

## Génération automatique d'exercices contextuels de vocabulaire

Thierry Selva

ILT – K.U. Leuven  
Dekenstraat 6  
3000 Leuven, Belgique  
thierry.selva@ilt.kuleuven.ac.be

### Résumé – Abstract

Cet article explore l'utilisation de ressources lexicales et textuelles ainsi que d'outils issus du TAL dans le domaine de l'apprentissage des langues assisté par ordinateur (ALAO). Il aborde le problème de la génération automatique ou semi-automatique d'exercices contextuels de vocabulaire à partir d'un corpus de textes et de données lexicales au moyen d'un étiqueteur et d'un parseur. Sont étudiées les caractéristiques et les limites de ces exercices.

This paper examines the use of lexical and textual resources and NLP tools in Computer-Assisted Language Learning (CALL). It tackles the issue of automatic or semi-automatic generation of contextual vocabulary exercises from a corpus of texts and lexical data by means of a tagger and a parser. It studies the characteristics and limits of these exercises.

### Keywords – Mots Clés

Exercices contextuels, lexique, corpus, ALAO.  
Contextual exercises, lexicon, corpus, CALL.

## 1 Introduction

Les activités pédagogiques concernant le vocabulaire sont désormais très nombreuses sur Internet (Jaser 2002, Didier 2002, La Passerelle 2002, etc., pour ne citer que ces sites). Le plus souvent, il s'agit d'exercices à trous ou de QCM, conçus à l'aide de systèmes auteur, un des plus connus étant *Hot Potatoes*. On peut aussi trouver des jeux sous forme de mots croisés ou bien des exercices d'appariement entre une image graphique et le nom de l'objet dessiné. Toutes ces activités ont un point commun : ce sont des transpositions à l'écran d'exercices conçus et rédigés à l'origine sur le papier. L'unique intérêt de l'ordinateur est l'interactivité

qu'il procure, c'est-à-dire qu'il compare les réponses données avec celles qu'il a en mémoire et attribue sur-le-champ un score.

L'objet de notre travail est d'étudier les exercices qu'il est possible de générer automatiquement à partir de ressources lexicales et textuelles (dictionnaire et corpus) et des outils TAL d'analyse textuelle (analyseur syntaxique et étiqueteur). Ce travail est réalisé dans le cadre du projet ALFALEX (*Active Leeromgeving Frans voor Anderstaligen Lexicon*, environnement d'apprentissage lexical interactif pour apprenants du français), en cours de développement à la K.U.Leuven. À côté du module de génération d'exercices que nous allons présenter, l'environnement comprend un dictionnaire électronique pour apprenants, le DAFLES<sup>1</sup> (Selva et al., 2002), et le corpus GRELEP<sup>2</sup> (Verlinde, Selva, 2001 et 2002).

Les exercices que nous cherchons à générer sont des exercices contextuels, c'est-à-dire des activités dans lesquelles le vocable<sup>3</sup> est entouré de son contexte (la phrase, dans le cadre de ce projet). 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. D'autre part, le contexte facilite la pratique d'inférences, une stratégie essentielle en apprentissage des langues.

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 celui 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 lexicales, lui permettent de retourner un diagnostic sur la réponse un peu plus élaboré que le classique *vrai* ou *faux*<sup>4</sup>. Par ailleurs, en dotant le système d'une « mémoire », en l'occurrence un dictionnaire personnalisé dans lequel sont consignées les annotations des apprenants et le détails de leurs résultats, on parvient à une meilleure gestion des énoncés en évitant, par exemple, de reposer une épreuve qui a déjà été résolue avec succès. Ainsi un tel environnement permet une plus grande individualisation de l'apprentissage.

Nous allons maintenant étudier les exercices obtenus en les classant en deux catégories : les exercices contextuels générés automatiquement et ceux générés semi-automatiquement.

