

Alfalex : un environnement d'apprentissage du vocabulaire français en ligne, interactif et automatisé

Serge Verlinde, Thierry Selva, Jean Binon

1. La genèse du projet Alfalex

Depuis quelques années, la K.U.Leuven mène une politique didactique qui vise à stimuler l'apprentissage semi-autonome et le développement d'outils informatiques qui favorisent ce type d'apprentissage. C'est dans cette perspective qu'ont été lancés par le professeur Debrock les projets OOI (Onderwijsgerichte onderzoeks-, ontwikkelings- en implementatieprojecten), financés par les autorités académiques de notre université.

Le projet OOI *Alfalex* (*Actieve leeromgeving Frans voor anderstaligen - lexicon*) s'inscrit dans cette perspective et constitue un prolongement des travaux en lexicographie pédagogique (dictionnaires d'apprentissage du français des affaires) entamés depuis une dizaine d'années par un petit groupe de recherche composé de professeurs de français de l'Institut interfacultaire des langues vivantes.¹

Le projet Alfalex poursuit deux objectifs bien distincts, mais complémentaires. Dans la perspective OOI, nous voulions développer un outil informatique qui permette aux étudiants en économie, en droit et en sciences sociales, qui ont un cours de français dans leur cursus, de mettre à niveau leurs connaissances de base du français, du moins en ce qui concerne le vocabulaire, sous tous ses aspects. Chaque année, nous constatons en effet d'énormes différences de niveau entre ces étudiants sans avoir pu, jusqu'à présent, trouver de remède adéquat. Le côté recherche appliquée du projet porte davantage sur l'exploitation que l'on peut faire de ressources informatisées. Le fait de disposer d'un dictionnaire électronique déjà bien avancé pour le français général² et d'un cor-

¹ Pour plus d'informations, voir le site www.kuleuven.ac.be/grelep.

² *Dafles : Dictionnaire d'apprentissage du français langue étrangère ou seconde* (www.kuleuven.ac.be/dafles). Voir Verlinde, Selva et Binon (2002) et Selva, Verlinde et Binon (à paraître) pour une description du *Dafles*. Actuellement sont décrits les sens des 5 000 lemmes, soit plus de 6 600 mots et 14 000 sens (voir 3.2.), les plus fréquents du français général tels qu'ils ont été relevés dans un corpus de 50 000 000 de mots (*Le Monde* et *Le Soir* 1998) (Verlinde et Selva 2001). Pour approximativement la moitié des sens, nous avons une phrase exemple accompagnée d'une représentation formelle de la structure syntaxique du mot (par exemple: *N1 V N2*: Jean aime Marie, *aimer*, verbe transitif ; *N en N3* : l'approvisionnement en carburant, etc.). En outre, quelque 13 000 combinaisons de mots (locutions, expressions, collocations, proverbes) ont été enregistrées et classées. Pour un quart de ces combinaisons, on trouve dès à présent dans la base une définition et/ou une phrase exemple. Les traductions vers le néerlandais sont disponibles pour plus d'un quart des sens.

pus de plus de 75 000 000 de mots, soit quelque 150 000 pages de texte³, nous a incités à examiner comment on pouvait utiliser ces deux importantes sources de données pour générer des exercices de la façon la plus automatisée possible.

Le projet, qui touche presque à sa fin, a donné lieu à un environnement d'apprentissage du vocabulaire qui est accessible en ligne et qui s'inspire de l'environnement d'apprentissage Alexia (Selva 1999), tout en y ajoutant de nombreuses fonctionnalités. L'environnement propose des exercices portant sur différentes propriétés des mots, une correction immédiate et, en même temps, un retour d'information (feedback) très performant. Dans ce qui suit, nous présentons les principales caractéristiques de l'environnement (3.), suivies du détail des différents exercices proposés (4.) et des fonctionnalités annexes (5.). Nous terminons par une présentation des premiers résultats de l'utilisation de l'environnement par nos étudiants (6.). Nous débutons toutefois par un bref aperçu des exercices de vocabulaire que l'on trouve sur le net (2.) pour mieux mettre en évidence les différences par rapport à notre environnement.

2. Les exercices de vocabulaire sur le net

Ces exercices sont particulièrement nombreux. Pour s'en rendre compte, il suffit par exemple de taper « exercice », « vocabulaire » et « français » sur le site de Google ou de se reporter aux répertoires d'adresses.⁴ Les exercices se présentent sous des formes très variées : QCM, phrases à trous, mots croisés, etc., mais ils portent essentiellement sur des éléments isolés du vocabulaire. Les critères de sélection des mots repris dans ces exercices ne sont pas transparents, si ce n'est qu'il s'agit éventuellement d'exercices sur le vocabulaire vu dans un dossier, un cours... En outre, la création de ce genre d'exercices est relativement fastidieuse. Même si l'on dispose actuellement d'excellents systèmes auteur, dont l'un des plus répandus est sans conteste Hot Potatoes (web.uvic.ca/hrd/halfbaked), le professeur est contraint de chercher ou de rédiger une phrase par exercice. La productivité est donc faible, ce qui explique que les exercices proposés sont généralement en nombre relativement restreint.

Exceptions notoires, les quelques sites qui proposent des exercices sur la conjugaison des verbes. En effet, comme il est possible de lister toutes les formes des verbes, et que ces formes peuvent être paramétrées (verbe, mode, temps, etc.), on peut générer une multitude d'exercices en faisant varier ces paramètres. L'exemple du site canadien *Le devoir conjugal* (<http://www.pomme.ualberta.ca/devoir/exer.htm>) en est une parfaite illustration.

Seul le deuxième type d'exercice semble vraiment constituer un outil de remédiation performant, avec des exercices suffisamment renouvelables, une couverture complète des cas et la possibilité, pour l'apprenant ou le professeur, de déterminer un véritable parcours d'apprentissage en faisant des exercices ciblés.

³ Le corpus comprend les textes originaux des numéros de l'année 1998 des journaux *Le Monde* et *Le Soir* et des années 1998 et 1999 du quotidien *Le Devoir* (Québec).

⁴ Par exemple Clicnet (<http://www.swarthmore.edu/Humanities/clicnet/fle.html>) ou le Weboscope (<http://millennium.arts.kuleuven.ac.be/weboscope/francais/index.htm>).

