

Diagnostic d'erreurs dans le projet FreeText

Sébastien L'HAIRE et Anne VANDEVENTER FALTIN

Université de Genève, Suisse

Résumé : Cet article présente une recherche, menée dans le cadre du projet FreeText, qui avait pour but de développer un système de diagnostic automatique d'erreurs pour apprenants du FLE. Après un bref résumé des caractéristiques principales du projet ainsi que du corpus d'apprenants récolté et exploité dans ce cadre, l'article traite du système de diagnostic d'erreurs, en particulier de deux de ses composantes, (1) un vérificateur syntaxique qui utilise deux techniques différentes pour détecter les erreurs de nature purement grammaticale et (2) un outil de comparaison de phrases qui compare la réponse de l'apprenant avec celle correcte stockée dans le système, dans le but de détecter des différences sémantiques possibles, comme les référents ou l'usage d'un mot. Puis nous comparons les résultats obtenus par notre système avec ceux d'un système commercial. Enfin nous discutons des avantages d'un système automatique de ce type et des pistes pour une recherche future.

- [1. Introduction](#)
- [2. Le projet FreeText](#)
- [3. Le corpus FRIDA](#)
- [4. Diagnostic d'erreurs](#)
- [5. Tests et comparaison du système](#)
- [6. Avantages attendus et perspectives de recherche](#)
- [Références](#)

1. Introduction

L'enseignement des langues assisté par ordinateur (ELAO) est un domaine qui attend beaucoup des outils de traitement automatique du langage naturel (TALN). Les domaines les plus vendeurs pour des logiciels d'ELAO sont la reconnaissance vocale et les synthétiseurs vocaux. Cependant, en ELAO, il y a de grands besoins en matière de diagnostic d'erreurs et de rétroaction immédiate et intelligente. Des systèmes de diagnostic d'erreurs fiables permettraient aux auteurs et utilisateurs de se libérer des contraintes imposées par les questionnaires à choix multiples et les exercices à trous, et aux apprenants de se voir offrir des tâches plus communicatives.

Dans le cadre du projet FreeText^[1], qui avait pour but de développer un système d'ELAO complet pour apprenants du français langue étrangère (FLE) de niveau intermédiaire à avancé, nous avons développé un système de diagnostic d'erreurs, composé d'un correcteur orthographique, d'un vérificateur syntaxique et d'un outil de comparaison de phrases. Les exercices et le système de diagnostic ont été basés sur l'étude d'un corpus d'apprenants du FLE. Dans cet article, nous discutons en détail des techniques utilisées pour le vérificateur syntaxique et pour l'outil de comparaison de phrases. Nous présentons également la sortie brute du système de diagnostic et quelques possibilités de la transformer en rétroaction qui fasse sens pour des apprenants de langue étrangère. Nous terminons par une présentation de quelques résultats des performances du système et par une comparaison de notre système avec un vérificateur grammatical existant.

Cet article est structuré comme suit : la section 2 donne un aperçu du projet FreeText ; la section 3 décrit le corpus FRIDA d'apprenants du FLE ; la section 4 détaille les techniques de diagnostic d'erreurs développées et implémentées dans le cadre de ce projet et montre comment la sortie de ce système peut être transformée en rétroaction intelligente ; la section 5 présente les résultats atteints par notre système sur des productions authentiques d'apprenants et les compare aux résultats d'un autre système ; enfin la section 6 présente quelques perspectives de recherches futures.

2. Le projet FreeText

Les recherches présentées dans cet article ont été menées dans le cadre du projet FreeText, qui avait pour but de développer un logiciel d'ELAO pour apprenants du FLE intermédiaires à avancés. FreeText contient quatre tutoriels présentant seize documents authentiques, allant du texte aux fichiers audiovisuels, qui illustrent différents actes communicatifs. Les exercices qui exploitent ces documents sont basés sur l'étude du corpus d'apprenants FRIDA ([voir section 3](#)), dans le but de traiter des erreurs réellement commises par le public cible. De plus, FreeText utilise des outils de TALN, comme un visualiseur de phrases ([voir section 4.3](#)), un synthétiseur vocal et un système de diagnostic d'erreurs, pour offrir une rétroaction intelligente pour des exercices de production (relativement) libre.

Le consortium FreeText regroupe des partenaires dans différents domaines d'expertise de quatre pays européens :

- le Department of Language Engineering, University of Manchester Institute of Science and Technology (UMIST), Royaume Uni, chargé du contenu didactique du logiciel ;
- le département de linguistique de l'Université de Genève, Suisse, qui développe des outils de TALN ;
- le Centre for English Corpus Linguistics (CECL), Université Catholique de Louvain, Belgique, responsable du travail autour du corpus et des différentes phases de validation ;
- Softissimo SA, Paris, France, pour l'intégration des différentes composantes dans une interface ergonomique.

FreeText est un système ouvert, où les enseignants peuvent introduire de nouveaux exercices et documents afin de répondre aux besoins particuliers de leurs apprenants. Le projet a démarré en avril 2000 et a pris fin en mars 2003. De plus amples informations sont disponibles sur le site du projet FreeText [[FreeText02](#)].

3. Le corpus FRIDA

Le corpus FRIDA (French Interlanguage Database) [[Granger03](#)] [[Granger01](#)] est un corpus d'apprenants du FLE recueilli et développé par le CECL dans le cadre du projet FreeText et rassemblé à partir de rédactions d'apprenants, qui contient quelque 450 000 mots. Une fois recueillis, les textes ont été saisis en format électronique afin de permettre leur traitement informatique. Les erreurs contenues dans les textes ont été ensuite codées d'après une typologie d'erreurs spécialement conçue pour les apprenants du français. Un trigramme d'étiquettes marque chaque erreur afin d'indiquer un domaine d'erreur, comme forme ou grammaire, une catégorie d'erreur, comme genre ou diacritique, et la catégorie lexicale du mot incorrect, par exemple adjectif ou verbe. Environ 300 000 mots ont été étiquetés de cette manière.

Une fois le travail d'étiquetage terminé, le CECL a extrait des statistiques de ce corpus afin d'en déterminer les erreurs les plus fréquentes. Des analyses détaillées de certaines catégories d'erreurs ont fourni des informations sur les contextes spécifiques où les erreurs avaient le plus de risques de se produire. Ces deux types d'informations ont été utilisés pour créer des exercices ciblés sur des domaines problématiques pour les apprenants et pour déterminer sur quelles catégories le système de diagnostic devait se concentrer.

4. Diagnostic d'erreurs

Cette section présente le système de diagnostic d'erreurs qui est actuellement implémenté dans le projet FreeText. Dans ce logiciel, les utilisateurs peuvent utiliser l'outil de diagnostic tant pour corriger des exercices spécifiques, où ils doivent écrire une production relativement libre, comme une phrase ou un court paragraphe, que pour une utilisation indépendante, sans tâche particulière, afin de contrôler la grammaticalité de leurs propres productions. Le diagnostic se passe en trois étapes : (i) correction orthographique ; (ii) diagnostic d'erreurs syntaxiques ; (iii) comparaison de phrases. A chaque étape, des erreurs peuvent être détectées et affichées pour les utilisateurs à qui l'on demande ensuite de corriger leur production avant de passer à l'étape suivante. Ce processus assure qu'il n'y a plus d'erreur orthographique lors du diagnostic syntaxique et, ensuite, qu'aucune erreur orthographique ou de grammaire n'a été détectée avant de commencer la comparaison de phrases.

