ALFALEX, un environnement d'aide à l'apprentissage lexical du français langue étrangère

Thierry SELVA, Serge VERLINDE, Jean BINON

Katholieke Universiteit Leuven (KUL),
tselva@free.fr; {serge.verlinde, jean.binon}@ilt.kuleuven.ac.be
ILT, Dekenstraat 6, B-3000 Leuven, Belgique

Résumé

Cet article décrit le projet ALFALEX, un environnement d'aide à l'apprentissage du lexique français langue étrangère (FLE). Il est composé d'une dizaine d'activités lexicales en contexte portant sur de nombreuses propriétés des mots. Les exercices sont générés automatiquement à l'aide d'outils issus du TAL (Traitement Automatique des Langues), de ressources dictionnairiques et d'un corpus journalistique de 75 millions de mots, ce qui permet un grand renouvellement des épreuves et une correction plus évoluée. L'environnement enregistre le travail de l'apprenant, revient sur ses erreurs, lui permet de prendre des notes dans un dictionnaire personnel et s'adapte à son niveau. Il dresse également des statistiques sur les résultats de l'ensemble des utilisateurs, ce qui permet d'orienter l'enseignement/apprentissage du FLE. ALFALEX est actuellement testé par près de 2000 étudiants de la K.U.Leuven.

Mots-clés : lexique, apprentissage, TAL

Abstract

Alfalex is a learning environment for French as a second language (FSL). It consists of different types of exercises on ten specific features of words. Exercises are generated automatically by combining NLP-tools, electronic dictionary resources and a press corpus of 75 million words. By this means, the exercises can be constantly renewed and the correction can be very detailed. The environment saves all answers of the users, allows to make exercises on previous errors and adapts the exercises to the level of the individual user. The user can also create his own personal dictionary. Statitstics and scores on all exercises are recorded for all users. This year, Alfalex has been tested by 2000 students of the University of Leuven.

Keywords: vocabulary, learning, NLP

Introduction

ALFALEX (*Acti**e**ve Leeromgeving Frans voor Anderstaligen Lexicon,* Environnement d'apprentissage actif de français pour allophones Lexique*)* est un environnement informatique en ligne d'aide à l'apprentissage lexical du français langue étrangère (FLE) (http://www.kuleuven.ac.be/alfalex). Il se compose d'exercices sur de nombreux aspects du lexique, d'un module de diagnostic d'erreurs et d'un dictionnaire personnel. En outre, il consigne l'ensemble des réponses des étudiants dans une base de données, ce qui permet à ceux-ci de refaire des exercices suivant certains critères. Du côté enseignant, il est possible d'obtenir des statistiques sur l'ensemble des réponses. Projet de recherche, financé par la K.U.Leuven, ALFALEX a été conçu pour stimuler l'apprentissage semi-autonome et compenser ainsi le nombre réduit d'heures de cours de français en économie, droit et sciences sociales à Louvain, tandis que l'on observe un nombre toujours croissant d'étudiants et une grande hétérogénéité de leurs connaissances. Les professeurs de FLE préfèrent donc réserver le peu d'heures d'enseignement présentiel à la communication orale et écrite et laisser aux étudiants le soin de faire des exercices de rémédiation à leur rythme et selon leurs besoins**.** L'environnement est actuellement expérimenté par près de 2000 étudiants de la K.U.Leuven.

Les exercices proposés ont deux caractéristiques principales : d'une part, ce sont, pour la plupart d'entre eux, des exercices contextuels, c'est-à-dire que le mot est placé dans un contexte, en l'occurrence la phrase. Le contexte permet en effet de préciser le sens du vocable et d'étudier comment celui-ci s'emploie et fonctionne dans la chaîne discursive. Les différentes propriétés du vocable sont ainsi illustrées de manière plus complète et plus naturelle que lors d'activités hors contexte. De plus, le contexte facilite la pratique d'inférences, une stratégie essentielle en apprentissage des langues. Enfin, il donne l'occasion à l'apprenant d'être confronté à d'autres mots et d'enrichir ainsi son vocabulaire.

D'autre part, ce sont des exercices générés automatiquement (avec nécessité de corrections manuelles en amont pour plus d'exactitude). La quantité de corrections manuelles est directement liée à la qualité des outils de traitement automatique utilisés lors de l'analyse du corpus de référence. L'intérêt de la génération automatique ou semi-automatique réside dans le fait que les exercices peuvent porter potentiellement sur tout le lexique (et donc en particulier, dans le cas de l'auto-apprentissage, sur le vocabulaire que l'apprenant est en train de travailler) et peuvent être refaits avec de nouvelles épreuves, de nouvelles phrases, sans intervention (manuelle) de la part des concepteurs. En

outre, le système possède des informations sur les énoncés qui, couplées avec des ressources dictionnairiques, lui permettent de retourner un diagnostic sur la réponse un peu plus élaboré que le classique *vrai* ou *faux* . Par ailleurs, en dotant le système d'une « mémoire », on parvient à une meilleure gestion des énoncés en évitant, par exemple, de reproposer une épreuve qui a déjà été faite avec succès. Ainsi un tel environnement permet une plus grande individualisation de l'apprentissage.

Les exercices s'appuient sur deux ressources essentielles développées au sein du GRELEP (Groupe de Recherche en Lexicographie Pédagogique) à Louvain : un dictionnaire d'apprentissage, le DAFLES (Dictionnaire d'Apprentissage du Français Langue Étrangère ou Seconde, http://www.kuleuven.ac.be/dafles) [1 ; 2] et un corpus journalistique récent (Le Monde, Le Soir et Le Devoir, année 98) de 75 millions de mots [3], lemmatisé par nos soins avec le logiciel Cordial 7.