3. Les caractéristiques d'Alfalex

Les principales caractéristiques des exercices d'Alfalex sont les suivantes :

3.1 Variété

Le Dafles a été conçu dès le départ comme un dictionnaire électronique sous forme d'une base de données relationnelle. Comme nous avons, dès le départ, tenu compte d'exploitations ultérieures des données de ce dictionnaire en structurant les informations de façon très hiérarchisée et très détaillée, elles sont directement utilisables comme point de départ pour la génération d'exercices :

- les aspects formels du mot
 - o la morphologie : terminaisons des noms et des adjectifs (*secret-secrète, muet-muette*)
 - o la conjugaison des verbes
 - o la dérivation : les familles de mots (*acheter, achat, acheteur*)
- le genre des noms
- la combinatoire des mots
 - o l'emploi des prépositions après les verbes, les noms et les adjectifs
 - o les collocations (*poser une question, une ambiance règne*)
- les champs sémantiques ou mots qui s'associent dans une situation de communication
 - o les synonymes et les antonymes
 - o les schémas actanciels (comment appelle-t-on l'action de tuer une personne ? *un assassinat* ; comment appelle-t-on la personne qui a tué une autre personne ? *un assassin* ; et la personne qui a été tuée ? *la victime*)
- la traduction
 - o en compréhension : français > néerlandais
 - o en production : néerlandais > français

L'éventail d'exercices pourrait encore être élargi en intégrant par exemple des exercices d'orthographe.

3.2 Performance

Sous 1., nous avons donné quelques chiffres concernant la taille actuelle du Dafles. Nous nous basons sur ces données pour générer les exercices. Ainsi, pour le genre, nous puisons dans un stock de plus de 3 700 noms. Compte tenu de la taille de notre corpus, nous disposons en outre d'une source quasi inépuisable de phrases pour alimenter les exercices. Pour l'exercice sur le genre par exemple, nous avons limité le nombre de phrases par mot à 10, ce qui est largement suffisant, même si l'on désire revenir plusieurs fois sur le même mot (voir ci-dessous).

Pour améliorer la performance de l'environnement et ne pas donner l'impression à l'apprenant qu'il va être submergé de données et d'exercices, nous avons tenté pour chaque exercice de les paramétrer en fonction de critères variés. Nous détaillerons le paramétrage lorsque nous passerons en revue les différents exercices proposés.

3.3 Pertinence

Le dictionnaire est la source des données sur lesquelles sont basés les exercices. Après correction, il sert également à alimenter le retour d'information (feedback) qui est proposé à l'apprenant. Ce retour est également généré de façon automatique en synthétisant les informations contenues dans le dictionnaire. Il a une fonction didactique importante en permettant à l'apprenant de découvrir des informations qui dépassent le cadre de la phrase de l'exercice.

La pertinence de ces informations est d'autant plus grande que l'on n'y a pas accès, ni dans les manuels d'apprentissage du français, ni lors de la simple consultation d'un dictionnaire papier ou électronique classique. En outre, ce retour rend superflue la rédaction de messages de remédiation.

Nous avons de bonnes raisons d'héberger l'environnement sur un serveur, qui est interrogeable à distance, plutôt que d'offrir les exercices sur un CD-ROM. En effet, le transfert des exercices et des réponses des apprenants par le web ajoute les deux avantages suivants à l'environnement:

3.4 Flexibilité

Le serveur qui héberge l'environnement récupère l'ensemble des réponses données par les apprenants. Chacun peut interroger la base de données à n'importe quel moment pour revenir sur des erreurs qu'il a commises pour n'importe quel exercice.

En outre, un suivi des résultats de chaque apprenant est également prévu sous la forme d'un tableau synoptique de ses scores pour tous les exercices.

Ces scores servent bien évidemment aussi au professeur qui veut savoir qui a fait quels exercices et quels sont les résultats obtenus. En outre, une synthèse des erreurs commises par l'ensemble des apprenants travaillant avec l'environnement lui offre un aperçu très détaillé des points forts et faibles de ses apprenants, avec la possibilité de cibler son enseignement sur les aspects qui semblent mal maîtrisés.

3.5 Individualisation

Comme on le verra par la suite, les exercices sont, dans la plupart des cas, proposés avec un double degré de difficulté (exercices sur les mots ou structures « simples » ou plus « difficiles »), afin de mieux s'adapter au niveau de l'apprenant. Le passage d'une catégorie à l'autre est prévu et dépend des résultats obtenus par l'étudiant à la session précédente (voir 4.).

L'environnement s'adapte donc automatiquement au niveau de l'apprenant et lui trace un parcours d'apprentissage sur mesure.

4. Les exercices d'Alfalex

Lorsque l'on accède au site, on est invité à s'identifier. Cette identification sert à personnaliser les réponses dans la base de données, et ainsi à retrouver par exemple les erreurs propres à chaque étudiant. Pour les étudiants de la K.U.Leuven, cette identification est authentifiée par un contrôle du numéro d'étudiant et du mot de passe dans la base de données centrale de la K.U.Leuven. Pour un utilisateur occasionnel, l'identifica-

tion est libre, voire optionnelle si l'on ne désire pas voir sauvegardées les erreurs au moment de quitter l'environnement.

La page d'accueil d'Alfalex est rédigée en français et en néerlandais, en deux colonnes parce que nous voulons simplifier au maximum l'accès à l'environnement (écran 1).

écran
1



La page se décompose en trois grandes parties. La première, sous le titre *Test*, propose un lien vers un test de closure. Ce type de test est reconnu comme indice fiable du niveau de connaissance d'une langue. Il se compose de quatre petits textes où, pour un mot sur deux, la moitié des lettres finales du mot ou la moitié plus une sont à compléter. En tout, il y a 100 cases à remplir.

Si vous voulez que l'environnement prenne en charge votre parcours d'apprentissage, vous êtes invité à faire l'exercice. En fonction du résultat obtenu (plus ou moins de 70/100)⁵, on se verra proposé automatiquement des exercices plus simples ou plus difficiles. Cela vaut uniquement pour la première série d'exercices que l'on fait. En effet, par la suite, le niveau de difficulté de chaque exercice proposé est déterminé par l'ordinateur en fonction du score obtenu lors de la session précédente. Nous avons placé provisoirement la césure à 80 % pour passer au niveau supérieur et à 50 % pour rétrograder.

On accède aux exercices par les liens dans la deuxième partie de l'écran d'accueil, qui fait l'inventaire de tous les exercices disponibles, comme sous 3.1.

La troisième partie de l'écran détaille quelques fonctionnalités supplémentaires sur lesquelles nous reviendrons sous 5.

⁵ Cette césure a été déduite de l'observation des résultats obtenus par deux groupes d'étudiants qui ont mis à l'essai l'environnement (voir 6.1).