Dans cet article, nous écartons la question de la vérification orthographique pour nous concentrer sur le diagnostic syntaxique et le système de comparaison de phrases, puis nous parlerons de l'interface du système automatique de diagnostic d'erreurs.

4.1. Diagnostic syntaxique

Le système de diagnostic est basé sur l'analyseur Fips, qui est brièvement décrit ci-dessous afin de faciliter la compréhension des deux techniques de diagnostic utilisées pour les erreurs syntaxiques, à savoir le relâchement de contraintes et la réinterprétation phonologique. Ces deux techniques sont décrites plus largement par la suite.

4.1.1. Les catégories d'erreurs

Dans [le tableau 1](#), nous décrivons les catégories d'erreurs qui peuvent être traitées par le composant syntaxique du système de diagnostic d'erreurs, illustrées par un exemple. Les erreurs sont marquées en gras ou par un soulignement.

Catégorie d'erreur	Acronyme	Exemple
Classe	CLA	Ce montre le problème.
Auxiliaire	AUX	Il a tombé.
Genre	GEN	Elle est beau.
Nombre	NBR	Les souris ont été mangée.
Personne	PER	Je peut partir.
Voix	VOI	Il _ révolte contre la société.
Euphonie – élision	EUFE	ce aspect
Euphonie – contraction	EUFC	de le chat
Euphonie – t analogique	EUFT	A- il mangé?
Complémentation de l'adjectif	CPA	Il est capable à faire cela.
Complémentation du verbe	CPV	Il est tombé _ la chaise.
Ordre	ORD	Ils ont les vus.
Ordre de d'adverbe	ORDAV	C'est lui plutôt qui chasse
Ordre de l'adjectif	ORDAJ	un arabe pays.
Redondance	RED	Il parle beaucoup des langues.
Oubli de ponctuation	OUB	Peut _ il partir?
Négation	NEG	C' _ est pas un problème.
Homophonie	HOM	Elles peuvent maître au monde des enfants.

Tableau 1 - Catégories d'erreur traitées par le système de diagnostic.

Toutes les erreurs, sauf celles d'homophonie, sont détectées par relâchement de contraintes, qui est activé en premier. La réinterprétation phonologique, qui n'est lancée qu'après le relâchement de contraintes, traite des erreurs d'homophonie. [2] La plupart de ces catégories d'erreur sont calquées sur la typologie de FRIDA, à part quelques exceptions, lorsqu'il a semblé possible d'affiner ces catégories, comme les erreurs d'euphonie ou d'ordre, compte tenu des particularités du système.

Les analyses détaillées du CECL sur le corpus ont montré, entre autres, que les erreurs d'ordre sont nettement plus fréquentes pour les adverbes et les adjectifs que pour les autres catégories lexicales. L'analyse fine menée sur l'ordre de l'adverbe, détaillant les classes d'adverbes et leurs positions légales et illégales, a été essentielle pour la mise en place du relâchement de la contrainte de la position de l'adverbe. Seuls les attachements erronés dans des positions attestées dans le corpus ont été autorisés, au lieu de permettre un attachement sauvage des adverbes dans n'importe quelle position, ce qui aurait eu comme conséquence une surgénération de structures gigantesque et ingérable.

4.1.2. L'analyseur syntaxique Fips

Fips, décrit en détail chez [Wehrli97], est un analyseur syntaxique basé sur la théorie des Principes et Paramètres [ChomskyLasnik95]. Cet analyseur fournit des analyses très riches et motivées linguistiquement. Par exemple, les mouvements des constituants de la structure profonde à la structure de surface sont repérés par l'insertion de traces ; les liens entre pronoms relatifs et antécédents sont également marqués.

L'algorithme d'analyse est celui d'attachement au coin droit et les alternatives sont traitées de manière pseudo-parallèle. Les constituants bien formés sont placés dans un graphe. Des scores sont attribués à chaque structure afin d'ordonner les alternatives et de choisir quelle analyse afficher lorsque l'analyse a réussi. Lorsque aucune analyse complète n'a été trouvée, l'analyseur présente néanmoins des analyses partielles de l'entrée, appelées aussi morceaux, tirées du graphe des constituants bien formés. Ainsi, l'analyseur produit une sortie même lorsque l'analyse de la phrase d'entrée échoue. La présence d'analyses partielles indique soit que l'entrée n'est pas grammaticale, soit que sa structure dépasse la couverture de l'analyseur.

Le processus d'analyse a lieu en trois étapes principales : (i) projection de structures sur la base des éléments lexicaux, d'après le schéma X-barre ; (ii) combinaison de ces structures ; (iii) application de contraintes pour rejeter les structures agrammaticales.

Développé depuis plus de dix ans à l'Université de Genève, Fips est un analyseur modulaire à large couverture qui est facilement adaptable pour des tâches particulières. Ainsi, les modules nécessaires au diagnostic d'erreurs ont pu être développés sur la base d'un moteur d'analyse pré-existant.

4.1.3. Relâchement de contraintes

Le relâchement de contraintes est une technique de diagnostic d'erreurs largement utilisée, qui connaît plusieurs variantes. Mentionnons notamment Weischedel et Black ([\[WeischedelBlack80\]](#) : 108) qui parle de "failable constraints", Matthews [\[Matthews93\]](#) de violation de principes, et Heinecke et al. [\[Heinecke88\]](#) de "graded constraints". Bien qu'il y ait des différences entre ces techniques, les principes sous-jacents sont les mêmes : normalement, on ne peut combiner des structures partielles que si certaines contraintes ou conditions sont remplies, comme l'accord sujet-verbe en nombre et en personne ; lorsque les contraintes sont relâchées, on permet l'attachement même si la contrainte n'est pas respectée. Pour utiliser cette technique dans un système de diagnostic, chaque contrainte relâchée doit être marquée dans la structure, afin que le type et le lieu de l'erreur détectée puissent être indiqués par la suite. Le marquage approprié de l'endroit où l'analyse a pu se poursuivre, grâce au relâchement spécifique d'une contrainte, constitue la base du diagnostic d'erreurs [\[Felshin95\]](#). La plupart des systèmes utilisant cette technique peuvent relâcher les contraintes une à une et ne le font que lorsque leurs conditions ne sont pas satisfaites. Un des avantages du relâchement de contraintes est qu'il fournit à la fois un diagnostic et une analyse complète de la phrase d'entrée.

Illustrons cette technique avec la phrase d'exemple en (1), tirée du corpus FRIDA, qui contient une erreur de genre. Les erreurs de genre sur les adjectifs arrivent en sixième position et représentent 2,81 % des erreurs du corpus.

(1)*L'héritage du passé est très forte et le sexisme est toujours présent.