Dans cet article, nous commencerons par passer en revue le « tronc commun », c'est-à-dire l'ensemble des exercices disponibles, avant d'aborder les fonctionnalités permettant de personnaliser l'environnement et d'individualiser l'apprentissage.

Description des Exercices

Les exercices couvrent plusieurs aspects du lexique : l'aspect formel (morphologie, conjugaison, dérivation), le genre, la combinatoire (prépositions et collocations), les réseaux sémantiques (synonymie et schémas actanciels) et la traduction.

Chaque exercice se déroule en trois parties : une page d'accueil dans laquelle est présenté l'exercice et sont précisés les consignes et le barème de correction, le formulaire que l'apprenant doit compléter et valider et la correction qui propose un retour, c'est-à-dire un complément d'information avec des liens vers la ressource de référence, le DAFLES.

Exercice de morphologie

Dans cet exercice, l'apprenant est confronté avec le problème de la congruence, si typique pour le français**.** Il doit fléchir correctement un adjectif ou un nom donné en indiquant le féminin singulier, le masculin pluriel ou le féminin pluriel suivant le contexte de la phrase, c'est-à-dire, dans le cas de l'adjectif, le genre et le nombre du nom qu'il qualifie, et dans celui d'un nom, la marque du pluriel (figure 1). L'exercice ne traite que les flexions « irrégulières » (féminin différent de +e et pluriel différent de +(e)s). L'inventaire de ces terminaisons a été dressé à partir des codages morphologiques du DAFLES. On obtient en tout 33 terminaisons pertinentes pour l'exercice (*créatrices*, *travaux*, *longue*, *aiguë*, *européennes*, *sèche, fraîche, grecque, yeux*, etc.). Le questionnaire comprend une dizaine de phrases à compléter, chacune proposant une terminaison différente.

Il ne s'agit pas ici uniquement de détecter des coquilles ou des erreurs d'accent mais de déterminer et d'accepter des réponses approchantes, grâce à l'utilisation d'un dictionnaire de synonymes par exemple.

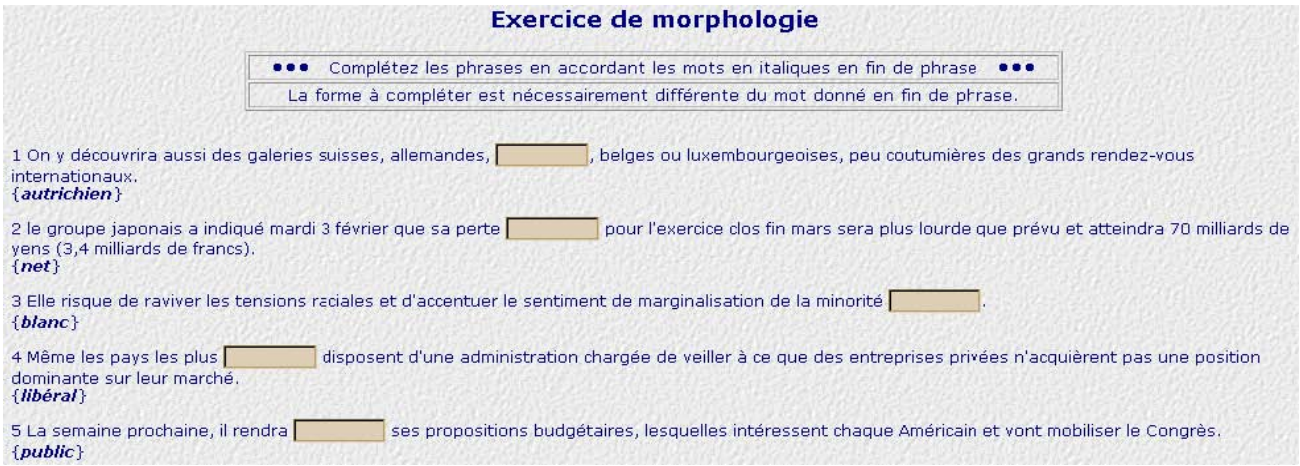


Figure 1 : extrait de l'exercice de morphologie

Du fait des accords article-nom et adjectif-nom dans la phrase, il est toujours possible de connaître d'après le contexte le genre et le nombre du mot à compléter et il n'y a pas d'ambiguïté.

Lors de la correction, le système indique l'ensemble des mots qui ont le même modèle de flexion.

Exercice hors contexte sur la conjugaison

Ce genre d'exercice n'est pas nouveau et nombreux sont les sites Internet qui proposent un entraînement à la conjugaison des verbes. C'est, en effet, pour qui dispose de l'ensemble des formes fléchies verbales, un exercice facile à mettre en œuvre en faisant varier les différents

paramètres (infinitif, temps, personne, mode). L'originalité de la conjugaison dans ALFALEX tient au fait que le lexique concerné est structuré et présenté d'une manière cohérente et pédagogiquement pertinente. Au lieu de tirer au sort une forme verbale parmi une grande quantité de verbes choisis indépendamment les uns des autres, les verbes ont été dans un premier temps regroupés par classe en fonction de leur modèle de conjugaison. On y retrouve les classiques distinctions qu'opère le Bescherelle. Ensuite, pour chacune de ces classes, seules les formes considérées comme délicates ont été retenues pour l'exercice. Ainsi, pour des verbes irréguliers comme *être* ou *avoir*, une grande partie du paradigme figure au programme tandis que, pour les verbes en *–cer* par exemple, ce sera les formes dans lesquelles le *c* du radical se transforme en *ç*.

Le retour de l'exercice permet d'avoir accès à l'ensemble des conjugaisons du verbe considéré (et non pas à celles d'un modèle de verbe) ainsi que la liste des verbes qui se conjuguent de la même manière.