---

<sup>1</sup> Dictionnaire d'Apprentissage du Français Langue Seconde ou Étrangère. La nomenclature prévue est d'environ 12 000 entrées (mai 2002, un tiers des entrées sont décrites). Il s'agit d'un dictionnaire en ligne (<http://www.kuleuven.ac.be/dafles>) qui n'est pas dérivé d'une version papier et dont l'interface, grâce à des scripts PHP, tente d'apporter plus d'interactivité et d'efficacité dans la consultation. Le dictionnaire repose sur une base de données relationnelle MySQL, ce qui permet d'extraire aisément les données et de les réutiliser dans d'autres applications.

<sup>2</sup> Ce corpus, qui regroupe une année (1998) des quotidiens *Le Monde* (France) et *Le Soir* (Belgique) et deux années (1998 et 1999) du quotidien *Le Devoir* (Québec), contient près de 80 millions de mots lemmatisés par l'analyseur Cordial Analyseur version 7 (Cordial, 2002).

<sup>3</sup> Mot polysémique ou monosémique qui regroupe un ensemble cohérent d'acceptations. *voler* fait référence à deux vocables.

<sup>4</sup> 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.

L'article expose les caractéristiques et les limites de ces exercices. Ceux-ci sont disponibles en ligne à l'adresse Internet : <http://www.kuleuven.ac.be/alfalex>.

## 2 Exercices contextuels générés automatiquement

Les trois premiers exercices, qui sont des exercices contextuels, ont été obtenus en faisant analyser les sorties de Cordial par des exécutable en C. Ces programmes ont produit des énoncés qui ont été stockés dans la base de données du système et qui sont ensuite affichés par des scripts PHP.

Quant au quatrième, il s'agit d'un exercice automatique hors contexte. Il permet d'étudier les relations d'actance entre plusieurs vocables (ou plus précisément entre leurs sens, les lexies).

### 2.1 Exercice de morphologie

Dans cet exercice, l'apprenant 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. 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 effectué à partir des codages morphologiques du DAFLES. On obtient en tout une trentaine de terminaisons pertinentes pour l'exercice (*créatrices, travaux, longue, aiguë, européennes*, etc.), au stade de la description du dictionnaire (4000 mots). Le questionnaire comprend une dizaine de phrases à compléter, chacune proposant une terminaison différente.

**Exercice de morphologie**

Remplissez les cadres avec les mots en italiques en fin de phrase  
en les adaptant au contexte

1 L'étude, essentiellement [ ], mais qui a prouvé par le passé qu'elle reflétait bien les tendances réelles, porte sur les prévisions de production en France. *qualitatif*

2 Mêlant, avec cet inimitable style hongkongais, action et sentimentalisme, Johnnie To ausculte avec un étonnant réalisme la vie [ ] d'une caserne de pompiers. *quotidien*

3 Il faut créer un climat qui permette aux jeunes entrepreneurs de libérer leur énergie [ ]. *créateur*

4 Il y a six ans, explique au Monde Matthew Barrett, le PDG de la Banque de Montréal, je disais déjà dans tous mes discours que des fusions étaient nécessaires entre les six grandes banques [ ]. *canadien*

5 Surplombant, au perchoir, une vingtaine d'élus pressés d'en finir, le vice-président de l'Assemblée nationale fait défiler à grande vitesse les [ ] liasses d'amendements de la droite. *dernier*

Figure 1 : extrait de l'exercice de morphologie

Cet exercice a été produit à partir d'une portion du corpus de 700 000 mots. Une fois les textes étiquetés, chacun de ces mots s'est vu attribuer un codage morphologique par une jointure externe sous MySQL avec la table de morphologie du DAFLES. Ensuite, pour diminuer la taille des données et les rendre plus maniables, ont seules été retenues les phrases contenant au moins une terminaison pertinente, avec un maximum de trente phrases pour chaque terminaison. Au total, on obtient ainsi une table de 48 600 enregistrements. Par la suite, le script PHP tire aléatoirement une dizaine de terminaisons, recherche les phrases qui