Pour que le syntagme adjectival (AP pour Adjectival Phrase) très forte s'attache au syntagme verbal comme attribut, il doit s'accorder en genre et en nombre avec le sujet de la phrase (TP pour Tense Phrase), à savoir le syntagme nominal (DP pour Determiner Phrase) l'héritage du passé. La vérification de l'accord en genre se fait en prenant l'intersection des ensembles des genres des deux constituants. Si le résultat est un ensemble vide, cela signifie que les deux éléments ne s'accordent pas, que la phrase est agrammaticale et qu'elle doit être rejetée. Cependant, lorsque le relâchement de contraintes est activé, si l'intersection de deux ensembles est un ensemble vide, cette contrainte spécifique est relâchée et l'on procède à la place à l'union des ensembles, permettant ainsi aux deux constituants de s'attacher malgré l'erreur détectée. Le lieu où la contrainte a été relâchée est marqué dans la structure afin de pouvoir être affiché pour les utilisateurs par la suite. Ce processus est schématisé dans [la figure 1](#).

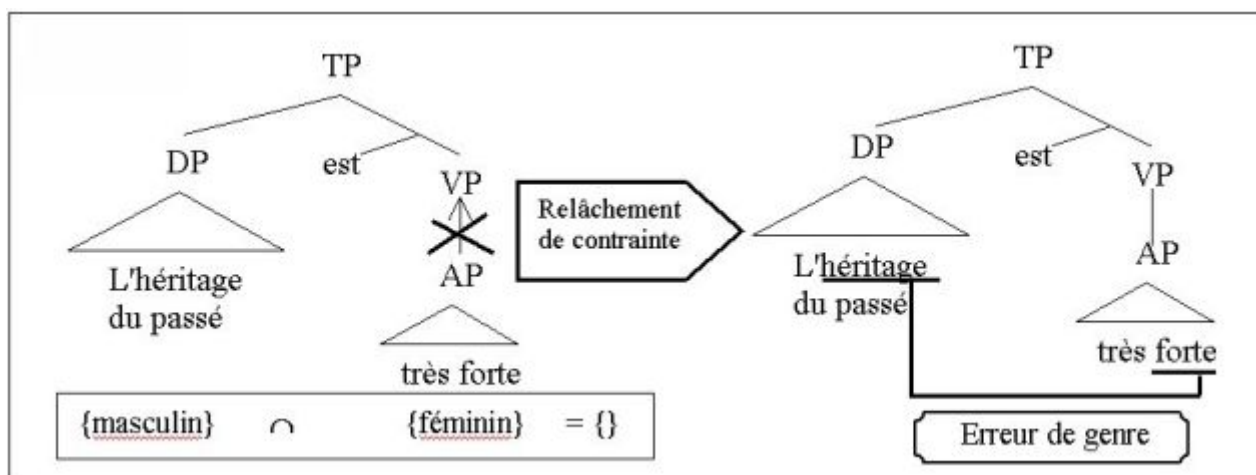


Figure 1 - Schéma du relâchement d'une contrainte en genre (structures simplifiées).

En pratique, cependant, l'utilisation du relâchement de contraintes aboutit à une surgénération massive d'analyses, ralentissant le système et diminuant ses performances en présentant des analyses (et donc des marques d'erreurs) inadéquates. Le système de score existant dans Fips pour classer les analyses est fortement perturbé par le relâchement de contraintes. Pour résoudre ce problème et identifier quelle analyse correspond le mieux à l'intention de l'apprenant, trois améliorations ont été apportées au système de classification existant : (i) un ensemble de règles heuristiques pour éliminer les analyses hautement improbables ; (ii) des mesures de classement basées sur la fréquence des mots ; (iii) des mesures de classement basées sur la fréquence de types d'erreurs. Un algorithme a été développé pour identifier l'analyse préférentielle sur la base de ces valeurs. [3]

4.1.4. Réinterprétation phonologique

La réinterprétation phonologique est basée sur l'idée que certains mots ont des homophones et que les apprenants d'une langue peuvent confondre des mots avec une prononciation similaire. Remplacer des mots par leurs homophones peut permettre au système d'obtenir une analyse complète pour une phrase donnée et en même temps d'obtenir un diagnostic d'erreurs – avec cependant la nuance que seuls les emplois erronés d'homophones de différentes catégories lexicales ou ayant des traits syntaxiques suffisamment distincts peuvent être détectés, car, comme le constate Vosse ([Vosse92] : 112), "unfortunately, a parser cannot detect substitutions by homophones which have the same syntactic properties".

La réinterprétation phonologique procède d'après l'algorithme suivant :

1. Un premier traitement de la phrase n'aboutit qu'à des analyses partielles ;
2. Les mots aux limites des analyses partielles sont récupérés ;
3. Pour chacun des mots récupérés :
4. On recherche sa représentation phonologique dans le lexique ;
5. Tous les mots avec la même représentation phonologique sont recherchés dans le lexique et insérés dans le graphe de l'analyseur ;
6. La phrase est analysée à nouveau avec tous les nouveaux mots alternatifs ;
7. Si une nouvelle analyse complète est trouvée, les mots réinterprétés que l'on a trouvés peuvent être proposés comme une correction possible.

Cette technique de diagnostic d'erreurs nécessite que le lexique utilisé par l'analyseur contienne la représentation phonologique de tous les mots et qu'on puisse faire une recherche sur cette clé. De plus, l'analyseur doit être non déterministe et accepter les alternatives de mots.

La phrase donnée en (2a) est une erreur authentique mais simplifiée, tirée du corpus FRIDA :

(2)a.*Il faut maître les enfants au monde.b. Il faut mettre les enfants au monde.c. [TP Il faut maître] [DP les enfants au monde]

L'analyse de la phrase incorrecte en (2a), et corrigée en (2b), donne les deux analyses partielles en (2c). Les deux mots en frontière de morceau d'analyse, maître et les, sont réinterprétés et donnent les mots mettre, mètre, mètres et maîtres pour maître et aucun pour les. Mètre en tant que substantif sera exclu (identité de catégorie lexicale et de traits syntaxiques), tandis que mètres et mètre comme verbes (indicatif présent 2^e et 3^e personne du singulier) et les substantifs pluriels maîtres et mètres ne permettront aucune combinaison grammaticale. Les est aussi un pronom clitique, mais a déjà été considéré comme constituant un seul élément dans le graphe de l'analyseur, puisque toutes les lectures possibles d'un mot sont gardées dans la liste des constituants et peuvent se combiner pour en constituer un plus grand.

Ainsi, une seconde analyse de la phrase, incluant les mots réinterprétés, peut donner une analyse complète de la phrase en utilisant mettre à la place de maître. L'erreur d'homophonie est détectée et une correction peut être proposée.

Pour conclure cette partie, le composant syntaxique du système de diagnostic connaît néanmoins de sérieuses limites. Il traite chaque phrase indépendamment les unes des autres. Le niveau sémantique n'est absolument pas considéré. La résolution de l'anaphore, source de nombreuses erreurs d'accords ne peut notamment pas être traitée. Néanmoins, les catégories d'erreurs couvertes fournissent un apport non négligeable à l'apprentissage.

4.2. Diagnostic de comparaison de phrases

Dans le domaine du diagnostic d'erreur, une vérification syntaxique seule s'avère parfois insuffisante pour fournir une rétroaction correcte à une question ouverte. Par exemple, les apprenants du FLE et même les locuteurs natifs confondent souvent le déterminant démonstratif pluriel ces avec le déterminant possessif pluriel ses, qui sont des homophones. Considérons un exercice de compréhension de texte où les apprenants doivent répondre à la question en (3a), avec la réponse attendue en (3b). Si un apprenant entre la phrase en (3c), aucune erreur syntaxique ne sera détectée, car la phrase est grammaticale et correcte en isolation :

(3)a. Qu'est-ce que Paul a fait ?b. Il a lu ses livres.c. Il a lu ces livres.

