entre réponse donnée et réponse attendue, en se limitant à des productions contrôlées, tant sur le plan de la forme qu'au niveau du contexte didactique. Ces limitations permettent selon nous de mettre en œuvre des techniques fiables dont les analyses n'outrepassent pas les capacités actuelles des systèmes de TAL. Cette approche, qualifiée de "moins disante" permet selon nous, lorsque le contexte didactique est suffisamment spécifié (de la définition d'une consigne à la mise en œuvre de rétroactions adaptées), de diagnostiquer des fautes d'orthographe, des confusions, des problèmes d'accord, de conjugaison, etc. Dans cette optique, nous avons implémenté et évalué une méthode simple d'analyse des réponses, avec une heuristique de désambiguïisation tirant parti des écarts avec les réponses attendues. Les résultats sont encourageants, avec une précision supérieure à 98%.

L'étape suivante consistera à mettre en œuvre des règles de diagnostic, afin de déterminer les causes probables des erreurs (p.ex. un accord erroné d'un participe passé avec le sujet de l'auxiliaire avoir). Pour qu'un tel système soit généralisable et puisse s'adapter facilement à des activités variées, il est important de définir un langage simple et déclaratif pour la mise en œuvre des règles de diagnostic : c'est à notre avis une condition indispensable à l'appropriation de ces techniques par les pédagogues, qui seuls sont qualifiés pour interpréter les erreurs, déterminer leurs causes, et ensuite définir des rétroactions adaptées.

Pour y parvenir, nous prévoyons de développer des techniques de désambiguïisation plus fines, basées sur une triangulation entre RA/RD et leur contexte linguistique. Cette étape, dépassant les outils génériques tels que l'étiquetage et la lemmatisation, constituera un module de TAL relativement autonome, et généralisable à de nombreux types d'activités en ALAO.

Références

- ANCTIL D. (2005). *Maîtrise du lexique chez les étudiants universitaires : typologie des problèmes lexicaux et analyse des stratégies de résolution de problèmes lexicaux*. Mémoire de M.A. Faculté des Sciences de l'Éducation. Université de Montréal (Québec).
- ANDRE T. (1986). *Problem solving and education*. San Diego, CA: Academic Press.
- ANTONIADIS G., ECHINARD, S., KRAIF O., LEBARBÉ T. & PONTON C. (2005). Modélisation de l'intégration de ressources TAL pour l'apprentissage des langues : la plateforme MIRTO. *ALSIC*, vol. 8, n° 1. pp. 65-79.
- ANTONIADIS G., KRAIF O., PONTON C., ZAMPA V. (2007). Un outil exploratoire de corpus d'apprenants. Actes de *UNTELE'07*. Université de Compiègne. 29-31 mars 2007 (à paraître).
- CHARNET C., PANCKHURST R. (1998). Le correcteur grammatical : un auxiliaire efficace pour l'enseignant ? Quelques éléments de réflexion. *ALSIC*, vol. 1, 2. pp. 103-114.
- DÉSILETS M. (1998). Que penser de l'utilisation des logiciels correcteurs à l'école? *Vie pédagogique*. No 107. avril-mai 1998. 9-11.
- GRANGER, S., VANDEVENTER, A., HAMEL, M.-J. (2001). Analyse des corpus d'apprenants pour l'ELAO basé sur le TAL. *Traitement automatique des langues*. 42 (2). 609-621.
- KRAIF O. (2006) Extraction automatique de lexique bilingue : application pour la recherche d'exemples en lexicographie, *Journées du CRTT*, Université Lyon 2, Lyon.
- KRAIF O. (2001) Exploitation des cognats dans les systèmes d'alignement bi-textuel : architecture et évaluation, *TAL* 42 :3, ATALA, Paris, pp. 833-867.
- L'HAIRE S. (2004). Vers un feed-back plus intelligent, les enseignements du projet Freetext. Actes de la *journée TALAL*. Grenoble. 1-12.
- MEUNIER L. E. (2000). La typologie des intelligences humaine et artificielle : complexité pédagogique de l'enseignement des langues étrangères dans un environnement multimédia. *Apprendre une langue dans un environnement multimédia*. Les éditions Logiques : sous la direction de Duquette L. et Laurier M. Outremont (Québec). 211-253.
- RÉNIÉ D., CHANIER T. (1993). La modélisation de l'acquisition, une étape dans la construction de systèmes d'EIAO des langues: le cas des interrogatives en français langue seconde. *Environnements Interactifs d'Apprentissage avec Ordinateur*. Tome 1. Eyrolles. 123-134.
- RÉZEAU J. (2001). *Médiatisation et médiation pédagogique dans un environnement multimédia. Le cas de l'apprentissage de l'anglais en Histoire de l'art à l'université*. Thèse de troisième cycle. Université Bordeaux II, France.
- SELVA, T., CHANIER, T. (2000). Génération automatique d'activités Lexicales dans le système ALEXIA. *Sciences et Techniques Educatives, (STE)*, 7 (2). 385-412.
- WYATT D. H. (1987). Applying pedagogical principles to CALL courseware development. *Modern Media in Foreign Language Education*. Wm. Flint Smith (dir.). Lincolnwood, IL : National Textbook. 85-98.