Exercice sur la dérivation

Le but de cet exercice est d'explorer les familles lexicales qui sont, malgré un certain nombre de régularités, parfois imprévisibles. Ainsi, si les noms en *–tion* sont généralement des noms d'action (*punir*, *punition*, *consommer*, *consommation*, etc.), ce n'est pas toujours le cas (*imagination* n'est pas l'action d'*imaginer*). Certains dérivés sont irréguliers (l'action de *travailler* est le *travail*), d'autres n'existent pas (**allocateur* n'est pas le sujet d'*allouer*), d'autres se confondent et se mélangent dans l'esprit des apprenants (une erreur courante chez les apprenants néerlandophones est d'employer *fondeur* (du verbe *fondre*) au lieu de *fondateur* pour le sujet de *fonder* ou **exporteur* au lieu de *exportateur*, par interférence avec le néerlandais *exporteur*). De tels phénomènes justifient un exercice sur les dérivés.

L'environnement dispose pour cela d'environ 200 familles lexicales et de 500 dérivés. Il tire au sort deux ou trois verbes et affiche l'ensemble des dérivés (verbe, noms, adjectifs) en contexte. L'apprenant doit alors retrouver ces dérivés et les placer correctement dans les phrases. Si les mots sont corrects mais mal fléchis, le système accorde la moitié des points.

Le retour de l'exercice ne propose pas une liste en vrac des dérivés du verbe considéré mais les dispose, tout comme l'exercice sur les schémas actanciels (voir 2.8), dans des tableaux qui mettent en valeur les relations entre membres d'une même famille en fonction de leurs acceptions. En effet, *employeur* n'est pas le sujet de tous les sens d'*employer* (*employer un outil*, *la force*, etc.).

Exercice sur le genre

Dans cet exercice, l'apprenant doit retrouver le genre des noms en complétant l'article et/ou l'adjectif qui s'accorde avec le nom. Ne sont pas pris en compte les articles et adjectifs invariables en genre (*chaque*, *souple*, *les*, *deuxième*, etc.) qui sont laissés tels quels dans la phrase. Le système contient environ 2850 noms communs. Les caractères qu'il faut retrouver sont, pour chaque mot, les désinences des diverses formes du masculin, féminin et pluriel. On enlève également la dernière lettre du radical pour obliger l'utilisateur à taper au moins une lettre afin qu'il n'obtienne pas une bonne réponse en laissant le champ vide.

Etant donné la masse de noms qui interviennent dans l'exercice, il n'est pas pédagogiquement pertinent de les tirer aléatoirement pour la constitution des énoncés : la quantité d'information à retenir serait trop grande pour couvrir l'ensemble des noms pris un à un. L'un des buts de cet exercice est d'attirer l'attention de l'apprenant sur certaines règles qui régissent le genre des mots. Dans le DAFLES, les noms ont été répartis en deux groupes. Un premier groupe contient tous les noms dont le genre est déterminé par la terminaison. Par exemple, tous les noms se terminant par *–t* sont masculins, sauf exceptions peu nombreuses. Il y a ainsi 1700 noms qui se répartissent en 22 classes. Les exceptions représentent 4% du total.

Un deuxième groupe contient tous les autres noms, dont une grande partie se terminent par *–e*. Un certain nombre a été identifié comme « posant problème » : systématiquement, un grand nombre d'étudiants néerlandophones se fourvoient sur le genre de *problème, système, interview*, etc. L'environnement insiste alors sur ces noms qui concernent un tiers des phrases réservées aux noms du deuxième groupe. La part d'arbitraire et d'intuition concernant la classification de ces noms « difficiles » sera levée peu à peu. En effet, en fonction du taux de réponses correctes, le statut des ces noms sera amené à être modifié pour devenir plus objectif.

La correction est ici assez sommaire puisque le système compare les lettres entrées par l'utilisateur avec celles réellement dans la phrase et attribue un vrai ou un faux. Par contre, le retour de l'exercice (figure 2) constitue un précieux outil d'apprentissage car il explicite les règles permettant de déterminer le genre des noms, ou du moins de dégager quelques régularités. Ainsi, l'apprenant saura (le plus souvent découvrira) que tous les noms se terminant par *–u* sont masculins sauf trois exceptions qui sont citées (*eau, tribu, vertu*).

réponse correcte	*votre réponse*	*renvoi vers le Dafles*	*nom à reproposer sous la forme d'un exercice ?*	*nom à ajouter au dico personnel ?*
Un type de la trempe de Robert Waseige, à qui le FCLiégois aurait dû ériger, en son temps, **une statue**, a lui aussi été licencié quelques mois avant de boucler un cycle de 9 années dans le même club.	**bonne** réponse	Genre des mots se terminant par -ue	☐	☐
Plus tard, ce sera au tour d'Eugen Drewerman, prêtre et psychanalyste, d'être suspendu de toutes ses fonctions pour avoir levé **le tabou**	**bonne** réponse	Genre des mots se terminant par -u	☐	☐
Les clubs évo comme Court en sont, aujo spécialement pressés de remonter.		re des mots se terminant par -age	☐	☐
Il a déclaré que, malgré **ce conservatisme**, les retombées de l'étude de 1997 étaient plus importantes que celles qui s'étaient dégagées d'une étude réalisée en 1990, un peu moins conservatrice.	**bonne** réponse	Genre des mots se terminant par -sme	☐	☐

Aide — Close

Les mots se terminant par **-u** sont **masculins** :
beau, matériau, flou, bijou, genou, milieux, dieu, revenu, niveau, noyau, bleu, trou, porte-drapeau, barreau, carreau, tombeau, tabou, gâteau, cadeau, couteau, jeu, individu, jeu, bienvenu, fardeau, etc.
Exceptions : eau, tribu, vertu.

