



HAL
open science

Génération automatique d'activités Lexicales dans le système ALEXIA

Thierry Selva, Thierry Chanier

► **To cite this version:**

Thierry Selva, Thierry Chanier. Génération automatique d'activités Lexicales dans le système ALEXIA. Sciences et Techniques Educatives, 2000, 7 (2), pp.385-412. edutice-00000282

HAL Id: edutice-00000282

<https://edutice.hal.science/edutice-00000282>

Submitted on 23 Nov 2003

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Génération automatique d'activités lexicales

Thierry Selva* - Thierry Chanier**

* Institut des Langues Vivantes, Katholieke Universiteit Leuven,
Dekenstraat 6, B-3000 Leuven, Belgique
thierry.selva@ilt.kuleuven.ac.be

**Laboratoire d'Informatique de l'Université de Franche-Comté,
16, route de Gray, F-25030 Besançon Cedex, France
thierry.chanier@univ-fcomte.fr

RÉSUMÉ. Cet article traite de la génération automatique d'activités lexicales au sein d'un environnement informatique d'aide à l'apprentissage lexical du français langue seconde : le système ALEXIA. Plus précisément, nous établissons en quoi les ressources lexicales intégrées au système et leur utilisation par des outils issus du traitement automatique du langage naturel conduisent à générer des activités informatisées d'un type nouveau, pertinentes et adaptées au travail déjà accompli par l'apprenant. Les ressources lexicales s'avèrent indispensables pour favoriser l'apprentissage en autonomie. Cependant, le système ne doit pas se résumer à ses ressources, mais doit s'appuyer sur des principes pédagogiques bien définis qui influent grandement sur sa conception.

ABSTRACT. This paper deals with the automatic generation of lexical activities within a computer assisted language learning environment of French as a second language: the ALEXIA system. We intend to show more precisely how the lexical resources which are fully integrated into the system and their use by tools originating from the natural language processing field allow us to generate new computational activities which are both relevant and learner centered. These lexical resources are of paramount importance to facilitate the autonomous learning process. The system however should not be limited to its resources. It should be based also on sound and well defined pedagogical principles which greatly influence its design.

MOTS-CLÉS : apprentissage lexical, Apprentissage des Langues Assisté par Ordinateur (ALAO), Français Langue Étrangère, activités lexicales, réseaux lexicaux, corpus, lexique mental.

KEY WORDS: lexical learning, Computer Assisted Language Learning (CALL), French as a Foreign Language, lexical activities, lexical networks, corpus, mental lexicon.

Sciences et Techniques Educatives, Hermes, 2000, 7 (2), pp.385-412

1 Introduction

Les activités pédagogiques dans les systèmes d'aide à l'apprentissage lexical consistent le plus souvent en la programmation d'exercices conçus à l'origine pour le papier : les matériaux sont créés par les concepteurs, les épreuves ne sont pas renouvelables et la sélection des mots sur lesquels portent les exercices ne se fait pas toujours en fonction du travail de l'apprenant et de ce qu'il a réellement acquis. Avec les progrès techniques, les systèmes peuvent maintenant incorporer de nombreuses ressources lexicales ainsi que des outils issus de la branche du traitement automatique du langage naturel tels que dictionnaires électroniques sous forme de réseaux, analyseurs morphologiques et syntaxiques, logiciels de traitement de corpus, etc. Il est donc nécessaire de s'interroger sur l'influence de ces ressources et de ces outils dans le cadre d'activités pédagogiques sur le lexique : peuvent-ils générer des activités informatisées d'un type nouveau ? Si oui, quelles sont les caractéristiques de ces activités, leurs limites ? Quelles ressources lexicales doivent être intégrées dans le système ? Et enfin, cette orientation dans la conception des systèmes d'aide à l'apprentissage lexical permet-elle de considérer davantage et de mieux exploiter le travail de l'apprenant (concernant les mots auxquels il a été exposé, déjà acquis ou en passe de l'être) ?