Dans la suite, nous passerons en revue les dix types d'exercices proposés. Pour chaque exercice, trois écrans se suivent :

écran Présentation de l'exercice (voir par exemple écran 2) :

- informations pratiques sur l'exercice (par exemple le nombre de phrases à faire)
- accès aux exercices avec informations concernant le degré de difficulté de l'exercice
- retour sur erreurs précédentes et sur phrases sélectionnées
- mode de calcul du résultat

écran Exercice (voir par exemple écran 3) :

- instruction
- exercice
- bouton d'envoi ou d'effacement des réponses

écran Correction (voir par exemple écran 5) :

- tableau de correction avec réponse exacte
correction et, éventuellement, réponse fautive donnée
retour d'information du Dafles
option de sauvegarde des phrases
dans le dictionnaire personnel
pour une session d'exercices ultérieure
- score obtenu

Pour tous les exercices, l'apprenant a des cases à remplir. Cela est sans conteste moins attrayant que ce que proposent beaucoup de sites (voir 2.). Le fait d'obliger systématiquement l'apprenant à taper les formes en entier nous semble se défendre du point de vue didactique, puisque nous tenons à sensibiliser l'apprenant à l'orthographe. Le fait de proposer uniquement des cases à compléter s'explique aussi par le mode de correction standardisé pour tous les exercices.

Une procédure de choix aléatoire règle l'apparition des phrases ou des formes. Chaque exercice créé à l'écran est donc différent des autres et, dans de nombreux cas, l'on ne retrouve pas les mêmes formes ou les mêmes phrases lorsque l'on refait l'exercice.

4.1 Les aspects formels du mot : la morphologie

Pour l'exercice sur la morphologie, nous n'avons retenu que les noms et adjectifs dont le modèle de flexion était irrégulier. On ne trouvera donc pas dans l'exercice un nom comme *exercice*, qui forme son pluriel par l'adjonction régulière d'un -s; contrairement à *travail*, qui comme *bail*, a un pluriel en -aux.

Ces modèles de flexion, limités à quelques dizaines pour les noms et les adjectifs, réduisent le nombre de cas de figure à proposer dans les exercices.

Nous avons ajouté sur l'écran d'accueil de l'exercice, là où cela était possible, un lien vers un site externe où l'apprenant trouve un aperçu synoptique d'informations concernant l'objet de l'exercice, la formation du féminin et du pluriel dans ce cas-ci (écran 2).

écran
2

Exercice de morphologie

nombre de phrases	10
type d'énoncé	Phrases avec terminaisons de noms et d'adjectifs à compléter
informations complémentaires	Rappel de la formation des formes des adjectifs et des noms (phrases)

exercice

[mots fréquents]
[mots moins fréquents]
[sans indice de fréquence]

retour sur exercices précédents

[exercices précédents] Tous les noms et adjectifs pour lesquels les terminaisons n'étaient pas corrects, proposés dans une nouvelle phrase
[noms/adjectifs à reposer] Tous les noms/adjectifs que vous avez demandé à reposer sous la forme d'un exercice

calcul du score

2 points : réponse correcte
1 point : forme correcte de l'adjectif, sans autres incorrect
0 point : mauvaise réponse
un mot donné (la forme que vous devez donner est nécessairement différente de celle donnée en fin de phrase)

accueil	exercices	morphologie	conjugaison	dictionnaire	grammaire	prépositions
accueil	exercices	exercices	exercices	exercices	exercices	exercices
accueil	exercices	exercices	exercices	exercices	exercices	exercices

L'exercice proposé est un exercice en contexte : le mot à forme irrégulière est placé dans un environnement naturel. Vu le niveau de nos étudiants, il nous a paru plus sensé de les confronter le plus possible à la langue réelle (écran 3).

écran
3

Exercices de morphologie

*** Compléter les phrases en accordant les mots en italiques en fin de phrases ***
La forme à compléter est nécessairement différente du mot donné en fin de phrases

1 Selon Durand, les organisateurs de la cérémonie ont renoncé à la *derrière* minute à montrer au public une séquence commandée aux auteurs des Guignols.
{*derrière*}

2 Malgré les efforts réalisés des producteurs, conseillés par une *journaliste* de CBS qui venait de couvrir un autre cirque mécatique le procès Simpson, Mad City est plus sage que la *folle* réalité des médias électroniques.
{*folle*}

3 Les enquêteurs n'ont pas pu trouver sur les *lieux* de perquisition tous les individus recherchés.
{*lieux*}

4 Les annonceurs *internationaux* représentent 20 % des recettes publicitaires de L'Express et de L'Expansion.
{*internationaux*}

5 Dans l'attente des réponses de Washington, on estime, à la Commission européenne, que la situation est suffisamment grave pour que soit interdite toute importation vers l'Europe de viandes *fraîches* et de volailles en provenance des États-Unis.
{*fraîches*}

6 Sous le logo international un tournesol, le logo du Mercantour vient de tenir de *longues* semaines la vedette du courrier.

accueil	exercices	morphologie	conjugaison	dictionnaire	grammaire	prépositions
accueil	exercices	exercices	exercices	exercices	exercices	exercices
accueil	exercices	exercices	exercices	exercices	exercices	exercices

On obtient la correction de l'exercice en cliquant sur un bouton Envoyer à la fin de celui-ci. Dans le cas de la morphologie, la correction est sensible aux erreurs de nombre, qui sont pénalisées moins sévèrement qu'une erreur due à l'ignorance de la forme irrégulière du mot.

Le retour d'information du Dafles se présente sous la forme d'un lien qui pointe vers une liste de mots ayant le même modèle de flexion. Chaque mot de cette liste est accompagné d'un lien qui renvoie l'apprenant vers le texte du Dafles (voir écran 5).

4.2 Les aspects formels du mot : les conjugaisons

Comme pour la morphologie, il nous a semblé nécessaire de paramétrer l'exercice plutôt que de puiser au hasard dans le stock de formes verbales.⁶ Pour ce faire, nous avons assigné à tous les verbes un code de classe, repris au Bescherelle. Ensuite, nous avons fait le total par classe des fréquences d'apparition des verbes dans notre corpus. Les verbes des classes avec 50 000 occurrences ou plus sont étiquetés « fréquents » ; les autres se retrouvent dans la catégorie « peu fréquents ». Dans la première, on retrouve, par ordre de fréquence décroissante : *être*, *avoir*, *tenir* (22 verbes), *faire* (5 verbes), *pouvoir*, *-{an/en/on/er/or}dre* (28 verbes), *-cer* (45 verbes), etc. ; dans la seconde : *-e.er* (23 verbes), *-uire* (18 verbes), *-tir* (12 verbes), *falloir*, *vouloir*, *-{f/v}rir* (7 verbes), etc. L'exercice porte sur plus de 1 100 verbes identifiés parmi les 5 000 lemmes les plus fréquents de notre corpus.

En outre, pour chaque classe de verbes, nous n'avons retenu que les formes qui nous paraissaient déterminantes. Pour *être* et *avoir*, pratiquement tout le paradigme a été retenu. Par contre, pour les verbes en *-cer*, nous avons préférentiellement sélectionné les cas où apparaît le *ç*.

L'exercice sur les conjugaisons est un exercice hors contexte, semblable à celui présenté sous 2. Contrairement au site mentionné dans cette section, il n'est pas encore possible à l'heure actuelle de s'exercer sur un verbe ou sur un mode ou temps en particulier.