Figure 2 : correction et retour de l'exercice sur le genre

Exercice sur les prépositions

Dans cet exercice, l'utilisateur doit compléter la préposition s'accordant avec un verbe, un nom ou un adjectif (*continuer à/de*, *envie de*, *apte à*, etc.). L'environnement dispose d'environ 4800 phrases pour 654 constructions différentes. La préposition peut parfois être vide dans le cas de la construction Verbe+infinitif (*espérer réussir*). Le repérage des groupes prépositionnels étant délicat et source de beaucoup d'erreurs par les logiciels de traitement automatique, seules les constructions où la préposition suit immédiatement le mot sont prises en compte.

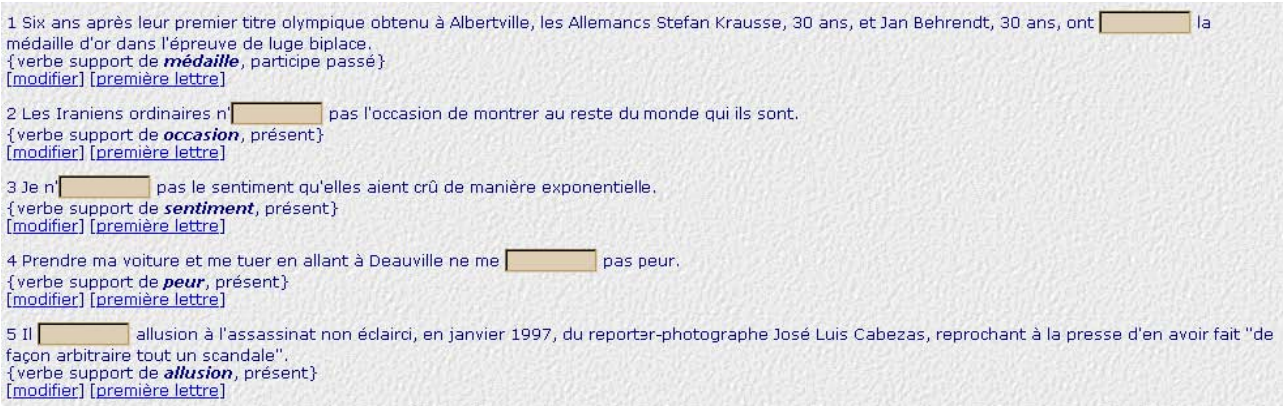
Le retour de l'exercice affiche les mots se construisant avec la même préposition.

Exercice sur les collocations

Le but de l'exercice est de compléter une des composantes d'une collocation à partir du contexte (et donc du reste de la collocation). Parmi toutes les combinaisons de catégories grammaticales, nous avons retenu les combinaisons nom+verbe, et plus précisément les verbes support sémantiquement vides (*poser une question*, *donner un ordre*, *jouer un rôle*, etc.). Ces combinaisons correspondent en fait aux fonctions lexicales *Oper* et *Func* [4] qui sont de loin les plus répandues et les plus faciles à comprendre. Pour d'autres fonctions, l'énoncé risquerait d'être trop ambigu et rendrait difficile un diagnostic précis. L'environnement propose environ 850 collocations.

Cet exercice est incontestablement le plus difficile parmi les exercices contextuels, surtout pour les collocations les moins fréquentes, car bien que très nombreuses et très utilisées dans la langue par les natifs, les collocations sont difficiles à maîtriser en production pour tout apprenant d'une langue. Les combinaisons de mots sont en effet arbitraires, imprévisibles et suivent une syntaxe précise. Leur emploi régulier par un locuteur dénote d'ailleurs sa bonne maîtrise de la langue.

Pour tenir compte de cette difficulté, l'exercice aide quelque peu l'utilisateur en proposant, sur demande, la première lettre du verbe conjugué (ce qui coûte quelques points). D'autre part, il est possible de modifier la phrase, si elle n'est pas comprise (figure 3).



1 Six ans après leur premier titre olympique obtenu à Albertville, les Allemancs Stefan Krausse, 30 ans, et Jan Behrendt, 30 ans, ont ☐ la médaille d'or dans l'épreuve de luge biplace.
{verbe support de ***médaille***, participe passé}
[modifier] [première lettre]

2 Les Iraniens ordinaires n'☐ pas l'occasion de montrer au reste du monde qui ils sont.
{verbe support de ***occasion***, présent}
[modifier] [première lettre]

3 Je n'☐ pas le sentiment qu'elles aient crû de manière exponentielle.
{verbe support de ***sentiment***, présent}
[modifier] [première lettre]

4 Prendre ma voiture et me tuer en allant à Deauville ne me ☐ pas peur.
{verbe support de ***peur***, présent}
[modifier] [première lettre]

5 Il ☐ allusion à l'assassinat non éclairci, en janvier 1997, du reporter-photographe José Luis Cabezas, reprochant à la presse d'en avoir fait "de façon arbitraire tout un scandale".
{verbe support de ***allusion***, présent}
[modifier] [première lettre]

Figure 3 : extrait de l'exercice sur les collocations

Contrairement au genre, la correction est ici délicate car beaucoup de facteurs entrent en jeu dans la réponse. Il y a tout d'abord le problème de la conjugaison du verbe et le système octroie tout de même des points si le verbe est correct mais mal conjugué. Ensuite, certains noms acceptent plusieurs verbes support synonymes (on peut tout aussi bien *jouer* un match que le *disputer*). Dans ces cas-là, le système parcourt la base des verbes support et vérifie que la réponse s'y trouve, à construction syntaxique identique. Si c'est le cas, il considère la réponse juste, bien que ce ne soit pas le même verbe dans la phrase. On mesure ici tout l'avantage de l'intégration du DAFLES dans l'environnement d'apprentissage. Enfin, certains verbes support ne sont pas sémantiquement vides, apportent une nuance et sont exprimés par d'autres fonctions lexicales (*Incep*, *Real*, etc.) moins intuitives (*aborder, résoudre, soulever une question* au lieu *de poser une question*). Le système en tient compte en accordant quelques points. Tout cela implique une grande cohérence et une grande précision dans le codage et la description des collocations dans le DAFLES.