les contiennent, effectue un deuxième tirage aléatoire parmi elles et les affiche. Du fait des accords article-nom et adjectif-nom dans la phrase, il est toujours possible de savoir d'après le contexte le genre et le nombre du mot à compléter et il n'y a pas d'ambiguïté. La marque du pluriel étant facilement repérable, le seul problème peut provenir, dans le cas d'un adjectif, du genre du nom qu'il qualifie. Par exemple, dans la phrase 5 de la figure 1, l'adjectif *dernier* se rapporte à *liasse*. Ce nom n'étant pas très fréquent, l'apprenant peut fléchir convenablement *dernier*, mais se tromper sur le genre. Le système doit donc émettre un diagnostic en deux parties : l'adjectif est-il correctement fléchi (i.e. la terminaison indiquée fait-elle partie des quatre possibles) ? Si non, y a-t-il une erreur de genre ?

Pour éviter ce problème, une possibilité consisterait à contrôler les noms que qualifient les adjectifs à fléchir. On pourrait facilement appliquer des critères de fréquence, information contenue dans le DAFLES, et afficher des phrases avec des mots compris dans telle ou telle tranche de fréquence, en fonction du niveau souhaité. Mais il faudrait alors déterminer que l'adjectif s'accorde avec le nom en question, ce que ne peut faire le lemmatiseur. La solution consiste alors à utiliser l'analyseur syntaxique de Cordial. C'est ce que nous avons fait pour l'exercice suivant qui porte sur le genre.

## 2.2 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. Le système contient une liste de 32 noms communs (*aspect*, *dialogue*, *problème*, *interview*, *loi*, *programme*, *voix*, etc.) sur lesquels les étudiants néerlandophones font de fréquentes erreurs de genre. 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.

**Exercice sur le genre**

Complétez les phrases en accordant les articles et adjectifs avec le nom commun en gras

- 1 Parmi les quatre **premiers**, le livre arrivait en tête pour une raison majeure qui tient à une caractéristique très originale : c'est lui qui possède le plus d'**articles** différen[ ] .
- 2 En Belgique, on a coutume d'insister sur c[ ] **aspect** des choses mais il y a d'autres pays de par le monde qui ont une structure fédérale comme la nôtre et qui vivent sans problème avec un CNO unitaire.
- 3 Pour s'attaquer au coffre-fort grand-ducal, Leekens privilégie la méthode douce, à charge pour Walem de trouver l[ ] bo[ ] **combinaison**.
- 4 Certes, elle est inscrite depuis longtemps à un organisme de dons d'organes mais, dit-elle, "ce fut l[ ] **conséquence** et non pas la cause du choix.
- 5 A terme, les praticiens pourront trouver sur le RSS un ensemble de **contenus** médica[ ] .
- 6 Le partage en réseau des données médicales devrait surtout favoriser l[ ] **dialogue** entre médecins afin d'améliorer la qualité et la rapidité des soins.
- 7 Hors d[ ] **élément** exceptionnelle[ ] , le bénéfice progresse de 24,7 %, à 3,417 milliards de francs.
- 8 Certain[ ] **pannes** résultent des difficultés de paramétrage : par **exemple**, impossible d'installer un CD-ROM 256 couleurs lorsque l'écran est paramétré en milliers ou en millions de couleurs.

Figure 2 : extrait de l'exercice sur le genre des noms

Pour produire ces énoncés, une partie du corpus a été analysée par le parseur de Cordial. Outre le lemme et la catégorie grammaticale, le logiciel indique le groupe syntagmatique du mot ainsi que le pivot du syntagme nominal. L'interrogation porte alors sur les adjectifs et les articles de ce syntagme, du moins ceux qui présentent une variation morphologique sur le

genre. Ainsi ne sont pas pris en compte les mots invariables en genre (*chaque, souple, les, deuxième*, etc.). Il faut aussi considérer les dépendances plus lointaines comme les cas où l'adjectif est attribut du sujet.