Lors de la correction, le retour d'information du Dafles est double. On a tout d'abord la possibilité de voir la conjugaison entière du verbe en question et non pas simplement celle d'un représentant type de la classe. En plus, on peut également obtenir la liste de tous les verbes qui se conjuguent sur le même modèle.

4.3 Les aspects formels du mot : la dérivation

Même si le français connaît certaines régularités dans le domaine de la dérivation (les mots en *-tion* identifient souvent le nom d'action : *consommation*, *production*, etc.), on ne peut pas prévoir, dans de nombreux cas, ni le sens d'un mot à partir de sa terminaison (*imagination* n'est pas le nom d'action du verbe *imaginer*), ni la façon dont il va s'écrire (*travail* exprime l'action du verbe *travailler*). Un exercice sur la dérivation semble donc avoir un sens.

Dans l'exercice, tel qu'il est proposé aux apprenants, Alfalex choisit au hasard une phrase par dérivé d'un verbe. Les dérivés potentiels ne sont pas donnés (écran 4). Dans

⁶ Nous tenons à remercier Piet Mertens qui nous a fourni les formes verbales des verbes contenus dans notre nomenclature, générées automatiquement à l'aide du logiciel Morlex (<http://bach.arts.kuleuven.ac.be/pmertens/morlex/>).

Dans le deuxième groupe de mots, dont une bonne partie se terminent en *-e*, nous avons distingué ceux qui nous paraissaient être des sources d'erreurs (*problème, système, etc.*) des autres (*masse, règle, etc.*). Ces deux sous-classes de noms apparaissent dans les exercices dans une proportion de 3 contre 1.

Comme nous l'illustrerons dans ce qui suit, la part d'arbitraire dans la sélection des mots sources d'erreurs ou non, pourra être corrigée au fur et à mesure de la récupération des réponses données par les apprenants.

Lors de la correction, nous attirons systématiquement l'attention de l'apprenant sur l'existence des classes de noms, dont la terminaison détermine le genre, afin qu'il en prenne conscience et qu'il puisse évacuer ainsi une bonne part des problèmes liés au genre en français. En cela, les liens qui apparaissent sur la page de correction constituent de précieux outils d'apprentissage, puisqu'ils récapitulent les principaux noms qui suivent la « règle », tout en mentionnant les exceptions (écran 5).

écran
5

réponse correcte	votre réponse	renvoi vers le Dafles	nom à reproposer sous la forme d'un exercice ?	nom à ajouter au dico personnel ?
QUERELLE SURRÉALISTE* La tentative de constitution de la liste RPR-UDF a tourné court après l'irruption sur la scène insulaire de Charles Pasqua.	réponse juste		<input type="checkbox"/>	<input type="checkbox"/>
L'édifice, qui doit mesurer 200 mètres de diamètre et atteindre une hauteur de 135 mètres, coûtera 850 millions de dollars (environ 5,185 milliards de francs)	réponse juste		<input type="checkbox"/>	<input type="checkbox"/>
L'équipe de France de football a inauguré "son" stade par une victoire	réponse juste		<input type="checkbox"/>	<input type="checkbox"/>
Mais ce nouvel hommage à l'équipe de France ne doit pas faire oublier le début de saison difficile que connaissent les tricolores.	réponse juste	Genre des mots se terminant par -age	<input type="checkbox"/>	<input type="checkbox"/>
Dans son prolongement exact, la petite salle fait face, via la scène, à sa grande soeur.	réponse juste	Genre des mots se terminant par -ite	<input type="checkbox"/>	<input type="checkbox"/>
Alors qu'il était aux commandes depuis plus d'un an, reconduisant son titre de monde sur longues			<input type="checkbox"/>	<input type="checkbox"/>

Aide
LES mots se terminant par **-ite** sont :
féminins : salle, bataille, bataille, table, capitale, grande, grande, école, table, grande, etc.
Exceptions : centrale, table, portable, grande, etc.

Comme toujours, dans le retour d'information, chaque mot constitue un lien qui mène l'étudiant vers le texte du Dafles, où il peut découvrir le sens de mots qu'il ne connaît pas.

4.5 La combinatoire des mots : l'emploi des prépositions

Il s'agit d'un des classiques de l'apprentissage du français : *espérer faire quelque chose, fort en quelque chose, une dette envers quelqu'un*. Un des intérêts de l'exercice est que l'on est confronté de temps en temps à des constructions sans préposition. Près de 350 constructions différentes sont déjà intégrées dans la base des exercices.

Comme pour la plupart des autres exercices, la construction est également intégrée dans une phrase authentique.

Le retour d'informations du Daflex consiste en un aperçu des autres mots de la base qui se construisent avec la même préposition (écran 6).

écran
6

réponse correcte	votre réponse	renvoi vers le Daflex	nom à reproposer sous la forme d'un exercice ?	nom à ajouter au dico personnel ?
QUERELLE SURRÉALISTE* La tentative de constitution de la liste RPR-UDF a tourné court après l'irruption sur la scène insultaire de Charles Pasqua.	réponse juste		<input type="checkbox"/>	<input type="checkbox"/>
L'édifice, qui doit mesurer 200 mètres de diamètre et atteindre une hauteur de 135 mètres, coûtera 850 millions de dollars (environ 5,185 milliards de francs)	réponse juste		<input type="checkbox"/>	<input type="checkbox"/>
L'équipe de France de football a inauguré "son" stade par une victoire	réponse juste		<input type="checkbox"/>	<input type="checkbox"/>
Mais ce nouvel hommage à l'équipe de France ne doit pas faire oublier le début de saison difficile que connaissent les tricolores.	réponse juste	Genre des mots se terminant par -age	<input type="checkbox"/>	<input type="checkbox"/>
Dans son prolongement exact, la petite salle fait face, via la scène, à sa grande soeur.	réponse juste	Genre des mots se terminant par -lie	<input type="checkbox"/>	<input type="checkbox"/>
Alors qu'il était aux depuis plus d'un an, reconduit son titre d' monde sur longues			<input type="checkbox"/>	<input type="checkbox"/>

Aide
Les mots se terminant par -lie sont :
féminisme
salle, parole, tournée, table, capitale,
grande, grande, école, salle, famille, lie,
salle, maternité, etc.
Exceptions : contrainte, salle, portabale,
contrainte, salle, portabale