Le retour lors de la correction affiche l'ensemble des collocations qui se construisent avec le verbe ou le nom. On peut ainsi se rendre compte facilement que *commettre* s'emploie avec des noms exprimant un délit ou des concepts négatifs. En cliquant sur le lien, on est dirigé vers l'endroit précis où la collocation est décrite dans le dictionnaire.

Pour évacuer le problème du temps en l'absence de contexte plus large (dans le cadre strict de la phrase, un futur peut tout aussi bien convenir qu'un présent), celui-ci est indiqué à la fin de la phrase.

Exercice hors contexte sur les synonymes

La maîtrise des synonymes est un élément important dans la richesse de l'expression tant orale qu'écrite. Le but de cet exercice est de trouver un des synonymes d'un mot donné, dans une acception donnée, précisée par un exemple. Il exploite en cela les groupes de synonymes (ou *synset* dans la terminologie WordNet [5]) constitués dans le DAFLES. Pour aider l'utilisateur, l'initiale du mot à trouver est fournie. Au stade actuel, l'exercice est imparfait car un certain nombre de mots du DAFLES n'ont pas encore d'exemple, ce qui ne facilite pas la tâche. Des essais ont été tentés pour replacer les mots dans de vraies phrases extraites du corpus. Malheureusement, le remplacement d'un mot par un synonyme est très fortement dépendant du contexte et il n'est pas possible d'isoler de manière automatique les phrases qui permettraient ces remplacements sans commettre de contresens. Dès lors l'exercice est non contextuel, l'exemple ne servant qu'à préciser l'acception. Afin de s'assurer de la bonne compréhension, il est envisagé de fournir la traduction dans la langue maternelle.

Le retour de l'exercice affiche l'ensemble des synonymes du mot en question.

Exercice hors contexte sur les schémas actanciels

L'exercice consiste à explorer l'axe syntagmatique et paradigmatique du lexique en demandant à l'apprenant quels sont les actions, sujets, compléments, etc. prototypiques d'un verbe donné (figure 4). Il reprend en cela les quelque 200 schémas actanciels présents dans le DAFLES.

Les schémas permettent de considérer le lexique non pas comme un « sac de mots » informe, mais comme des ensembles structurés de mots reliés par la morphologie, la syntaxique, la sémantique ou le contexte dans une situation de communication donnée.

L'élément innovant de cet exercice est la fonction d'aide qui apparaît lorsqu'on promène la souris sur l'un des formulaires HTML et qui pose une question en langage naturel. Ainsi, dans la figure 4, on demande à l'apprenant le sujet prototypique du verbe *construire* (*constructeur*) dans le sens 2, *réaliser un appareil*.

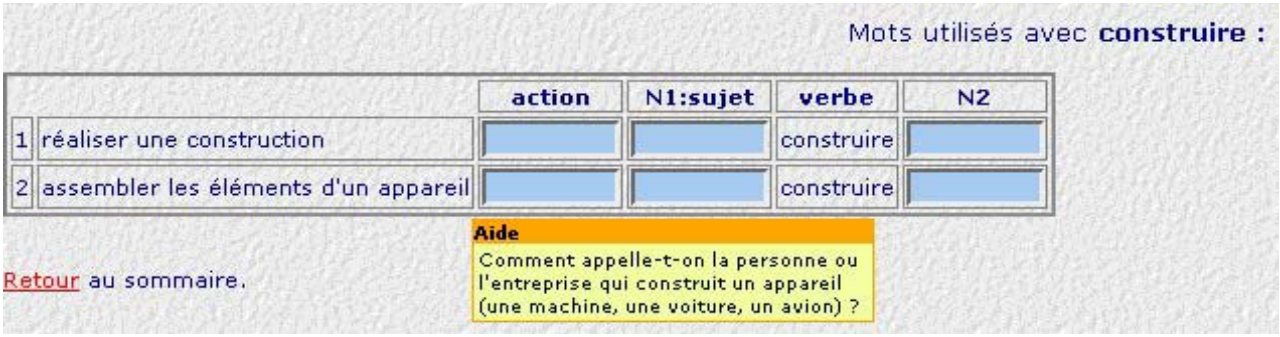


Figure 4 : exercice sur le schéma actanciel de *construire*

Ces questions sont disponibles pour l'action et le sujet du verbe. Elles sont générées automatiquement par le système et exploitent la systématicité des définitions du DAFLES. Il s'agit de définitions phrastiques dans le style rédactionnel du Collins COBUILD [6]. Elles explicitent la construction syntaxique du verbe et ses actants, notamment le sujet. La question est composée de deux parties : les compléments du verbe, s'il y en a, qui sont

tirés de la définition du verbe, et son sujet, tiré de la définition de l'acception du nom défini comme sujet par les schémas actanciels. Pour les autres fonctions syntaxiques (COD, COI, compléments circonstanciels, de manière, etc.), seule la définition du mot est proposée car la forme des questions est trop variable. De même que pour les dérivés, le retour de l'exercice oriente l'utilisateur vers les schémas actanciels du DAFLES.