La production de l'exercice est dépendante des erreurs d'analyse du parseur. Les erreurs concernent la mauvaise constitution du syntagme nominal, le plus souvent à cause d'une erreur de catégorie grammaticale. Le groupe nominal « cassé », il n'est plus possible de déterminer les mots qui s'accordent avec le nom en question et l'énoncé peut indirectement donner la réponse. C'est le cas de la phrase :

*Le récit s'attarde sur l'aspect "flou", "glauque", "trouble" ou "diffus" de certaines surfaces.*

dans laquelle *trouble* est étiqueté comme verbe. A ce moment-là, *diffus* ne fait plus partie du syntagme d'*aspect*, sa terminaison n'est pas occultée et il est possible de savoir que le mot qu'il qualifie est masculin. D'autres erreurs de catégorie grammaticale peuvent produire des énoncés déviants. C'est le cas de « Le jour de l'élection, chaque Allemand vote deux fois » où *vote* est considéré comme nom et *Allemand* comme adjectif (le système demande alors d'accorder *Allemand* à *vote*). Ces cas sont difficiles à prévoir mais sont néanmoins marginaux. Les énoncés déviants peuvent aussi être produits lorsqu'il y a une ambiguïté syntaxique. C'est le cas des mots à la fois nom et adjectif comme *futur*. Dans la phrase « C'est un futur lauréat », *futur* peut aussi bien être nom qu'adjectif car le mot avec lequel il s'accorde, *lauréat*, est lui aussi ambigu. S'il l'on veut exclure ces cas, la seule solution consiste à contrôler manuellement les occurrences des mots pouvant avoir une double catégorie grammaticale.

Un autre problème concerne les cas d'anaphore dans des phrases comme

*Les lampes donneraient aux automobilistes une impression de sécurité largement illusoire, au même titre que le système ABS quand il est mal maîtrisé.*

ou bien le fameux accord du participe passé placé après le COD, qui hante nos mémoires d'écoliers, dans la phrase « Tout est parti d'une interview que mon entraîneur a donnée ». Dans les deux cas, Cordial n'indique pas que *maîtrisé* et *il* réfèrent à *système* et *donnée* à *interview*, ce qui laisse ces mots tels quels dans l'énoncé et permet de déduire le genre du nom antécédent. Là encore, il est difficile de prévoir ces cas et de retirer ces concordances. Cependant, d'un point de vue didactique, il n'est pas inintéressant de voir si l'apprenant est capable de tirer parti de ces mots anaphoriques pour résoudre son problème, si toutefois l'information pertinente n'est pas placée juste à côté de l'antécédent.

### 2.3 Exercice sur les collocations

Le but de l'exercice est de compléter une des composantes d'une collocation<sup>1</sup> à partir du contexte (et donc du reste de la collocation). Étant donné le caractère plus libre des collocations par rapport à d'autres unités polylexicales comme les locutions ou les

---

<sup>1</sup> Association privilégiée de plusieurs vocables (*poser une question, travailler dur, petit boulot*, etc.)

expressions semi-figées, l'énoncé est suivi d'une indication sous la forme d'une fonction lexicale (Mel'čuk et al., 1995) et de son argument pour aiguiller l'apprenant vers la bonne solution. Pour l'instant ont été sélectionnées les fonctions *Magn* (intensification) et *Oper* et *Func* (verbe support sémantique vide), car elles 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. Par exemple, *question* peut se combiner avec *poser*, qui est le verbe support sémantiquement vide, mais aussi avec *aborder*, *résoudre*, *soulever*, etc. qui apportent une nuance et qui sont exprimés par d'autres fonctions lexicales (*Incep*, *Real*, etc.) moins intuitives. Ceci n'exclut pas pour *Magn* et *Oper* les cas de synonymie et la possibilité de pouvoir donner plusieurs réponses exactes (on peut tout aussi bien *jouer* un match que le *disputer*). Pour établir un diagnostic adéquat, le système explore les informations consignées dans le DAFLES, et compare la réponse de l'apprenant avec tous les verbes support pour le mot en question. Si elle en fait partie, elle sera acceptée même si ce n'est pas le verbe effectivement employé dans la phrase d'origine.