Pour répondre à ces questions, nous nous proposons dans cet article d'exposer la méthodologie employée lors de la conception et de la réalisation du système ALEXIA, environnement informatique d'aide à l'apprentissage lexical du français langue seconde. Issu des travaux de l'atelier « Modélisation de l'acquisition lexicale en langue seconde » du GDR sciences cognitives de Paris, ce système a été développé au sein du Laboratoire de Recherche sur le Langage à l'Université Clermont2 à partir de fin 1994, puis au Laboratoire d'Informatique de l'Université de Franche-Comté de 1996 à 1999. ALEXIA, qui est pour l'instant en l'état de prototype, est un système composé de plusieurs parties ou modules : corpus de textes, dictionnaire général, dictionnaire personnel, activités lexicales, modèle de l'apprenant sous forme de traces et analyseur morphologique ([CHA 95], [CHA 98] et [SEL 99]). Nous avons décidé d'explorer les concepts sous-jacents par la focalisation sur l'un de ses modules, le générateur d'activités lexicales. Les activités étant produites à partir des informations contenues dans les autres éléments du système, ce module rend ainsi compte de sa cohérence.

Après un bref récapitulatif historique des environnements d'apprentissage lexical existants, nous exposerons les principes pédagogiques qui ont présidé à la conception d'ALEXIA en mentionnant les principaux résultats des travaux en psycholinguistique et en didactique des langues. Nous effectuerons ensuite une étude concise des différents modules avant de passer aux activités lexicales proprement dites. Leur type ainsi que leur apport pédagogique dans l'environnement sera alors justifié. Enfin, nous verrons comment les informations et les outils que les modules de l'environnement contiennent permettent de générer des activités pertinentes et adaptées, qui seront à ce moment-là exposées en détail.

Avant de continuer, nous précisons la terminologie que nous employons concernant les *mots* : un *vocable*, qui correspond à un mot polysémique (ou monosémique dans certains cas), est un tout sémantique considéré comme la somme de ses sens ou acceptions (travail est un vocable qui a 11 sens dans la figure 3, voler fait référence à deux vocables, planer et dérober). Une *lexie* est chacune des acceptions d'un vocable.

2 Les environnements informatiques d'aide à l'apprentissage lexical

Les premiers programmes d'aide à l'apprentissage lexical consistaient pour la plupart en des versions informatiques de jeux faisant intervenir les mots tels que les anagrammes ou les mots-croisés. Ils permettaient de développer le vocabulaire de l'apprenant mais de façon fortuite [KEN 90]. En raison de l'interactivité limitée de ces programmes (pas d'interface graphique, il fallait taper les vocables en entier) et de la limitation des algorithmes (l'ordinateur stockait toutes les solutions), l'accent a surtout été mis sur les vocables courts et sur l'orthographe et non le sens. Ce qui semait le doute sur leur supériorité par rapport aux exercices papier. Par la suite, beaucoup a été fait pour rendre ces programmes plus attrayants et plus motivants (ajout de couleur, de sons, de graphiques, mise en place de systèmes de notes) mais, comme il a été noté [CLA 92], ces programmes mettaient plus l'accent sur l'évaluation par des tests de ce qui avait déjà été appris et non pas sur le processus d'apprentissage lui-même. Certains programmes étaient toutefois porteurs d'idées pertinentes, s'appuyant sur les résultats de recherches en pédagogie, psychologie ou études en acquisition (par exemple Wordchip [DEC 93], qui permettait de sélectionner les vocables à étudier dans de vastes bases lexicales, ou StoryBoard [HIG 84] sur la reconstruction de textes qui mettait en œuvre la capacité d'inférence des apprenants).

Avec les progrès technologiques, les environnements d'aide à l'apprentissage ont bénéficié de nombreuses améliorations. Tout d'abord, l'emploi d'interfaces graphiques les a rendus plus conviviaux et plus intuitifs. D'autre part, la capacité de traitement s'étant accrue, ils permettent maintenant de pouvoir organiser le travail des apprenants et donc de personnaliser leur apprentissage. Enfin, l'amélioration des performances techniques a permis l'incorporation de ressources lexicales (corpus de textes, dictionnaires, concordanceurs, etc.) ce qui, dans les environnements d'aide à l'apprentissage lexical, favorise l'apprentissage en autonomie.