Exercice sur les traductions

Le but de cet exercice est de vérifier l'un des éléments les plus importants de la connaissance d'un mot : son sens. Pour cela, l'utilisateur est invité à fournir un équivalent dans sa langue maternelle. L'exercice est composé de phrases à compléter à la fin desquelles est indiqué le mot à traduire. De la même manière que pour les synonymes, la phrase affichée est l'exemple du DAFLES pour le mot et l'acception donnés, en raison de la difficulté de la désambiguïsation sémantique automatique. En réalité, l'exercice est possible dans les deux sens : de la langue étrangère vers la langue maternelle et *vice versa*. Suivant le cas, on indique à la fin de la phrase le mot français ou sa traduction.

Pour l'instant, cet exercice concerne uniquement le néerlandais (2200 traductions disponibles). Mais d'autres langues viendront s'ajouter au fur et à mesure de l'avancement des descriptions du DAFLES.

Personnalisation

Sur la base commune que nous venons de décrire, l'environnement est personnalisable de nombreuses façons et chaque apprenant l'utilisera de manière différente.

L'environnement est disponible sous deux modes : un accès libre et un accès contrôlé. L'accès libre est le mode par défaut et concerne tout utilisateur. Dans un premier temps, lors de la connexion, le système demande une identification et réserve un emplacement dans la base de données. Si l'utilisateur revient sur le site, il retrouve les résultats de ses sessions précédentes. S'il se connecte en tant qu'anonyme (identification vide), il a accès à tout l'environnement mais ses résultats sont effacés au bout de deux jours.

L'accès contrôlé est réservé à des groupes d'utilisateurs qui se seront déclarés à l'administrateur du système (contact par courrier électronique disponible sur la page d'accueil). C'est le cas pour les étudiants de la K.U.Leuven qui se connectent en indiquant leur numéro d'étudiant et leur mot de passe habituel. Pour l'instant, le système ne gère les mots de passe que pour les étudiants de Louvain. L'accès au système se fait via une adresse différente qui détermine un numéro de groupe. Il est alors possible d'avoir des statistiques sur le travail de ce groupe. D'autre part, le score de chacun est calculé de manière plus précise (nous reviendrons sur ces points un peu plus loin).

Niveau de difficulté

La plupart des exercices (sauf la morphologie, les synonymes et les schémas actanciels) possèdent deux niveaux de difficulté, déterminés par la fréquence de l'information en jeu dans le corpus de référence. Il y a un rapport direct entre la fréquence d'une information et sa maîtrise puisqu'on observe en moyenne une différence de performance de 15% entre les deux tranches. Ainsi des collocations de faible fréquence (*vaquer à une occupation*) seront très difficiles à trouver, au contraire de *jouer un rôle* ou *poser une question*, beaucoup plus fréquentes. De même pour le genre, entre un mot connu et un mot moins courant. Tandis qu'en accès libre, l'utilisateur peut s'exercer dans la tranche qu'il désire, l'accès contrôlé impose une tranche en fonction des résultats de l'apprenant. S'il obtient de bons résultats (80% et plus), il pourra essayer le groupe le plus difficile, sinon il continue avec des informations de niveau de difficulté comparable. De même, le niveau baissera s'il n'obtient pas d'assez bons résultats (50% et moins). Dans un premier temps, un test (exercice à trous indépendant) est proposé lors de la première connexion en accès contrôlé (aucun autre exercice n'est disponible). En fonction du résultat à ce test, l'utilisateur sera orienté vers des exercices de niveau facile ou difficile. Le but est bien évidemment de s'adapter au niveau de l'apprenant en lui proposant des épreuves qui soient, autant que possible, ni triviales, ni hors de portée.

Statut des informations

Après validation des réponses, une correction automatique est effectuée et, pour chaque utilisateur, chaque information traitée se voit attribuer un statut (maîtrisé ou non maîtrisé) en fonction du résultat et d'un barème. L'environnement garde donc en mémoire tout ce qui est fait, ce qui lui permet de présenter toujours de nouvelles épreuves avec de nouvelles informations. Ainsi, d'une session à l'autre, on évite de reposer les mêmes questions ou de reproposer les mêmes phrases.

A l'issue de l'exercice, si le mot est considéré comme non maîtrisé, il sera mis dans une liste d'erreurs et l'utilisateur pourra revenir dessus quand bon lui semblera. Tout de suite, le lendemain ou dans un mois. Cela lui permet de vérifier s'il a retenu la correction. Tant que la réponse est fausse, le mot reste dans la liste d'erreurs.

Au contraire, si le mot est considéré comme maîtrisé, il ne sera plus proposé. Il se peut toutefois que l'étudiant ait hésité avant de donner la bonne réponse et ne soit pas sûr de bien connaître le mot en question. Dans ce cas, afin de renforcer sa connaissance, le système offre la possibilité à l'apprenant d'être réinterrogé sur ce mot. Là aussi l'exercice peut être refait n'importe quand. Si l'étudiant se trompe, le mot va dans la liste des erreurs. S'il réussit et ne souhaite plus refaire d'exercice sur ce mot, celui-ci n'est plus proposé.

Dictionnaire personnel

Le dictionnaire personnel est une fonctionnalité importante de l'environnement. Il permet à l'étudiant de conserver dans sa base de données les informations qu'il a travaillées si celles-ci lui semblent importantes (figure 5). Il lui suffit de cocher les informations correspondantes lors de la correction. Avec l'information sont conservés la réponse de l'apprenant, la réponse juste et l'énoncé de l'exercice. L'apprenant a également la possibilité de mettre, de modifier ou d'effacer une annotation.