**Exercice sur les collocations**

Remplissez les cadres à l'aide des informations données  
entre parenthèses à la fin des phrases

1 En plus, maintenant, elle ne se contente plus de [ ] des questions sur l'actualité. (verbe support de *question*)

2 M. Bardon a donc [ ] recours à une pratique admise par l'administration mais interdite par le législateur avant la loi du 10 juillet 1991. (verbe support de *recours*)

3 S'il n'y avait pas un effort de l'Algérie pour se distinguer du Maroc en [ ] appel à son histoire, à des mythes fabriqués de toutes pièces sur son identité, il serait difficile de les distinguer. (verbe support de *appel*)

4 Le groupe pharmaceutique Astra indiquait, peu après Ericsson, qu'il pourrait délocaliser une [ ] partie de ses activités de recherche et développement vers la Grande-Bretagne et les Etats-Unis. (intensification de *partie*)

5 A Cardiff, l'Europe des nations [ ] des points face aux tenants du fédéralisme. (verbe support de *point*)

6 On [ ] le ménage dans sa tête et on s'emplit de beauté. (verbe support de *ménage*)

Figure 3 : extrait de l'exercice sur les collocations

En raison de la variation des collocations (flexion des composantes, transformations syntaxiques), il n'est pas possible d'extraire directement les phrases où elles occurrent. Il faut passer par un codage intermédiaire inspiré de (Segond, Tapanainen, 1995) qui est ensuite interprété par le programme d'extraction des phrases du corpus. Ainsi, nous avons opté pour un codage simple indiquant les parties fixes (précédées de +), variables (laissées telles quelles) et optionnelles (précédées de :). De même, nous indiquons quel est le mot à retrouver (c'est-à-dire soit le verbe support, soit l'intensificateur) par le symbole \* :

- \*gagner :ADV POSS +vie (*gagner sa vie*, ADV adverbe optionnel, POSS adjectif possessif et *vie* singulier)
- \*jouer DET match (DET article, démonstratif, adjectif possessif, etc.)
- ++bonne +partie +de (intensification de *partie*)

Bien entendu, ce codage suit l'ordre linéaire des mots dans les phrases et ne rend pas compte d'éventuelles transformations syntaxiques pour les collocations verbales comme les passivations ou les mises en relative. Ainsi, le programme ne repèrera pas *poser une question* dans « la question que je me suis posée ». Néanmoins, ce problème n'est pas gênant car la majorité des occurrences des collocations suivent a priori l'ordre « simple » du codage et

nous ne cherchons pas à extraire toutes les occurrences mais un nombre suffisant (une trentaine) pour avoir une certaine variété dans les phrases exemples. Le problème est plus ennuyeux lorsqu'il s'agit de compter les collocations (pour que l'exercice ne soit pas trop difficile, seules les dix<sup>1</sup> collocations les plus fréquentes sont utilisées) car il n'est pas possible d'avoir un compte exact. Néanmoins, on obtient un comptage relatif qui permet d'établir une liste de fréquence des collocations plus ou moins exacte<sup>2</sup>.

## 2.4 Exercice hors contexte sur les schémas actanciels

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

Mots utilisés avec **construire** :

	action	N1:sujet	verbe	N2
1	réaliser une construction		construire	
2	assembler les éléments d'un appareil		construire	

[Retour](#) au sommaire.

**Aide**  
 Comment appelle-t-on la personne ou l'entreprise qui construit un appareil (une machine, une voiture, un avion) ?

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

Ces questions sont disponibles pour l'action et le sujet du verbe. Elle sont générées automatiquement par le système et exploitent la systématisme des définitions du DAFLES. Celles-ci sont sous forme de phrases dans le style rédactionnel du Collins COBUILD (Sinclair, 1995). 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 la lexie décrite comme sujet par les schémas actanciels. Ainsi pour l'exemple de la figure 4, on a (*constructeur* étant déclaré sujet de *construire* dans le DAFLES) :

construire (sens 2) : lorsqu'une personne ou une entreprise construit *un appareil* (*une machine, une voiture, un avion*), elle le réalise en assemblant les différents éléments qui le composent selon un plan préétabli.