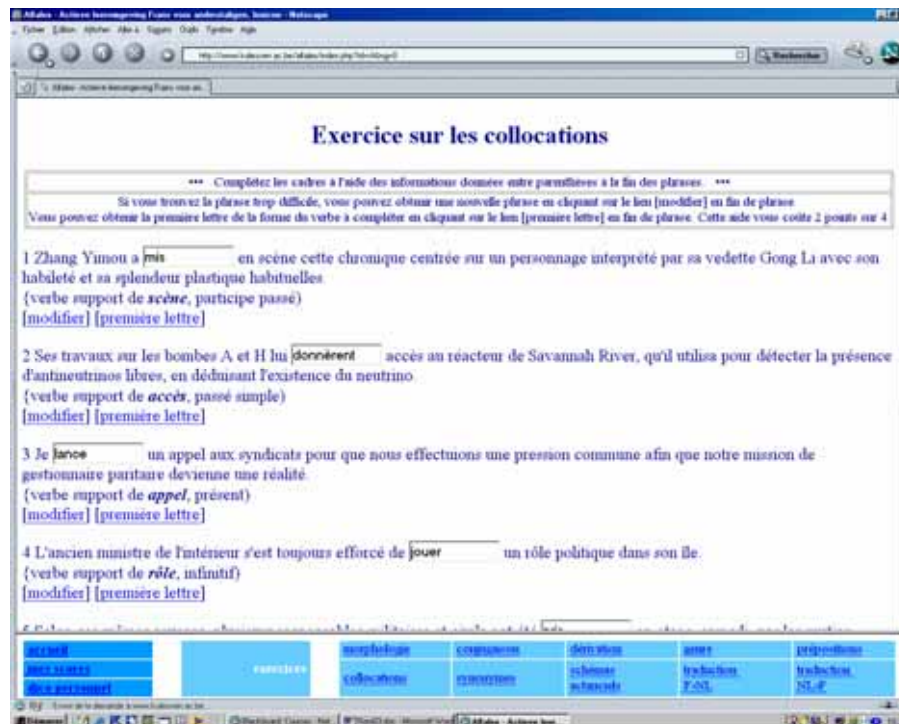
4.6 La combinatoire des mots : les collocations

Pour l'instant, cet exercice n'intègre que quelque 300 constructions à verbe support (*poser une question* = questionner ; *une ambiance règne* = il y a une ambiance).⁷ Vu la difficulté de l'exercice, surtout pour des collocations peu fréquentes⁸, nous donnons à l'apprenant la possibilité de modifier la phrase et d'appeler la première lettre de la forme verbale qu'il doit remplir. Ces deux aides sont possibles au prix d'une (légère) pénalisation lors de la correction (écran 7).

⁷ Un exercice analogue pourrait par exemple être conçu pour des intensificateurs : *de vives critiques*, *largement insuffisant*, etc.

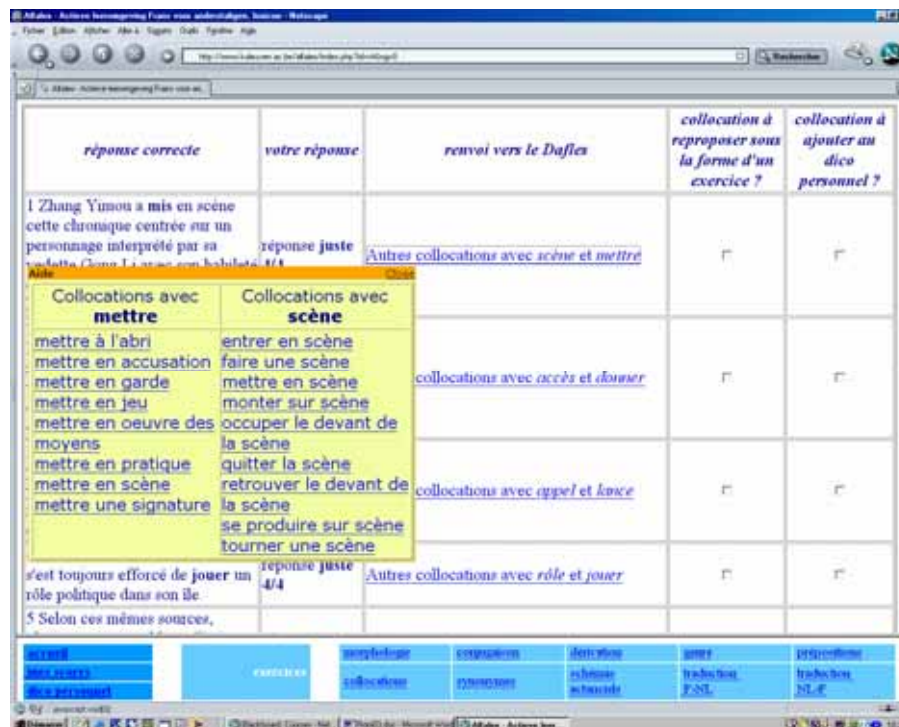
⁸ Ici aussi, la fréquence est tirée de notre corpus. À notre connaissance, il s'agit du premier comptage de fréquence effectué sur des combinaisons de mots pour le français. Ce comptage met en évidence la fréquence relativement élevée de ces combinaisons de mots et donc leur importance pour des apprenants de niveau intermédiaire et supérieur.

écran
7



Après correction, le retour d'information est composé d'une liste synoptique de collocations construites avec les deux mots en question (écran 8).

écran
8



En cliquant sur l'une de ces collocations, l'apprenant tombe à l'endroit exact du Dafles où elle est traitée.

4.7 Les champs sémantiques : les synonymes

Dans le Dafles, de très fines relations sont tissées entre les sens des mots : relations de synonymie, d'antonymie exclusive (*mort* <> *vivant*), d'antonymie gradée (*chaud* <> *tiède* <> *froid*), d'hyponymie (*construction* > *bâtiment* > *maison*), etc. Les mots qui entrent dans ces relations peuvent faire l'objet d'exercices visant à élargir le stock lexical de l'apprenant. Pour l'instant, seul un exercice sur les synonymes est disponible, et pour un nombre limité de mots, compte tenu de l'état d'avancement de la description de ces relations dans la base de données. Dans cet exercice, l'apprenant remplit un mot synonyme du mot donné et dont le sens est illustré à l'aide d'une phrase exemple tirée du Dafles. Nous prévoyons de fournir à l'apprenant la traduction en néerlandais du mot donné afin de faciliter la recherche d'un synonyme.

4.8 Les champs sémantiques : les schémas actanciels

Un schéma actanciel est un tableau avec comme colonne centrale un verbe, avec autant de lignes que le verbe a de sens. Le verbe étant le pivot de la phrase, c'est lui qui détermine l'apparition et la fonction des compléments essentiels : *briller*, intransitif, avec comme seul complément un sujet ; *donner* avec trois compléments : *quelqu'un*, *quelque chose*, à *quelqu'un*. Le verbe détermine donc la structure formelle de la phrase. À cette couche syntaxique vient s'ajouter une couche sémantique : chaque verbe détermine aussi le sens des compléments qui l'accompagnent. Le caractère de ces restrictions est peu contraignant (le sujet du verbe *travailler* dans son sens premier peut être n'importe quelle personne) ou au contraire très contraignant (pour le verbe *se jeter*, dans un de ses sens, on trouve la dénomination d'un cours d'eau comme sujet et d'une vaste étendue d'eau comme complément prépositionnel introduit par *dans*).