Nous pouvons aussi imaginer des exercices de pronominalisation où la réponse attendue est (4a) et la réponse de l'apprenant est (4b). Cette réponse est grammaticalement correcte, puisque la règle de l'accord du participe passé avec un pronom accusatif antéposé est respectée, mais elle est incorrecte pour l'exercice en question.

(4)a. Je les ai vues.b. Je les ai vus.

Un diagnostic d'erreurs syntaxique est donc insuffisant et d'autres techniques doivent être implémentées. Dans la plupart des systèmes d'ELAO, la technique de "mise en correspondance de formes" (pattern matching) est utilisée pour des réponses plus ou moins ouvertes. Les tuteurs doivent entrer une réponse modèle avec une formule complexe et doivent prendre en compte une grande variété de réponses possibles. Cette tâche peut être fastidieuse et des réponses possibles et correctes risquent de ne pas être traitées. De même, une mauvaise formule peut accepter des réponses tout à fait incorrectes. Ces raisons font qu'il est difficile d'envisager des exercices avec des phrases complètes lorsqu'un logiciel ne dispose que de cette technique de mise en correspondance.

Dans le projet FreeText, nous avons tenté d'implémenter une autre technique appelée "diagnostic de comparaison de phrases". Les tuteurs ne doivent entrer qu'une seule réponse modèle dans la base de données des exercices. Ensuite, le système extrait une structure "sémantique" profonde de la phrase modèle et de celle de l'apprenant. Après les avoir comparées l'une à l'autre, le système renvoie une rétroaction.

Cette technique pourrait être utilisée avec les exercices de transformations actif/passif ou déclaratif/interrogatif, de topicalisation, pronominalisation, transformation phrase nominale/phrase complète ou infinitif/impératif, ainsi que pour des exercices sur l'accord.

Les structures utilisées pour comparer les phrases sont appelées "structures pseudo-sémantiques" (pseudo-semantic structures, PSS). Elles sont extraites directement des structures de sortie de Fips et servent de structures d'entrée au générateur GBGen [EtchegoyhenWehrle98]. Ce système est principalement utilisé pour la traduction automatique [Lhaire00]. Les PSS sont des structures hybrides qui combinent à la fois des informations abstraites et des éléments lexicaux.

Les PSS sont divisées en trois sous-types : (i) les CLS (clause structures) qui correspondent aux états et aux événements et contiennent des verbes ou des adjectifs ; (ii) les DPS (DP-Structures) qui correspondent aux structures nominales et contiennent des noms ; (iii) les CHS (Characteristic structures) qui correspondent aux modificateurs de syntagmes et contiennent des adjectifs ou adverbes. Les PSS peuvent se combiner entre elles pour former des structures hiérarchiques ; elles contiennent une liste non ordonnée d'autres PSS, appelées satellites, qui représentent les arguments des verbes, les ajouts, les modificateurs de syntagme, etc.

La figure 2 ci-dessous montre un schéma simplifié des PSS des deux phrases en (5) et (6).

(5) C'est Jean qui a mangé cette pomme rouge.

(6) Une pomme rouge a été mangée par Jean.

Une structure de PSS semblable peut être dérivée de ces deux phrases (figure 2). Les champs qui divergent sont désignés en gras. "Delta" signifie qu'une propriété est indéfinie. Les flèches dénotent la relation "est-le-satellite-de".

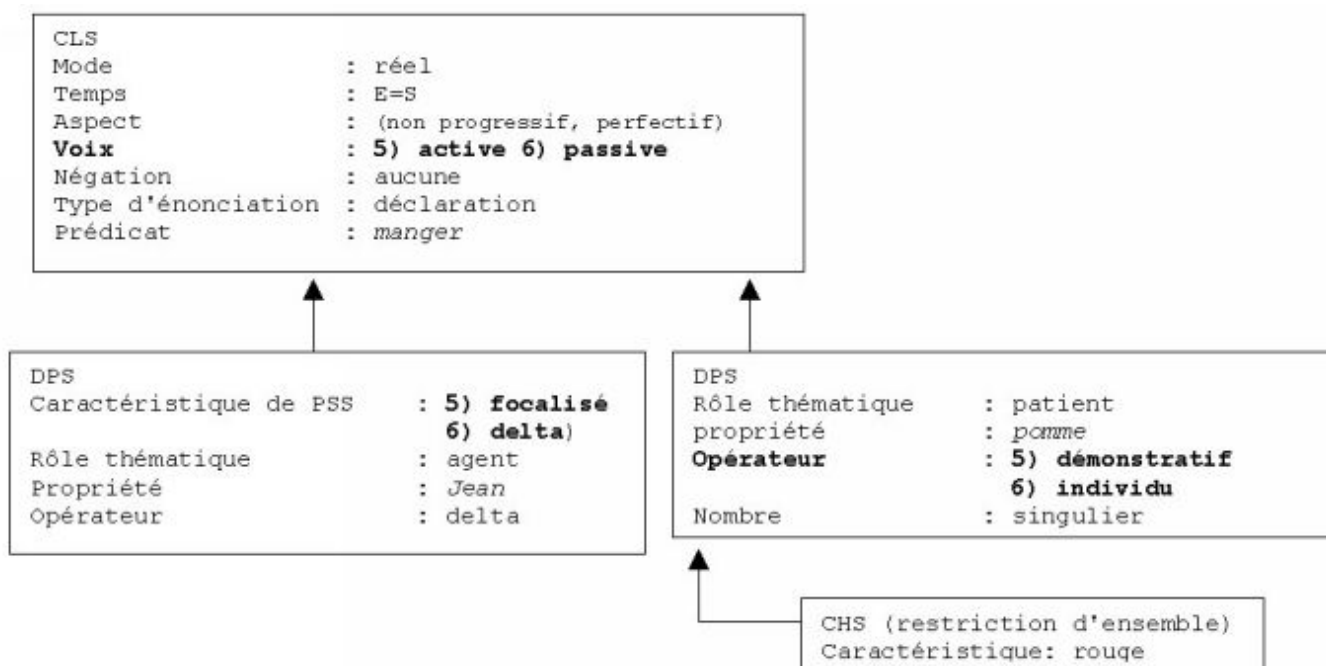


Figure 2 - Schéma simplifié de PSS pour les exemples (5) et (6).

Nous pouvons constater que seuls quelques paramètres varient entre les deux phrases, alors qu'elles n'ont pas la même structure syntaxique. Les champs modifiés sont "voix", "caractéristique de PSS" et "opérateur". Pour une phrase interrogative, le champ "type d'énonciation" de la CLS serait modifié, pour une négative le champ "négation", etc. [Le tableau 2](#) montre quelques constructions qui peuvent être traitées par le système et la rétroaction qui doit être fournie.