<sup>1</sup> Sur un petit échantillon de 200 collocations *Magn* et *Oper* extraites du DAFLES.

<sup>2</sup> D'après un comptage portant sur une dizaine de collocations, la proportion de passivations et de mises en relative est de l'ordre de 5 à 20 % suivant les cas.

constructeur (sens 2) : un constructeur est *une personne ou une entreprise* qui réalise des appareils (une machine, une voiture, un avion, un ordinateur) en assemblant les différents éléments qui les composent selon un plan préétabli.

La question est alors générée de la manière suivante :

« Comment appelle-t-on » + [complément de l'actant sujet] + « qui » + [verbe conjugué] + [complément du verbe] ?

soit (les articles indéfinis du sujet sont remplacés par les articles définis) :

Comment appelle-t-on la personne ou l'entreprise qui construit un appareil (une machine, une voiture, un avion) ?

Dans ce cas précis, il n'est pas nécessaire de tirer l'information *une personne ou une entreprise* de l'actant sujet, étant donné qu'elle figure déjà dans la définition de *construire*. Cependant le sujet dans la définition du verbe n'est pas toujours un hyperonyme, et il faut alors le chercher dans celle du sujet. Par exemple, pour la définition d'un des sens de *jouer* : *lorsqu'un musicien joue un morceau de musique*, c'est la définition de *musicien* qui indique qu'il s'agit d'une personne.

Quant aux compléments à découvrir (colonne N2 de la figure), il est plus délicat de formuler des questions. Plutôt que de poser une question, l'aide alors proposée est en fait la définition avec le mot à découvrir masqué.

L'un des problèmes est que la réponse au complément N2 se trouve parfois dans la glose portant sur le sujet. Ainsi dans l'exemple de la figure 4, en lisant la question pour le sujet, on peut aisément savoir que ce qu'on construit est un appareil (*machine* peut aussi être accepté en exploitant l'axe paradigmatique par le biais des liens de synonymie ou d'hyper/hyponymie décrits dans le DAFLES). Ce cas n'est toutefois pas systématique et là aussi, comme dans l'exercice sur le genre, il n'est pas inintéressant de vérifier que l'apprenant utilise l'information à sa disposition pour résoudre l'exercice.

### 3 Exercices contextuels semi-automatiques

Contrairement aux précédents, les trois exercices suivants (déjà abordés dans (Selva, Chanier, 2000)) ne peuvent être générés automatiquement car ils nécessitent une désambiguïsation sémantique du vocable dans les phrases de contexte. Dès lors, les énoncés de base sont extraits du corpus manuellement à l'aide d'un concordanceur et mis dans la base de données. Pour plus de variété, chaque lexie possède une ou plusieurs phrases exemple en fonction du rang de fréquence du vocable. Naturellement les mêmes phrases sont utilisées dans les trois exercices, ce qui diminue la charge de travail d'extraction et rentabilise un peu l'opération. Ces exercices mettent en jeu les synonymes, les dérivés et, de manière classique, les traductions. Toutes ces informations sont consignées dans le DAFLES. Pour les synonymes et les dérivés, le principe est le suivant : à partir d'un mot donné ou plutôt d'une lexie donnée, sélectionnée par l'apprenant ou par l'enseignant, le système parcourt le dictionnaire pour en extraire les synonymes ou les dérivés. Ensuite, il extrait du corpus les phrases qui contiennent ces mots dans le sens adéquat et les affiche en enlevant les occurrences des mots de départ. L'apprenant doit donc replacer les synonymes ou les dérivés suivant le cas en fonction du



contexte des phrases. Pour les traductions, c'est le même principe, sauf que les mots ne sont pas forcément reliés par une relation synonymique ou dérivationnelle. Ce dernier exercice est plutôt destiné à des vocables un peu « isolés » dans les graphes sémantiques et dérivationnels ou qui ne sont pas reliés les uns aux autres (pour récapituler le travail d'une session par exemple).