Dictionnaire personnel de Brecht Adriaens

date	exercice	mot	réponse correcte	votre réponse	renvoi vers DAFLES	vos commentaires	effacer
2003-11-07	dérivés	informer	Les ministères de l'intérieur et de la justice en avaient été **informés**, après que la brigade financière se fut procurée les différentes adresses du président du Conseil constitutionnel.	forme existante mais accord incorrect votre réponse : informées	dérivés et actants de informer	ajouter un commentaire	effacer la ligne
2003-11-07	dérivés	intéresser	La meilleure coopération qui élimine tous les conflits d'**intérêt**, c'est la fusion.	votre réponse : intéresses	dérivés et actants de intéresser	ajouter un commentaire	effacer la ligne
2003-11-07	dérivés	informer	La concurrence des nouveaux médias électroniques et des chaînes câblées a bouleversé les règles du jeu et accéléré le rythme de l'**information**.	votre réponse : informatique	dérivés et actants de informer	ajouter un commentaire	effacer la ligne
2003-11-07	morphologie	rugbymar	" Keep the ball in hand ! ", répètent à satiété les **rugbymen** anglais.	rugbyhommes	autres mots sur le modèle de *rugbyman*	ajouter un commentaire	effacer la ligne
2003-11-07	morphologie	bref	La rencontre a été **brève**, mais suffisamment symbolique pour satisfaire M. Madelin.	brèves	autres mots sur le modèle de *bref*	ajouter un commentaire	effacer la ligne
2003-11-07	morphologie	faux	L'étude attentive des éléments obtenus çà et là en Algérie révèle que la première impression de massacres commis au hasard est **fausse**.	fause	autres mots sur le modèle de *faux*	ajouter un commentaire	effacer la ligne
2003-11-07	morphologie	fou	C'est une période de ma vie où je prenais mon pied à faire des choses **folles**, violentes."	foux	autres mots sur le modèle de *fou*	ajouter un commentaire	effacer la ligne
2003-11-07	conjugaisons	être	étant	pas de réponse	verbe(s) du même modèle	ajouter un commentaire	effacer la ligne

Figure 5 : dictionnaire personnel d'un étudiant

Pour faciliter le parcours de ce dictionnaire et le rendre ainsi plus pratique, l'étudiant peut le classer suivant différents critères : par exercice, par date ou par mot. En procédant à un classement par mot, il peut voir si celui-ci a été travaillé dans plusieurs exercices (verbe de la collocation et conjugaison, genre et traduction, etc.). Cette fonctionnalité semble très appréciée des étudiants, puisque nombreux sont ceux qui disposent d'un dictionnaire de plusieurs dizaines d'enregistrements (en moyenne, plus de 20 par étudiant).

Scores

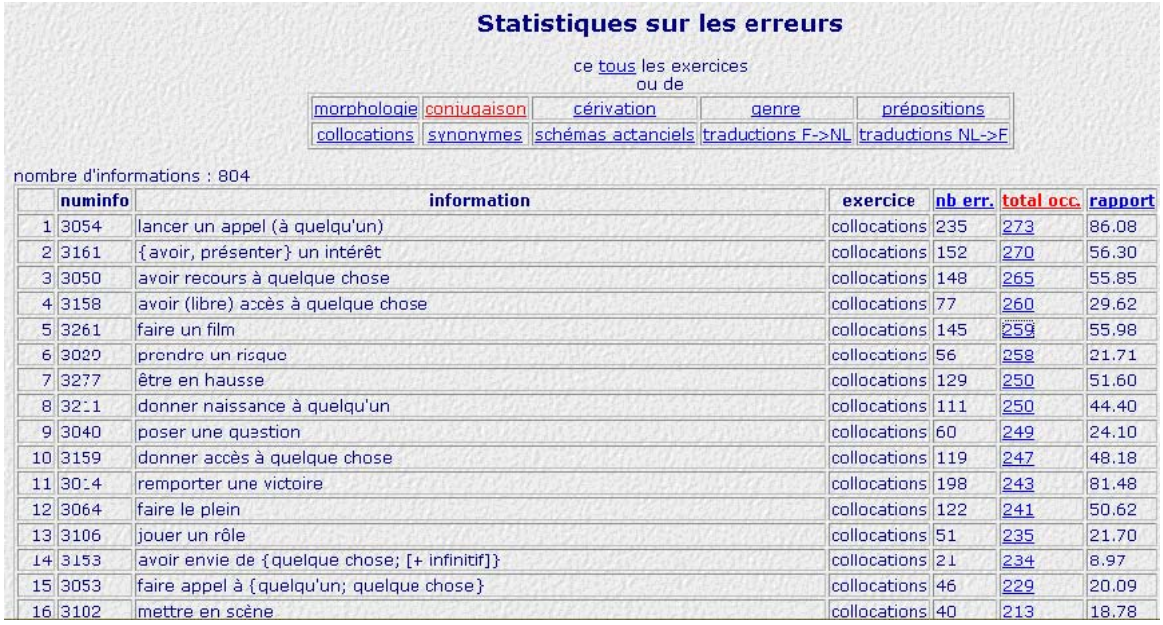
Chaque étudiant peut visualiser ses résultats qui sont proposés sous forme de tableau synthétique. Les notes, sous forme de pourcentages, sont affichées en fonction de l'exercice et de la difficulté. Les pourcentages sont calculés pour chaque mois. Une dernière ligne indique les erreurs restantes dans chaque exercice.

Il est à noter une différence dans le calcul des scores entre l'accès libre et l'accès contrôlé. Tandis que l'accès libre comptabilise l'ensemble des réponses données aux exercices, l'accès contrôlé n'opère que sur les premières réponses, sans prendre en compte le retour sur les erreurs. Cela donne des résultats plus objectifs, car il serait facile, après avoir vu la correction, d'améliorer ses performances.