Construction	Réponse modèle	Phrase de l'apprenant	Rétroaction
Usage incorrect du pronom clitique	Je les ai vues.	Je les ai vus.	Erreur de genre du pronom
Erreur de signification	La grosse souris est grise.	La grosse souris est blanche.	Erreur de signification sur l'adjectif
Question globale (oui/non)	Jean a-t-il lu ce livre?	Est-ce que Jean a lu ce livre?	Pas de rétroaction: construction équivalente
Questions partielles	Quel livre est-ce que Jean lit ?	Quel livre Jean lit-il ?	Pas de rétroaction: construction équivalente
Erreur de voix	Le chat poursuit la souris.	La souris est poursuivie par le chat.	Erreur de voix sur le verbe
Relatives	La sœur du capitaine qui a été tuée.	La sœur du capitaine qui a été tué.	Erreur d'accord pronom relatif-verbe
Clivée	C'est Pierre que Jean a vu.	Jean a vu Pierre.	Erreur de clivage
Topicalisation	Avec ses jumelles, Jean a observé cet homme.	Jean a observé cet homme avec ses jumelles.	Erreur de topicalisation
Adjonction	Jean a regardé l'homme avec ses jumelles.	Jean a regardé l'homme.	Constituant manquant
Verbes à montée	Jean semble dormir	Il semble que Jean dort.	Pas de rétroaction: construction équivalente
Pronominalisation	Le chien est blanc et noir.	Il est blanc et noir.	Avertissement sur le pronom
Ellipse	Le chien est blanc et noir.	Blanc et noir.	Avertissement: verbe manquant

Tableau 2 - Constructions traitées par l'outil de comparaison de phrases.

L'on remarquera que, dans le cas des verbes à montée et des questions ouvertes et fermées, des PSS identiques sont extraites à partir de structures syntaxiques différentes. Il pourrait cependant arriver que des enseignants veuillent que les apprenants utilisent une construction et non une autre. Dans le cas des phrases interrogatives, la structure en "est-ce que" sera préférée par les apprenants à la structure d'inversion sujet-verbe, qui est plus difficile à produire. Nous prévoyons d'introduire des paramètres pour permettre de tolérer ou non des constructions équivalentes, les ellipses, etc.

Ainsi, les PSS sont particulièrement bien appropriées pour comparer des phrases et procéder à une analyse un peu plus profonde. Contrairement aux structures syntaxiques produites par l'analyseur, ces structures sont relativement simples et peuvent être facilement comparées entre elles. Les tuteurs ne doivent entrer qu'une seule réponse modèle, la plus complète possible, et le système pourra traiter et comparer des constructions des types mentionnés ci-dessus.

Cependant, le système de comparaison de phrases n'a pas de réel composant sémantique. Les apprenants peuvent répondre à des questions ouvertes de manière très variée, et le système atteindra rapidement ses limites. Considérons la question (7a) et supposons qu'elle soit illustrée dans le logiciel par une image. L'enseignant entre comme modèle la phrase (7b) et l'apprenant répond (7c).

(7)a. De quelle couleur est le chien ? b. Le chien est blanc et noir.c. Le caniche est noir et blanc.

Nous admettrons que l'apprenant connaisse le mot désignant ce chien particulier et que l'image représente effectivement cette race. Le système détectera une erreur de signification, car chien est différent de caniche. Pour améliorer les performances du système, nous avons imaginé utiliser des dictionnaires sémantiques contenant des hyponymes, synonymes et hyperonymes. Cependant, dans le cadre du projet, nous n'avons pas pu tester cette technique, par manque de temps et faute d'un dictionnaire sémantique étendu. En outre, l'exactitude de la réponse ne pourra être assurée, car les apprenants pourraient par exemple utiliser la mauvaise race de chien pour désigner l'image (hyponyme) ou utiliser l'hyperonyme "animal".

Cependant, dans l'ensemble, cette technique innovante semble suffisamment intéressante et fiable pour être implémentée et utilisée dans un logiciel d'ELAO : elle pourrait améliorer la qualité de la rétroaction et permettre ainsi aux enseignants de créer facilement des nouveaux types d'exercices plus intelligents, vu que le diagnostic d'erreur automatique offre une rétroaction plus précise. Cette technique, sans offrir un diagnostic véritablement sémantique, permet de franchir la barrière de la syntaxe et d'aller un peu plus loin. Néanmoins, la comparaison de phrases n'a pas été activée dans le système de diagnostic d'erreurs et est considérée comme résultat secondaire du projet FreeText.

4.3. Interface du système

Dans une application d'ELAO, une bonne rétroaction est cruciale car les apprenants sont souvent seuls avec une machine et ne peuvent s'appuyer sur des enseignants et tuteurs pour les aider. Dans cette sous-section, nous parlons de la sortie du système et de la manière dont nous l'utilisons pour la rétroaction.

4.3.1. Sortie XML

Le système de diagnostic produit un document XML pour chaque analyse. Le formalisme XML, largement répandu et évolutif, peut être facilement adapté et étendu. Considérons [la figure 3](#), un extrait de la sortie pour le chats a dormi, qui contient une erreur d'accord :

```
<PROJ cat="DP" attachedAs="spec" colorGramm="sujet" tree="GN">
<HEAD cat="D" tool="yes" gender="masc" number="sin" lexeme="le" ortho="le"
colorGrammLex="Det">
<ERROR index="Aa01" manypart="yes" category="NBR">le</ERROR>
</HEAD>
<PROJ cat="NP" attachedAs="comp" tree="hidden">
<HEAD cat="N" tool="no" gender="masc" number="plu" lexeme="chat" ortho="chats"
colorGrammLex="Nom">
<PARTERROR antecedent="Aa01">chats</PARTERROR>
</HEAD>
</PROJ>
</PROJ>
```

Figure 3 - Extrait du diagnostic pour "le chats a dormi".

Les balises PROJ marquent les projections syntaxiques et contiennent des informations syntaxiques : la projection au sommet a la fonction grammaticale de sujet (attribut colorGramm) et sera analysée comme groupe nominal (attribut tree). Les balises HEAD marquent les têtes de projection et contiennent des informations sur les mots, comme le genre, le nombre, le temps, la catégorie lexicale ou le lexème (forme de base d'un mot, pour un substantif la forme au singulier, pour un verbe l'infinitif, etc.). Les erreurs sont marquées avec les balises ERROR et parfois PARTERROR si l'erreur porte sur plusieurs mots. Ici, nous avons une erreur de nombre (marquée par l'attribut category) entre le et chats. Ces deux balises sont co-indicées, car plusieurs erreurs peuvent être détectées dans une seule phrase.

4.3.2. Rétroaction intelligente

Bien entendu, la sortie XML du traitement de diagnostic ne peut être utilisée telle quelle. Une sortie XML peut être analysée par de nombreux outils et langages de programmation disponibles sur la Toile. Dans les figures 4 à 6 ci-dessous, nous montrons trois interprétations différentes de la sortie illustrée dans la figure 3 ci-dessus, à savoir une grammaire en couleur et le diagnostic d'erreurs ainsi qu'une représentation arborescente.