Remplissez les cadres avec les mots suivants (les mots doivent être conjugués) :  
**habiter, résider, demeurer, être\_domicilié, occuper**

1 En moyenne, sur les onze villes étudiées, le locataire qui a emménagé en 1997 paie environ 43 francs au m<sup>2</sup>, celui qui [ ] son logement depuis dix ans ne débourse que 29 francs.

2 Et près d'un Français sur quatre [ ] à moins de 10 kilomètres d'un parc naturel régional, huit sur dix à moins de 50 kilomètres.

3 Le site est [ ] en Suisse, mais les courriers électroniques qui lui sont adressés sont acheminés vers Londres.

4 La victime, elle, [ ] dans le home de val Ubraye où elle termine ses vacances.

5 Mais il n'a pas oublié que, à quelques centaines de mètres en contrebas de l'appartement qu'il [ ], près de 250 militaires, pour la plupart des Français du deuxième régiment du génie de Metz, ont remis en état une des écoles de Kruja, cité historique perchée dans les montagnes à une quarantaine de kilomètres de Tirana.

Envoyer Effacer

Figure 5 : exercice sur les synonymes

## 4 Conclusion

L'article s'est principalement focalisé sur la génération des énoncés, ce qui correspond en fait à l'avancement des travaux. Pour l'instant donc, ces exercices sont essentiellement des tests qui ne font que vérifier la maîtrise par l'apprenant des vocables en jeu et de leurs propriétés morphologiques, syntaxiques et sémantiques. Comme évoquée dans l'introduction, l'étape suivante consiste à faire en sorte que le système puisse établir un diagnostic « évolué » en discernant des fautes de frappe ou en détectant des réponses approchantes (comme des synonymes) par exemple. Le système doit également donner des éléments d'aide partiels mais significatifs sur la forme et le sens des mots (définitions, partie de définitions, concordances, informations rentrées dans un dictionnaire personnel, etc.) de manière à faciliter l'inférence et à aiguiller l'apprenant vers la bonne solution. Dès lors, ces exercices ne seront plus de simples tests mais de véritables activités pédagogiques favorisant l'apprentissage lexical du français.

## Références

COBUILD. Sinclair J. (éd.) (1995) : *Collins Cobuild English Dictionary*. London, Harper Collins.

Cordial (2002), *Cordial Analyseur*, version 7.0, Synapse Développement : <http://www.synapse-fr.com>

Didier (2002), *Accord, exercices autocorrectifs*, consulté en février 2002 : <http://www.didieraccord.com>

Jaser (2002), *Chez Jaser*, consulté en février 2002 : <http://site.ifrance.com/jaser/>

La Passerelle (2002), *The Half-Baked Interactive Tests*, consulté en février 2002 : <http://lapasserelle.com/lm/pagespeciales/half.baked/halfbakedtests.index.html>

Mel'čuk I., Clas A., Polguère A. (1995), *Introduction à la lexicologie explicative et combinatoire*, Louvain-la-Neuve, Duculot.

Segond F., Tapanainen P. (1995), Using a finite-state based formalism to identify and generate multiword expressions, *MLTT Technical Report Rank Xerox Research Center*, <http://www.xrce.xerox.com/publis/mltt/mltttech.html>

Selva T., Chanier T. (2000), Génération automatique d'activités lexicales dans le système ALEXIA, *Sciences et Techniques Éducatives*, Vol. 7 n°2, pp. 385-411.

Selva T., Verlinde S., Binon J. (2002), Le DAFLES, un nouveau dictionnaire pour apprenants du français, Actes du 10<sup>e</sup> congrès international EURALEX'2002 (*European Association for Lexicography*), à paraître.

Verlinde S., Selva T. (2001), Corpus-based vs intuition-based lexicography: Defining a word list for a French learner's dictionary, *Proceedings Corpus Linguistics 2001, Lancaster*, pp. 594-598.

Verlinde S., Selva T. (2002), Nomenclature de dictionnaire et analyse de corpus, *Cahiers de Lexicologie*, à paraître.