Parmi ces environnements de deuxième génération, nous pouvons citer Mayday [SUS 94] et Lexica [GOO 94]. Le premier traite du phénomène de la composition de vocables en anglais à l'aide d'affixes, de préfixes et de suffixes. Il repose sur un jeu d'activités lexicales (exercices sur les affixes, mots-croisés, correspondances de sens entre une série de définitions et une autre de vocables composés à l'aide d'affixes, de préfixes et de suffixes, exercices à trous) et sur plusieurs outils lexicaux complémentaires (un dictionnaire pour apprenants, un dictionnaire de synonymes, un corpus d'exemples, une présentation sonore des vocables et un module de

construction de vocables composés à l'aide d'une racine). L'apprenant a donc à sa disposition plusieurs sources d'informations et a la possibilité d'appliquer plusieurs stratégies pour résoudre les exercices. Ce système aborde donc véritablement la question de l'apprentissage lexical et les types d'interactions que les apprenants peuvent avoir avec de tels programmes. Le deuxième, Lexica, repose sur un modèle (Kukulska-Hulme, voir figure 1 plus bas) comprenant trois étapes : sélection de vocables lors de la lecture de textes, regroupement dans un dictionnaire personnalisé et exercice de rappel. De même que plus haut, l'environnement comporte des ressources lexicales (corpus de textes, dictionnaire bilingue et concordanceur). Pour Goodfellow, les vocables ne sont pas des étiquettes arbitraires mais sont reliés entre eux de manière systématique et forment des familles ou champs sémantiques. Il s'agit donc d'aider à la mémorisation, en vue d'une utilisation productive, en essayant de faire prendre conscience de ces relations lexicales concernant des vocables nouvellement rencontrés. De même que Mayday, Lexica est un environnement ouvert, dans le sens où l'apprenant choisit le contenu lexical à étudier et la manière dont il l'étudie, c'est-à-dire les stratégies qu'il utilise.

Cependant, aucun de ces environnements n'intègre d'outils de traitement automatique spécifiques. Les dictionnaires sont uniquement utilisés pour la consultation humaine, les exercices sont de simples activités de rappel ou proviennent d'un stock déjà constitué et non renouvelable et les ressources textuelles ne sont pas exploitées de manière systématique et automatisée. Il convient donc d'examiner l'apport de ces outils et la façon dont ils peuvent tirer profit des différentes ressources lexicales de l'environnement pour la génération d'activités d'un type nouveau.

3 Principes pédagogiques

Comme nous l'avons vu précédemment, même si elles sont indispensables pour un apprentissage en autonomie, un environnement ne doit pas se résumer à ses ressources, mais s'appuyer sur des principes pédagogiques avérés. Concernant ALEXIA, nous avons repris les principaux résultats qu'ont mis en évidence les travaux de recherche en psycholinguistique sur l'organisation du lexique ainsi que ceux en didactique des langues.

Il en ressort que :

- loin d'être une multitude de connaissances non structurées, le lexique mental est hautement organisé [AIT 87], ce dont témoignent ses performances (taux de reconnaissance, quantité d'informations contenues). Il a été mis en évidence par l'étude des lapsus et des associations que le lexique mental est un vaste maillage de concepts exprimés non pas par les mots, mais par leur sens (théories des toiles verbales [AIT 87]). Les sens des mots, les lexies, sont reliés par des relations lexicales d'ordre sémantique, morphologique ou contextuel (synonymie, hyperonymie, dérivation, orthographe, etc.)