Un schéma actanciel est une tentative de visualisation de ces informations, en présentant, aux différentes positions autour du verbe (sujet, compléments essentiels, mais aussi compléments circonstanciels [lieu, temps, moyen]), les noms ou collocations prototypiques, et en donnant également les noms de résultat et d'action correspondants. En d'autres termes, il s'agit en quelque sorte des mots qui viendraient en premier à l'esprit lorsque l'on va former une phrase avec ce verbe, utilisé dans ce sens particulier.

Ces schémas actanciels constituent une source d'exercices intéressants, parce qu'ils forcent à concevoir le vocabulaire non comme un amalgame de faits isolés, mais comme des ensembles structurés d'éléments qui se tiennent et s'appellent dans une situation de communication, ce qui facilite la mémorisation.

La copie d'écran suivante montre comment est conçu l'exercice, ainsi que l'aide qui est fournie sous forme d'un petit écran en surimpression qui apparaît dès que l'on passe la souris sur l'une des cases à compléter (écran 9).

écran
9

Exercice sur les schémas actanciels

*** Complétez les cadres à l'aide de noms et d'adjectifs ***

Mots utilisés avec **brûler** :

	action	N1sujet	verbe	N2	prép N3	Nlien	Nrésultat	Nnoyau
1a	être attaqué par le feu	qqch	brûler				un e	le feu
1b	se faire une blessure par le feu	qqq => un brûlé	se brûler			une partie du corp	Aide: ... est un grand feu non contrôlé qui affecte un bâtiment, une forêt, un véhicule.	une brûlure
1c	mettre le feu à quelque chose	qqq	brûler	qqch				le feu
2	éprouver un sentiment très fort	qqq	brûler	un e un e				
3	ne pas s'arrêter	un e un e	brûler	un feu un stop		un e un e	un e	
4	déclarer une agression de violence	une partie du corps est brûlée	brûler					une inflammation la e

Mots utilisés dans le **français** :

accueil	connexion	morphologie	conjugaison	dérivation	sens	traduction
accueil	connexion	collocations	expressions	schémas syntaxiques	traduction F->NL	traduction NL->F

On notera que tous les dérivés, que l'on a vu apparaître dans l'exercice sur la dérivation (4.3), servent à compléter le tableau et fournissent en même temps une aide complémentaire lors de la recherche des mots manquants. Le Dafles comporte actuellement les données nécessaires pour créer plus de 200 de ces tableaux.

À la correction, l'environnement affiche les tableaux entiers. Un lien sur le verbe en haut du tableau mène l'apprenant vers le Dafles, où tous les mots du tableau sont munis d'un lien qui permet d'afficher leur sens, la structure syntaxique du mot, une phrase exemple et les synonymes et antonymes éventuels.

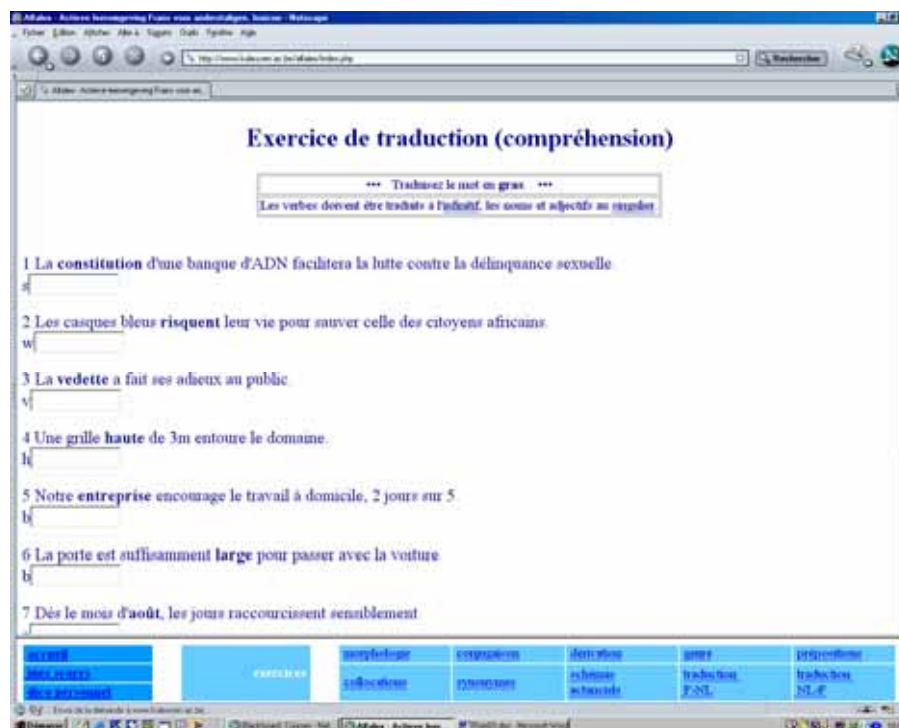
4.9 La traduction

Le Dafles est un dictionnaire semi-multilingue : il fournit, pour tous les emplois répertoriés du mot et pour chaque sens, une traduction vers la langue maternelle de l'étudiant. Contrairement aux autres dictionnaires semi-bilingues existants, les traductions ne sont pas simplement reliées aux sens, mais bien aux différentes structures syntaxiques décrites pour chaque sens, ce qui augmente la pertinence et l'exactitude de celles-ci. Ainsi, les deux constructions différentes du verbe *augmenter* au sens de « accroître » donnent donc une traduction différente : X augmente Y (*doen stijgen*) et Y augmente (*stijgen*). Toutes les traductions, répertoriées dans un index, facilitent l'accès au texte du dictionnaire. Un dictionnaire semi-bilingue combine donc dans une certaine mesure les avantages des dictionnaires monolingues, riches en sémantique et en contextes, et des dictionnaires bilingues, riches en correspondances entre deux langues, mais sans contextes.

Sur base de ces données, un double exercice de traduction français - néerlandais (compréhension) et néerlandais - français (production) est réalisé en utilisant non pas des

phrases sorties du corpus, mais les phrases exemples du Dafles (écran 10).

écran
10



La correction de l'exercice de traduction vers le français est sensible aux erreurs de forme, puisqu'une forme incorrecte du mot (pluriel au lieu de singulier par exemple) est moins lourdement pénalisée qu'un mot fautif.