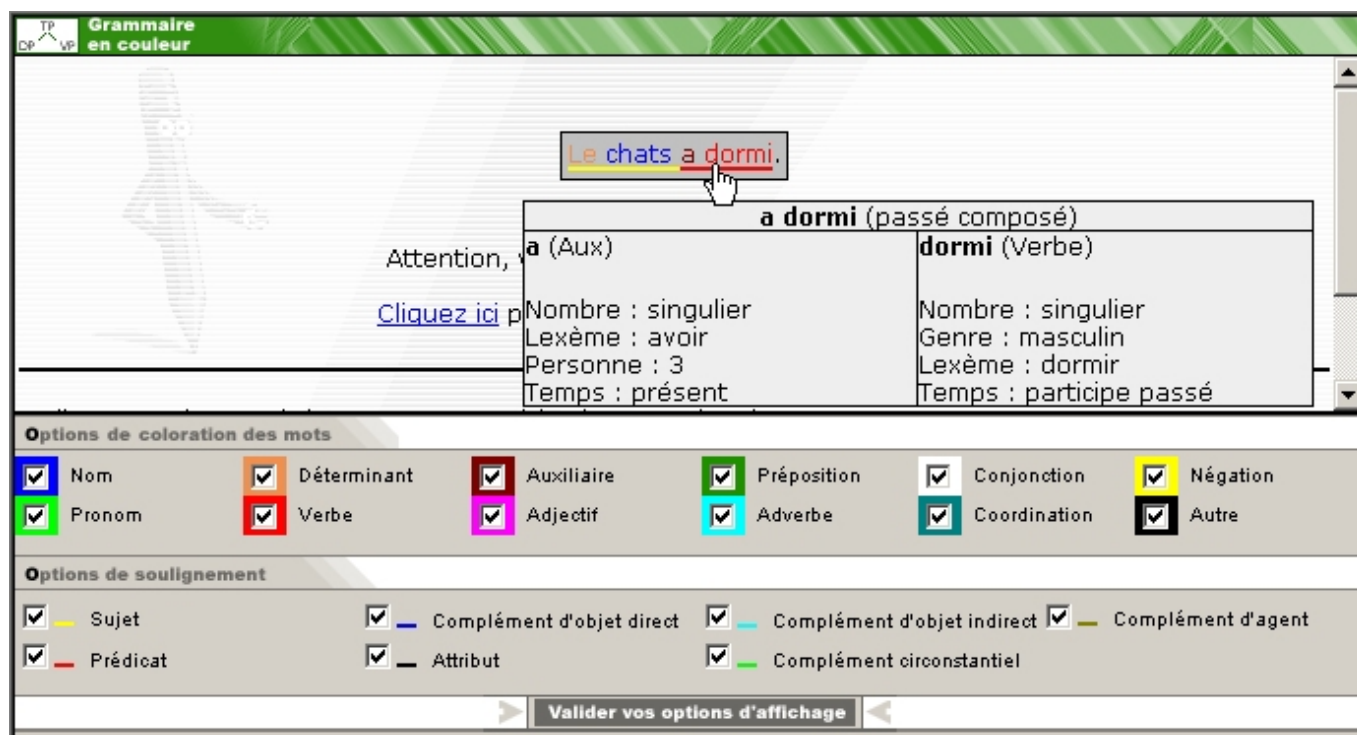


Figure 4 - Grammaire en couleur.

La figure 4 montre une grammaire en couleur. Les fonctions syntaxiques sont indiquées par des soulignements en couleur, les catégories lexicales par différentes couleurs de police. Si le curseur de la souris passe sur un mot, toutes les informations lexicales associées sont affichées. Dans cet exemple, le temps du verbe est décrit comme passé composé, et ses deux éléments sont affichés et analysés séparément. Comme la phrase contient une erreur, un avertissement invite l'apprenant à activer le diagnostic d'erreurs en cliquant sur un lien hypertexte. Il peut également accéder par un lien à l'outil de conjugaison fourni par le système.

le chats a dormi.

- **Erreur** sur le mot **le** avec le(s) mot(s) **chats** (Erreur de nombre).

Figure 5 - Système de diagnostic d'erreurs.

La figure 5 montre comment est affiché le diagnostic d'erreurs. Si les apprenants passent sur le mot Erreur avec le curseur de la souris, les deux parties de l'erreur sont mises en évidence en vert. S'ils passent sur les mots de la phrase, la même information lexicale que dans la grammaire en couleur est affichée. Si le système n'a pas trouvé d'erreur, un message annonce aux utilisateurs qu'aucune erreur n'a été détectée. Un fichier d'aide qui détaille quelles erreurs sont bien détectées, partiellement détectées et non détectées doit être disponible, afin que les apprenants et les enseignants soient conscients des limites du système.

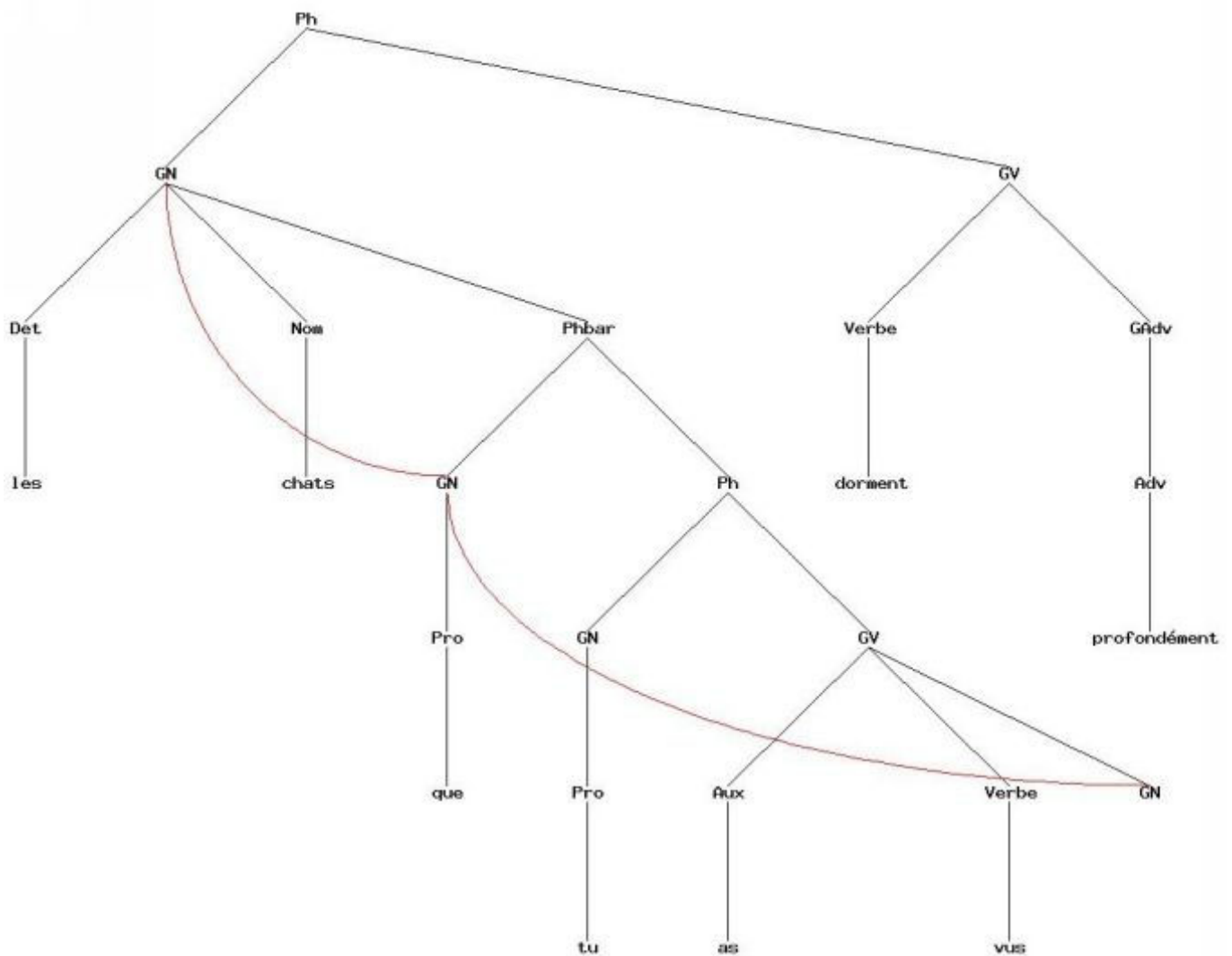


Figure 6 - Arbre syntaxique pour la phrase "les chats que tu as vus dorment bien".