- l'organisation du lexique mental est personnelle et évolue constamment dans le temps, ce qui rend difficile sa représentation.
- l'assimilation de nouvelles connaissances dépend du niveau du traitement : plus la tâche d'apprentissage est complexe, plus l'enregistrement dans la mémoire qui en découle sera riche, détaillé et précis. L'enregistrement d'un vocable n'est pas un phénomène ponctuel et définitif, mis en place une fois pour toute. Il doit être réactualisé pour subsister. Or, plus la trace mémorielle est riche et précise, plus elle a de chances d'être retrouvée, réutilisée et, par ce fait même, renforcée [BOG 94]. Ainsi, un travail en profondeur sur le sens des vocables est plus profitable qu'un travail sur leur forme qui restera superficiel. De ce fait, les liens avec les autres éléments du lexique mental seront plus forts.
- l'apprentissage est un processus graduel et lent : il est nécessaire de répéter l'exposition et de faire réviser le vocabulaire déjà acquis ou en passe de l'être.
- il faut apprendre à apprendre : l'inférence (trouver le sens d'un mot à partir du contexte) est une stratégie d'apprentissage en autonomie. L'environnement devra favoriser cette stratégie.
- les mots n'ont pas le même statut face à l'apprentissage : l'influence de la langue maternelle a un rôle important pour les congénères et pour les faux-amis. Il y a aussi les expressions (prendre le taureau par les cornes, travailler au noir, etc.) qui posent à la fois des problèmes de compréhension (leur signification ne peut pas toujours être déduite à partir de leurs composants) et surtout de production car elles sont imprédictibles. Leur utilisation témoigne du degré de maîtrise de la langue cible par l'apprenant.

Ces principes ont guidé fortement la conception du système. Ainsi, le dictionnaire du système tente de reproduire, en mettant en association les vocables et surtout les lexies qui le composent, une partie d'un « lexique mental de référence », c'est-à-dire contenant des informations linguistiques standard et servant de base pour l'organisation lexicale personnelle des apprenants. Le rôle du dictionnaire personnel qui, comme son nom l'indique, assure la personnalisation de l'apprentissage, est également primordial dans le sens où il permet de conserver le vocabulaire étudié par l'apprenant et facilite de ce fait les révisions. En outre, la construction même de ce dictionnaire, par regroupement de vocables présentant des similitudes sémantiques ou syntaxiques, est une tâche linguistique complexe qui nécessite de la part de l'apprenant une prise de conscience sur les relations entre vocables et donc un travail sur la langue elle-même. Cette tâche est renforcée par des activités lexicales qui tentent de développer les facultés d'inférence, en faisant travailler l'apprenant sur des vocables en contexte (dans des phrases et non pas isolés) et en mettant en jeu des relations sémantiques complexes telles que la synonymie ou l'actance par exemple.

Pour intégrer ces principes dans ALEXIA, nous nous sommes appuyés, tout comme le système Lexica de Goodfellow, sur le modèle de Kukulska-Hulme [KUK 88] : « voyage d'un item lexical » (figure 1).

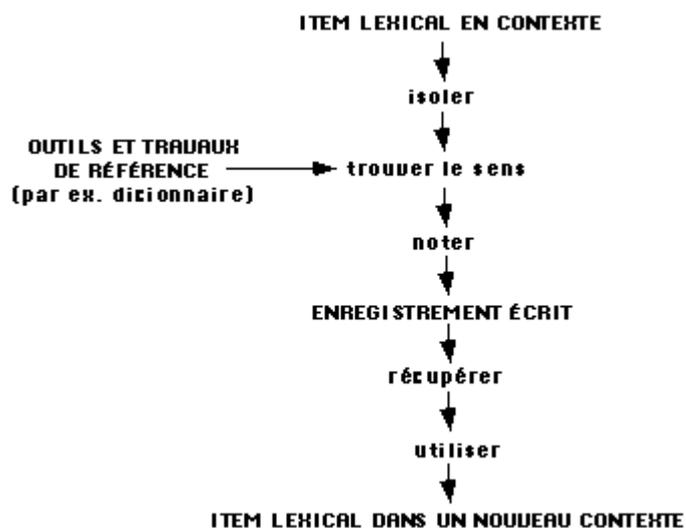


Figure 1. *Le modèle de Kukulska-Hulme, « voyage d'un item lexical »*

D'après ce modèle, la maîtrise d'un vocable comporte trois phases : exposition et compréhension d'un vocable dans un contexte, rétention et production dans un nouveau contexte. Le système prend en compte chacune de ces trois phases : textes et dictionnaire, dictionnaire personnalisé et activités lexicales. L'utilisation type d'ALEXIA est donc le choix et la lecture de textes, la compréhension à l'aide du dictionnaire, la sélection de vocables et le regroupement dans le dictionnaire personnel et enfin la résolution d'activités lexicales générées automatiquement à partir des informations dans les différents modules du système. Nous nous proposons à présent de présenter ces différentes parties avant de nous attarder sur les activités lexicales.