4.10 Retour sur les exercices

Comme nous l'avons signalé, l'environnement gère et sauvegarde toutes les réponses. L'exploitation de toutes ces données se fait de deux façons différentes. Tout d'abord, il y a les erreurs commises. Celles-ci sont identifiées par Alfalex et l'apprenant se voit proposer à nouveau ces cas, sous la forme d'un exercice qu'il peut choisir de faire quand il le veut et auquel il accède à partir d'un lien sur l'écran de présentation de l'exercice (écran 2). Le travail sur les erreurs se fait indépendamment des exercices réguliers et peut être poursuivi jusqu'à épuisement de celles-ci. Dans la mesure du possible, les cas sont intégrés dans une nouvelle phrase.

Outre les erreurs, il y a des cas sur lesquels, malgré une réponse correcte, l'apprenant aimerait revenir plus tard parce qu'il sent qu'il continue à avoir des doutes, ou qu'il a simplement donné la bonne réponse par chance. Pour ces cas, deux fonctionnalités particulières ont été prévues sur l'écran de correction. Ainsi, l'apprenant a la possibilité de choisir de placer une ou plusieurs phrases dans son dictionnaire personnel, auquel il a accès à n'importe quel moment par un lien sur la barre de navigation. D'autre part, il peut également se constituer un exercice personnalisé avec des phrases sélectionnées, accessible à partir d'un lien sur l'écran de présentation (écran 2).

5. Fonctionnalités annexes

5.1 Analyse des erreurs

Les réponses données aux exercices constituent des traces dans une table de la base de données Alfalex. Il est évident que les réponses (fautives et exactes) ainsi recueillies présentent également un intérêt pour le concepteur de l'environnement/le professeur. En effet, comme l'illustre l'écran 11, il est possible d'obtenir, pour chacun des dix exercices proposés, une liste des formes ou des (combinaisons de) mots apparus dans les exercices soumis aux apprenants, avec le nombre d'erreurs commises, le nombre total d'occurrences et un calcul du pourcentage d'erreurs pour chaque cas isolé.

écran
11



The screenshot shows a web browser window with the title 'Statistiques sur les erreurs'. Below the title, there are several tabs for different error categories: 'morphologie-conjugaison', 'dérivation', 'genre', 'prépositions', 'collocations', 'synonymes', 'schémas actanciels', 'traductions F->NL', and 'traductions NL->F'. The 'genre' tab is selected. Below the tabs, it says 'nombre d'informations : 481'. A table with the following columns is displayed: 'numinfo', 'information', 'exercice', 'nb. err.', 'total occ.', and 'rapport'.

numinfo	information	exercice	nb. err.	total occ.	rapport
1182	espace	genre	13	13	100.00
794	caractère	genre	8	8	100.00
2263	véhicule	genre	5	5	100.00
1269	forêt	genre	5	5	100.00
1143	emploi	genre	4	4	100.00
1764	patrimoine	genre	3	3	100.00
1096	drame	genre	3	3	100.00
538	aure	genre	3	3	100.00
633	athlétisme	genre	3	3	100.00
1817	port	genre	3	3	100.00
2329	vote	genre	10	11	90.91
1719	offre	genre	7	8	87.50
2241	univers	genre	7	8	87.50
1238	favueur	genre	6	7	85.71
863	chiffre	genre	5	6	83.33
915	commerce	genre	9	11	81.82

La liste ne tient pas compte des retours sur les erreurs ni des exercices personnalisés créés (voir 4.10) et constitue donc un véritable hit-parade des erreurs courantes, ou, inversement, un hit-parade des cas qui ne posent pas problème. Plus le nombre d'utilisateurs d'Alfalex augmentera, plus le nombre de cas soumis sera important et plus la pertinence de cette liste augmentera.

L'analyse de ces erreurs peut encore être poussée plus loin. Un lien sur le nombre total des occurrences donne en effet accès à la liste complète des réponses fautives données. On constate certaines régularités (prévisibles), comme la difficulté de la formation du féminin des noms en *-teur* (*-trice* ou *-teuse*), ainsi que des cas isolés : le pluriel de *bail*, malgré le parallélisme frappant avec *travail*, **épaïse* au lieu de *épaisse*, etc.

Une fois analysées, ces informations serviront à optimiser l'enseignement du vocabulaire dans tous ses aspects. On peut par exemple concevoir de petits aide-mémoire très ciblés, dispersés judicieusement dans un manuel ou dans les notes d'un cours avec une

récapitulation des problèmes inventoriés. Ou encore un ou plusieurs cours, avec exercices, pour travailler certains aspects moins connus du vocabulaire. L'avantage sera évidemment que l'on traitera non pas ce que l'on croit être problématique, mais bien ce qui l'est réellement.

À terme, on peut également concevoir que ces informations permettront d'optimiser l'environnement. Dans l'exercice sur le genre, par exemple, nous avons catalogué toute une série de noms se terminant en *-e* de « difficiles » sur base de notre intuition. Elle s'avèrera peut-être erronée pour certains mots, auxquels il faudra enlever l'étiquette au profit d'autres, considérés à tort comme ne posant pas de difficultés.

5.2 Tableau synoptique des résultats

L'environnement procède à une comptabilisation des résultats par utilisateur, accessible par un lien sur la barre de navigation (écran 12).

écran
12

**Résultats pour Serge Verlinde
Instituut Levende Talen (ILT)**

mois	type d'exercice			
date	pourcentage de réponses exactes		nombre de phrases reçues	
	mots fréquents mots moins fréquents		(mots fréquents mots moins fréquents)	
erreurs	nombre d'erreurs qu'il vous reste à refaire pour chaque type d'exercice			

Accès depuis Blackboard

mois	genre	collocation	morphologie	synonymes	dérivés	traductions F.->NL	traductions NL->F	prépositions	conjugaison	schémas actants
Fév 2001-47/48 (9/12/1)			100/0% (10/10)		30/0% (2/0)			80/0% (10/0)	100/0% (10/0)	
erreurs	122	14	2		27	4	1	2	46	41

Dans le tableau synoptique figurent les scores pour les différents exercices par niveau de difficulté. Ces scores sont calculés uniquement pour les premiers exercices, non pour les retours sur des cas précédents (erreurs ou cas sélectionnés), sinon, ici aussi, le score ne reflèterait pas le véritable niveau de l'apprenant.

Le nombre d'erreurs qui restent est comptabilisé pour chaque exercice séparément.

Ces informations sont également disponibles pour le professeur, qui dispose ainsi d'un moyen de contrôle pour voir le nombre d'exercices résolus par les différents apprenants et leur niveau global.

Jusqu'à présent, mises à part les utilisations en interne, l'environnement a déjà été testé une première fois, dans une version réduite, auprès d'une vingtaine d'étudiants qui suivaient un cours d'été de français à la Katholieke Universiteit Brussel (KUB) au mois de septembre 2002.