[La figure 6](#) montre une structure d'arbre syntaxique que les apprenants ont la possibilité d'afficher. Ph indique la phrase, GN le groupe nominal, GV le groupe verbal et Phbar la phrase subordonnée, tandis que des étiquettes lexicales marquent les mots. Des traces relient également différents éléments de la phrase entre eux afin de montrer les relations syntaxiques. Ici, les arcs indiquent que le pronom relatif que reprend le groupe nominal les chats et qu'il peut être attaché au groupe nominal objet dans le groupe verbal de la phrase subordonnée. Les traces sont aussi affichées dans le cadre des pronoms clitiques, du mouvement du passif, etc.

Ainsi, les trois sorties présentées ici fournissent à l'apprenant une rétroaction de sa production suffisamment complète, fiable et immédiate, donc utile.

5. Tests et comparaison du système

Le système de diagnostic n'a malheureusement pas pu être testé directement par des apprenants dans sa version finale. Toutefois, nous avons pu mener une série de tests de la partie syntaxique du système de diagnostic sur la base d'une étape de validation antérieure durant laquelle des apprenants avaient répondu à certains exercices de production libre dans le logiciel. Le corpus contient 120 phrases et 1574 mots, soit environ 13 mots par phrase en moyenne. Seules les fautes d'orthographe ont été corrigées avant de soumettre ce corpus authentique au composant syntaxique du système de diagnostic. Le même corpus a été soumis au correcteur de grammaire de Microsoft Word 2000® (ci-après Word), à des fins de comparaison. [Le tableau 3](#) indique le nombre d'erreurs dans le corpus pour les catégories d'erreurs pouvant être détectées par les deux systèmes, ainsi que le nombre d'erreurs correctement détectées par chaque système.

Erreurs	Corpus	FreeText	Word
GEN	28	26	22
HOM	12	7	2
OUB	10	0	4
PER	7	7	6
NBR	7	6	4
EUF	5	5	4
AUX	4	2	3
NEG	1	1	1
Total	74	54	46

Tableau 3 - Erreurs correctement détectées.

Nous constatons tout d'abord que le système de diagnostic implémenté dans le cadre du projet FreeText détecte plus d'erreurs que son concurrent. Seules les catégories d'oubli du tiret (OUB) et du choix de l'auxiliaire (AUX) sont mieux traitées par Word.

Nous pourrions nous étonner qu'aucune des dix erreurs de la catégorie OUB ne soit détectée par notre système. Un examen du corpus nous a montré que dans le cas présent ces oublis avaient lieu entre nom et adverbe (jour-là), et entre deux parties du mot composé (grand-mère). Or, uniquement les erreurs de manque du tiret entre un verbe et le pronom conjoint le suivant (pleut-il) avaient été prévues dans notre système de diagnostic. Nous avons également constaté que trois oublis sur le mot composé grand-mère étaient considérés comme des erreurs de genre par notre système.^[4]

Outre les catégories d'erreurs communes entre FreeText et Word, notre système détecte également d'autres catégories d'erreurs (voir [tableau 1](#)), y compris les erreurs de complémentation verbale (CPV), un sujet particulièrement délicat pour les apprenants et qui n'est pas traité par les correcteurs pour locuteurs natifs. La phrase en (8) est tirée du corpus.

(8) Elle s'est mis à téléphoner les hôpitaux.

Le système FreeText détecte à la fois l'erreur de genre du participe passé et l'erreur de complémentation verbale entre téléphoner et les hôpitaux, alors que Word n'indique que l'erreur de genre.

Pour être complet, il nous faut également mentionner le phénomène de la sur-détection d'erreurs. Dans certains cas, notre système, et dans une moindre mesure Word, indique des erreurs là où il n'y en a pas. Cela est dû généralement à une mauvaise analyse de la phrase qui peut être trop complexe et/ou contenir d'autres erreurs. Un exemple typique est le manque d'accent sur un participe passé, comme en (9).

(9) Quand j'étais petite j'ai beaucoup aime la crème glacée.

La seule erreur de cette phrase tirée du corpus est le manque d'accent sur aime.^[5] Perturbé par cette erreur qu'il n'est pas équipé pour traiter, le système FreeText détecte, de manière erronée, trois erreurs en (9). Une première erreur d'homophonie sur ai, pris comme et, une deuxième erreur d'homophonie sur glacée dont la version masculine glacé est retenue, et du coup une erreur de genre entre crème et glacée. Nous pouvons donc voir à quel point une erreur non traitée peut faire échouer toute l'analyse de la phrase. De son côté, Word ne détecte aucune erreur. Pour l'exemple (10) où les apostrophes sont manquantes, à la fois notre système et Word sont mis en difficulté.

(10) Quand j'étais petite j'aimais nager.

FreeText prend les deux j comme étant la lettre j et trouve des erreurs d'accord en personne avec les verbes. Word, par ailleurs, considère étais comme un nom et détecte une erreur d'accord en nombre cette fois avec l'adjectif petite qu'il prend pour un attribut du nom étais.

Les résultats du composant syntaxique du système de diagnostic de FreeText sont bons, pris en isolation ou comparé avec ceux de Word. FreeText pêche toutefois encore par un trop grand nombre de sur-détection d'erreurs, certaines, mais pas toutes, étant liées à des erreurs de catégories non traitées par le système. De nouveaux tests sur un corpus plus large seraient nécessaires pour le confirmer, mais il semble, a priori, qu'il pourrait être intéressant de désactiver certains relâchements de contraintes qui donnent lieu à une trop grande proportion de sur-détection par rapport aux erreurs réelles détectées. C'est le cas, en particulier, pour les erreurs de complémentation adjectivale : il n'y en a aucune dans le corpus, mais deux erreurs de cette catégorie ont été sur-détectées.

6. Avantages attendus et perspectives de recherche

Dans une phase antérieure du projet, un premier test du système a été mené avec 23 enseignants du FLE. A ce moment-là, le système de diagnostic d'erreurs ne détectait que quelques erreurs d'accord, et la grammaire en couleur n'était pas non plus à un stade avancé. Les enseignants se sont montrés intéressés par l'accès à des outils de TALN qui fournissent un bon diagnostic, mais le prototype n'était pas suffisamment bon. La large couverture de la version actuelle du système de diagnostic et la qualité de la grammaire en couleur devraient grandement améliorer ces résultats. Notre système de détection d'erreurs devrait permettre aux enseignants de construire rapidement et facilement des exercices qui sollicitent des réponses d'apprenants dans un contexte plus libre.