4 Les modules d'ALEXIA

4.1 *Le corpus*

La constitution du corpus est une étape importante qui doit être effectuée avec soin. En effet, le thème du corpus va orienter celui de l'environnement. Il a donc été nécessaire de circonscrire le cadre d'étude : nous ne désirons pas que l'apprenant travaille sur la langue en général car, d'une part, il est préférable de se recentrer sur un thème déterminé et suffisamment généralisable pour plus d'efficacité, et d'autre part, cela nécessiterait des ressources lexicales en conséquence : corpus diversifié

répertoriant tous les mots (plusieurs dizaines de millions d'occurrences sont alors indispensables) et dictionnaire décrivant toute la langue. Comme aucun corpus de cette taille n'est disponible en français (FRANTEXT est un corpus principalement littéraire et non actuel) et comme il n'existe aucun dictionnaire concluant de français pour apprenants contenant les informations dont nous avons besoin (voir [BOG 98] sur le déficit en matière de dictionnaires pour apprenants du français), nous avons donc dû constituer ces ressources. Le thème choisi a été celui du *travail*, de l'*emploi* et du *chômage*, car c'est un domaine du français courant auquel tout natif adulte est en principe confronté et qui n'est donc pas réservé à telle ou telle classe sociale.

Visant l'apprentissage du français actuel, les textes font partie principalement des années 1994 à 1996. Nous avons voulu constituer un corpus représentatif, en variant les sources et le niveau de langue. Ainsi, nous avons regroupé des textes de la presse nationale et quotidienne, des magazines, des livres et même des journaux vendus dans la rue par des sans domicile fixe. Au total, le corpus est constitué d'environ 400 textes, ce qui représente 450 000 occurrences. L'ensemble des textes a été concaténé en un seul fichier qui a été transmis au serveur SILFIDE [SIL 00] afin d'être mis à la disposition de la communauté de chercheurs. En retour, le corpus a été converti au format SGML.

Outre son contenu, chaque texte possède ses propres références, parmi lesquelles les mots-clés, utiles comme critère de sélection du texte à lire.

4.2 Dictionnaire du système ou base lexicale de référence

La finalité du corpus n'est pas uniquement l'exposition aux vocables lors de lectures. Nous nous en sommes servis comme base pour la description lexicographique du dictionnaire de l'environnement. Nous avons ainsi isolé, par analyse statistique, les vocables caractéristiques du domaine ainsi que certaines collocations. Le dictionnaire comprend environ 200 vocables, ce qui représente près de 400 lexies. Ce nombre de 200 entrées peut paraître dans un premier temps nettement insuffisant, en regard des quelques 60 000 entrées environ que possède habituellement un dictionnaire monolingue. Cependant, il faut avoir à l'esprit que tous les vocables de notre dictionnaire appartiennent au même champ notionnel et constituent en quelque sorte un sous-domaine de la langue. Certains vocables sont définis avec d'autres vocables du dictionnaire et on retrouve les structures lexicosémantiques classiques du lexique (qui, cependant, sont loin de figurer dans tous les dictionnaires habituels). Ainsi, chaque entrée (voir figure 2) contient deux types d'informations. D'une part les informations génériques propres à toutes les lexies d'un vocable (catégorie grammaticale, genre, graphie, différenciation entre homonymes si besoin est, fréquence dans le corpus), d'autre part les informations propres à chaque lexie (numéro de sens, de sous-sens si nécessaire, définition abrégée et détaillée, syntaxe, contraintes sémantiques sur les arguments du verbe, registre, lien vers un fichier d'exemples extraits du corpus). De plus, pour chaque lexie, sont indiquées les relations sémantiques et syntaxiques qui la relie aux

autres lexies : quasi-synonymie, hypo/hyponymie, synonymie intersective, antonymie, relations d'actance et de dérivation morphologique (avec catégorie grammaticale et numéro d'argument des actants et dérivés) et fonctions lexicales [MEL 92]. Ces informations sont affichées à la fois en mode texte (figure 3, pour les définitions) et en mode graphique sous forme de réseaux lexicaux interactifs (figure 4) [CHA 98].