Côté enseignant

L'enseignant a accès à un petit menu d'utilitaires : il peut ainsi procéder à diverses opérations de maintenance sur le système comme l'effacement d'étudiants qui n'ont plus utilisé l'environnement depuis plus de sept mois. Mais il peut aussi prendre connaissance des scores de l'ensemble des étudiants (affichage dans un tableau lisible par Excel) ainsi que des statistiques sur les réponses données par groupe (figure 6).

Les statistiques indiquent pour chaque information le nombre de fois qu'elle est apparue dans les exercices, le nombre d'erreurs et le taux de réussite. Le tableau peut être classé suivant ces critères. En cliquant sur un lien, on a également le détail de la réponse (réponse donnée mais aussi le score, le contexte, le numéro d'élève et la date). Ceci est très important pour voir quelles sont les connaissances exactes des apprenants, où en est leur interlangue. Cet outil met en lumière leurs forces et leurs faiblesses et permet ainsi d'optimiser le contenu des cours et l'approche de l'enseignement / apprentissage du FLE.



Statistiques sur les erreurs

ce tous les exercices
ou de

morphologie	conjugaison	dérivation	genre	prépositions
collocations	synonymes	schémas actanciels	traductions F->NL	traductions NL->F

nombre d'informations : 804

	numinfo	information	exercice	nb err.	total occ.	rapport
1	3054	lancer un appel (à quelqu'un)	collocations	235	273	86.08
2	3161	{avoir, présenter} un intérêt	collocations	152	270	56.30
3	3050	avoir recours à quelque chose	collocations	148	265	55.85
4	3158	avoir (libre) accès à quelque chose	collocations	77	260	29.62
5	3261	faire un film	collocations	145	259	55.98
6	3029	prendre un risque	collocations	56	258	21.71
7	3277	être en hausse	collocations	129	250	51.60
8	3211	donner naissance à quelqu'un	collocations	111	250	44.40
9	3040	poser une question	collocations	60	249	24.10
10	3159	donner accès à quelque chose	collocations	119	247	48.18
11	3014	remporter une victoire	collocations	198	243	81.48
12	3064	faire le plein	collocations	122	241	50.62
13	3106	jouer un rôle	collocations	51	235	21.70
14	3153	avoir envie de {quelque chose; [+ infinitif]}	collocations	21	234	8.97
15	3053	faire appel à {quelqu'un; quelque chose}	collocations	46	229	20.09
16	3102	mettre en scène	collocations	40	213	18.78

Figure 6 : statistiques sur les 15 premières collocations les plus fréquentes dans les exercices

Dernières Fonctionnalités

Actuellement, trois fonctionnalités sont en train d'être ajoutées :

- un retour trois semaines après sur les erreurs commises et corrigées par les apprenants, afin de mesurer le degré d'appropriation du vocabulaire. Il s'agit ici d'évaluer en toute objectivité le rendement de l'environnement,
- un dispositif permettant de travailler sur les erreurs les plus fréquentes de l'ensemble des étudiants. Il s'agit ici d'optimiser l'apprentissage en mettant l'accent sur le vocabulaire posant généralement des difficultés. Ces mots peuvent varier en fonction des groupes d'apprenants et sont sélectionnés sur un critère de fréquence à partir des réponses de ces groupes,
- un système d'aide visant à renforcer le rôle pédagogique de l'environnement. Pour les exercices les plus difficiles (synonymes, schémas actanciels, collocations, etc.), les erreurs les plus fréquemment commises par l'ensemble des apprenants sont retenues et présentées, avec la solution, sous forme de QCM. Il d'agit ici de proposer des alternatives plausibles et de donner une piste pour résoudre l'exercice.

Conclusion

Face à la quantité incroyable d'informations que représente le lexique d'une langue, l'environnement ALFALEX tente de présenter celui-ci de manière structurée et cohérente. Les différentes régularités sont explicitées autant que possible, ceci afin de favoriser l'apprentissage et la rétention du vocabulaire. Testant dans un premier temps les connaissances de l'apprenant, le système assure un suivi personnalisé (retour sur les erreurs, possibilité de refaire un exercice pour telle ou telle information, dictionnaire personnel) permettant de répéter l'exposition aux mots mal maîtrisés. Il se distingue ainsi des nombreux sites sur Internet qui ne dispensent pour la plupart qu'un travail ponctuel sur le vocabulaire.

Références

[1] Selva, T. ; Verlinde, S. ; Binon, J. 2002. Le DAFLES, un nouveau dictionnaire pour apprenants du français. In Actes du dixième congrès EURALEX'2002 (European Association for Lexicography). Copenhague. 199-208.

[2] Selva, T. ; Verlinde, S. ; Binon, J. 2003, Vers une deuxième génération de dictionnaires électroniques. Traitement Automatique des Langues : Les dictionnaires électroniques :: pour les personnes, les machines ou pour les deux ?, vol. 44, n° 2.

[3] Verlinde, S. ; Selva, T. 2001. Nomenclature de dictionnaire et analyse de corpus. Cahiers de Lexicologie, vol. 79, n° 2. 113-139.

[4] Mel'čuk, I. ; Clas, A. ; Polguère, A. 1995. *Introduction à la lexicologie explicative et combinatoire*. Louvain-la-Neuve : Duculot.

[5] WordNet. 2003. WordNet – a Lexical Database for English, consulté en avril 2004 : http://www.cogsci.princeton.edu/~wn/

[6] Cobuild. Sinclair, J. éd. 1995. *Collins Cobuild English Dictionary*. London : Harper Collins.