Du point de vue d'un apprenant, la possibilité d'obtenir une rétroaction quasi-immédiate est une caractéristique intéressante. Les apprenants ne devront pas attendre un tuteur dans une salle de classe, ni la correction papier de leurs devoirs, ni encore la réponse par courriel d'un tuteur à distance.

Au-delà de la grammaire et de la sortie actuelle du système de diagnostic, il devrait être possible de fournir une rétroaction plus spécifique aux apprenants, grâce à la richesse de la sortie XML. Nous prévoyons des liens directs vers des sections spécifiques d'une grammaire de référence et vers des exercices de remédiation liés à des erreurs particulières. L'intégration de dictionnaires sémantiques pourrait aussi améliorer le système de diagnostic d'erreurs. Bien que le système ne fournisse pas un diagnostic d'erreurs parfait, les résultats obtenus jusqu'ici sont encourageants et semblent être suffisamment bons pour être utilisés dans une application réelle.

Remerciements

Les auteurs tiennent à remercier le Dr. Catherine Walther Green pour sa lecture attentive de précédentes versions de cet article et pour ses commentaires très utiles. Nos remerciements vont aussi aux relecteurs pour leurs suggestions et corrections.

Références

Bibliographie

[ChomskyLasnik95]

Chomsky, N. & Lasnik, H. (1995). "The theory of Principles and Parameters". In Chomsky, N. (dir.). *The Minimalist Program*. Cambridge, Mass.: MIT Press. pp. 13-127.

[EtchegoyhenWehrle98]

Etchegoyhen, T. & Wehrle, T. (1998). "Overview of GBGen: A large-scale, domain independent syntactic generator". In *Proceedings of the 9th International Workshop on Natural Language Generation*. Niagara-on-the-Lake : Association for Computational Linguistics. pp. 288-291.

[Felshin95]

Felshin, S. (1995). "The Athena Language Learning Project NLP System: a multilingual system for conversation-based language learning". In Holland, V. M., Kaplan, J.D. & Sams, M.R. (dir.). *Intelligent Language Tutors: Theory Shaping Technology*. Hove, Royaume Uni : Erlbaum. pp. 257-272.

[Granger01]

Granger, S., Vandeventer, A. & Hamel, M.-J. (2001). "Analyse des corpus d'apprenants pour l'ELAO basé sur le TAL". *Linguistique de corpus*. Numéro spécial de *Traitement automatique des langues*, vol. 42, n°2. pp. 609-621.

[Granger03]

Granger, S. (2003). "Error-tagged learner corpora and CALL: a promising synergy". In Heift, T. & Schulze, M. (dir.). *Error Analysis and Error Correction in Computer-Assisted Language Learning*. Numéro spécial de *CALICO*, vol.20, n°3. pp. 465-480.

[Heinecke88]

Heinecke, J., Kunze J., Menzel, W. & Schröder, I. (1998). "Eliminative parsing with graded constraints". In *Proceedings of Coling-ACL'98*, vol. 1. pp. 526-530.

[Lhaire00]

L'haire S., Mengon, J. & Laenzlinger, C. (2000). "Outils génériques et transfert hybride pour la traduction automatique sur Internet". In *Actes de TALN 2000*. Lausanne : ATALA. pp. 253-262.

[Matthews93]

Matthews, C. (1993). "Grammar Frameworks in Intelligent CALL". *CALICO*, vol. 11. pp. 5-27.

Potter, Keith R. (2002). "Taming overgeneration in a constraint-relaxed parser". Communication à *13th meeting of Computational Linguistics in the Netherlands*, Groningen, 28 novembre.

[Vosse92]

Vosse, T. (1992). "Detecting and correcting morpho-syntactic errors in real texts". In *Proceedings of the Third Conference on Applied Natural Language Processing*. Trento, Italie: ACL. pp. 111-118.

[Wehrli97]

Wehrli, E. (1997). *L'analyse syntaxique des langues naturelles: Problèmes et méthodes*. Paris: Masson.

[WeischedelBlack80]

Weischedel, R.M. & Black, J.E. (1980). "Responding intelligently to unparsable inputs". *American Journal of Computational Linguistics*, vol. 6, n°2. pp. 97-109.

Site Internet

[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 le 15 juillet 2003. <http://www.latl.unige.ch/FreeText/>

Notes

[1] Le projet FreeText relève du programme IST du 5^e programme-cadre de la Commission Européenne, contrat IST-1999-13093. Le volet suisse du projet FreeText reçoit le soutien financier de l'Office Fédéral de l'Education et de la Science, contrat 99.0049. Le contenu de cet article n'engage que ses auteurs et ne représente pas l'opinion de la Communauté Européenne. La Communauté Européenne n'est pas responsable de l'usage qui pourrait être fait des données figurant dans cet article. Les informations de ce document sont présentées telles quelles et aucune garantie n'est donnée que ces informations sont adéquates pour n'importe quelle utilisation particulière. Dès lors, l'utilisateur emploie ces informations à ses propres risques.

[2] La réinterprétation phonologique pourrait détecter d'autres catégories d'erreurs comme le genre ou le nombre, lorsque ces erreurs n'impliquent pas un changement de prononciation. Cependant, comme la réinterprétation phonologique n'est lancée qu'après le relâchement de contraintes, ces erreurs sont traitées au préalable.

[3] Un corpus de 608 phrases a été sélectionné, dont 233 contenaient des erreurs détectées par le système (38%). Au total, l'analyseur a fourni 4854 analyses avec des erreurs détectées, de 1 à 238 analyses par phrase. Les règles heuristiques ont permis d'éliminer plus de 43% de ces analyses, qui n'étaient plus que 2743. Puis les deux autres filtres de classement ont été appliqués pour sélectionner la meilleure analyse. Au final, le système s'accordait à 97% avec la meilleure analyse sélectionnée par un expert humain [Potter02].

[4] A vingt reprises, notre système de diagnostic a d'ailleurs détecté des erreurs mais leur a assigné une catégorie d'erreurs qui n'était pas la leur. Ces vingt erreurs détectées, mais mal diagnostiquées, ne sont pas prises en compte dans le tableau 3.

[5] Selon toute vraisemblance, tous les accents de la phrase ont été ajoutés à l'aide du correcteur orthographique. Aime est resté inaccentué car c'est un mot du français.

À propos des auteurs

Sébastien L'HAIRE est assistant d'enseignement et de recherche au Département de Linguistique de l'Université de Genève. Il travaille sur l'implémentation d'outils de TALN pour l'ELAO et en particulier sur le diagnostic de comparaison de phrases.

Anne VANDEVENTER FALTIN est maître assistante au Département de Linguistique de l'Université de Genève. Elle a obtenu son doctorat en septembre 2003, en se spécialisant sur l'implémentation d'outils de TALN pour l'ELAO, en particulier les vérificateurs grammaticaux pour les apprenants de langue.

Adresse : Département de Linguistique, 2, rue de Candolle, 1211 Genève 4, Suisse.

Courriel : sebastien.lhaire@lettres.unige.ch ; anne.vandeventer@lettres.unige.ch

Toile : <http://www.latl.unige.ch/personal/seb.html> ; http://www.latl.unige.ch/personal/anne_f.html

[ALSIC](#) | [Sommaire](#) | [Consignes aux auteurs](#) | [Comité de rédaction](#) | [Inscription](#)

© Apprentissage des Langues et Systèmes d'Information et de Communication, décembre 2003