Depuis lors, deux groupes d'étudiants de licence des facultés de sciences de communication et de sciences appliquées, respectivement une soixantaine et une vingtaine d'étudiants, ont soumis l'environnement à un examen critique. Leurs suggestions nous ont permis d'optimiser plusieurs aspects de l'environnement.

En outre, grâce à leurs résultats, nous avons pu faire quelques constatations intéressantes. Nous nous sommes aperçus tout d'abord que la corrélation entre le résultat pour le test initial et la moyenne obtenue pour l'ensemble des résultats est acceptable : elle tourne autour de 0,60. C'est-à-dire qu'un étudiant qui obtient un bon résultat sur le test obtient en règle générale un bon résultat moyen sur l'ensemble des exercices.

Par contre, cette corrélation disparaît complètement lorsque l'on met en rapport les résultats au test et les résultats pour chacun des dix exercices proposés : un (moins) bon niveau de langue n'implique pas nécessairement un (moins) bon niveau de langue pour chacun des aspects du vocabulaire testés. Un étudiant qui a obtenu un faible score au test peut très bien obtenir un excellent score sur la conjugaison ou la morphologie. De là la nécessité d'un parcours d'apprentissage aussi individualisé que possible. Dans Alfalex, nous avons tenté d'y répondre en permettant le passage d'exercices faciles à des exercices difficiles, et inversement, en fonction du score obtenu (voir ci-dessus). La pertinence de la distinction entre les deux niveaux de difficulté, basée en grande partie sur des critères de fréquence, est confirmée par le fait qu'en moyenne, le score des deux groupes d'étudiants mentionnés ci-dessus a baissé de plus de 15 % pour les exercices difficiles.

Si l'on compare les moyennes obtenues pour les différents exercices, on s'aperçoit que le genre, la morphologie et la conjugaison sont les domaines pour lesquels les scores sont les meilleurs (>70 % pour les exercices faciles, >55 % pour les exercices difficiles). Les collocations, les synonymes et les exercices de traduction s'avèrent les plus difficiles, avec des moyennes sous les 50 % pour les exercices difficiles. Cela s'explique sans aucun doute par le fait qu'il n'y a pas d'enseignement systématique prévu pour ces aspects, contrairement par exemple à la conjugaison des verbes.

Depuis le mois de février 2003, quelque 350 étudiants de première candidature en droit utilisent l'environnement afin de tester les dernières modifications apportées à celui-ci et la robustesse de l'ensemble.

À partir de septembre 2003, l'environnement sera aussi mis à la disposition des autres étudiants de la K.U.Leuven, de la KULAK et du personnel par le biais de la plate-forme d'apprentissage électronique Blackboard.

6.2 Intégration de l'environnement dans l'enseignement/l'apprentissage

L'environnement est avant tout conçu pour fonctionner dans un contexte d'apprentissage semi-autonome où l'apprenant prend en charge la mise à niveau de ses connaissances dans le domaine du vocabulaire. L'assistance est livrée sous la forme d'informations tirées du Dafles, le dictionnaire d'apprentissage du français lié à Alfalex. Les traces laissées par les utilisateurs facilitent éventuellement un suivi détaillé de leur travail et peuvent, en retour, alimenter le contenu des cours.

7. Conclusion

Comparé aux environnements d'apprentissage (du vocabulaire) existants, Alfalex a l'avantage d'offrir une couverture très complète des différents aspects du vocabulaire le plus fréquent du français, grâce au fait qu'il se fonde sur la description lexicographique d'un dictionnaire d'apprentissage (électronique) du français langue étrangère ou seconde. Les phrases proposées dans la plupart des exercices sont tirées (semi-) automatiquement d'un important corpus de phrases, ce qui garantit leur authenticité et un apport constant de nouvelles phrases.

La correction des exercices est très sensible, puisqu'elle est paramétrée de façon à reconnaître des réponses possibles ou partielles et à tenir compte d'éventuelles aides demandées par l'apprenant au cours de l'exercice.

Le retour d'information après correction est particulièrement riche, puisqu'il propose une synthèse d'informations pertinentes puisées dans le Dafles.

Alfalex offre en plus un dispositif très performant de retour sur les exercices précédents. Ce retour se fait automatiquement pour les erreurs. Pour les réponses correctes, l'apprenant a le choix de les renvoyer à un exercice ultérieur, de les enregistrer dans son dictionnaire personnel ou de les écarter définitivement.

Avec Alfalex nous avons tenté d'explorer toutes les possibilités offertes actuellement par l'outil informatique dans la création d'un environnement d'apprentissage du vocabulaire. Le résultat est un environnement dynamique et interactif particulièrement performant permettant de contribuer à un véritable auto-apprentissage.

Post-scriptum

Les différents aspects du vocabulaire qui font l'objet d'exercices dans Alfalex et la distinction de deux niveaux de difficulté rendent l'environnement adaptable à de nombreux publics. Nous pensons concrètement à des centres de formation en langue, à l'enseignement supérieur ou au dernier cycle de l'enseignement secondaire, où il pourrait facilement servir d'outil de remédiation en dehors des cours réguliers.

Nous aimerions pouvoir tester l'environnement avec des publics autres que le nôtre, et en tirer les enseignements nécessaires. Pour cette raison, si cela vous tente, n'hésitez pas à nous contacter (serge.verlinde@ilt.kuleuven.ac.be ou 016 32 56 85) pour déterminer les modalités pratiques de l'utilisation gratuite par vos élèves/étudiants d'une version « light » de l'environnement Alfalex. Par « light », nous entendons l'environnement avec tous ses exercices dans leur état actuel, avec, pour chaque apprenant, un aperçu de ses scores et un dictionnaire personnel. En contrepartie, nous aimerions bien évidemment recueillir vos réactions et les éventuels problèmes ou erreurs qui seraient apparus afin de perfectionner l'environnement.

Références

(Le nouveau) Bescherelle. *1. L'art de conjuguer. Dictionnaire de 12 000 verbes*. Paris : Didier-Hatier.
Selva, T. 1999. *Ressources et activités pédagogiques dans un environnement informatique d'aide à l'apprentissage lexical du français langue seconde*. Thèse de doctorat non publiée. UFR des Sciences et techniques de l'Université de Franche-Comté (Besançon).

- Selva, T., S. Verlinde et J. Binon. (à paraître). Vers une deuxième génération de dictionnaires électroniques. Numéro spécial de la revue *TAL* consacré aux dictionnaires électroniques.
- Verlinde, S. et T. Selva. 2001. Nomenclature de dictionnaire et analyse de corpus. *Cahiers de lexicologie* 79.2. 113-139.
- Verlinde, S., T. Selva et J. Binon. 2002. Dictionnaires électroniques et apprentissage du vocabulaire. In J. Binon et al. *Tableaux vivants. Opstellen over taal-en-onderwijs, aangeboden aan Mark Debrock*. Leuven : Universitaire Pers. 179-199.