Loin de représenter un ensemble fragmentaire et incohérent, le lexique que nous avons isolé va nous servir à illustrer et à valider certaines hypothèses que nous pouvons formuler sur les stratégies d'accès lexical et l'utilisation de dictionnaires par des apprenants d'une langue étrangère.

| |
|---|
| <p>% DIRIGER catégorie grammaticale: verbe genre: non graphie: diriger homonyme: non fréquence dans le corpus: 138</p> <p>Sens 1a : définition abrégée : "contrôler une entreprise, une organisation, une opération" définition : "lorsqu'on dirige une entreprise, une organisation ou une opération, on en est responsable, on la contrôle, on donne des ordres et on prend des décisions pour que le travail soit fait correctement" syntaxe : n1,V,n2 contraintes : n1 humain registre: courant exemple:172 hyponyme : être responsable de, être à la tête de, mener, gouverner, régir, administrer synonyme intersectif : gérer actant : (dirigeant,nom,1),(dirigeant,adjectif,1)</p> <p>Sens 1b : définition abrégée : "contrôler un groupe de personne" définition : "lorsqu'on dirige une personne ou un groupe de personne, on les commande, on leur dit ce qu'il faut faire" syntaxe : n1,V,n2 contraintes : n1 humain, n2 humain registre : courant exemple: 173 hyponyme : commander, être responsable de, être à la tête de, mener actant : (employé,nom,2),(ouvrier,nom,2)</p> <p>Sens 2 : définition abrégée : "conduire un véhicule" ... quasi-synonyme: conduire ...</p> <p>Sens 4b : définition abrégée : "orienter de manière hostile" définition : "lorsqu'on dirige quelque chose contre quelqu'un, on l'oriente vers lui de manière hostile et agressive" syntaxe : n1,V,n2,contre,n3 contraintes : n1 humain, n2 non humain, n3 humain registre: courant exemple: 177</p> |
|---|

Figure 2. Extrait de l'entrée diriger

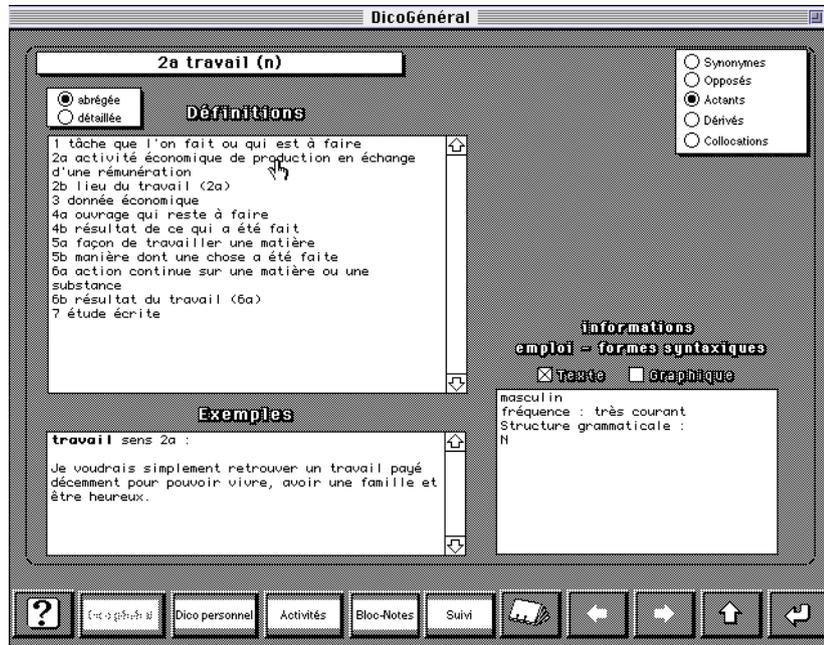


Figure 3. Article (définitions abrégées) du vocable travail

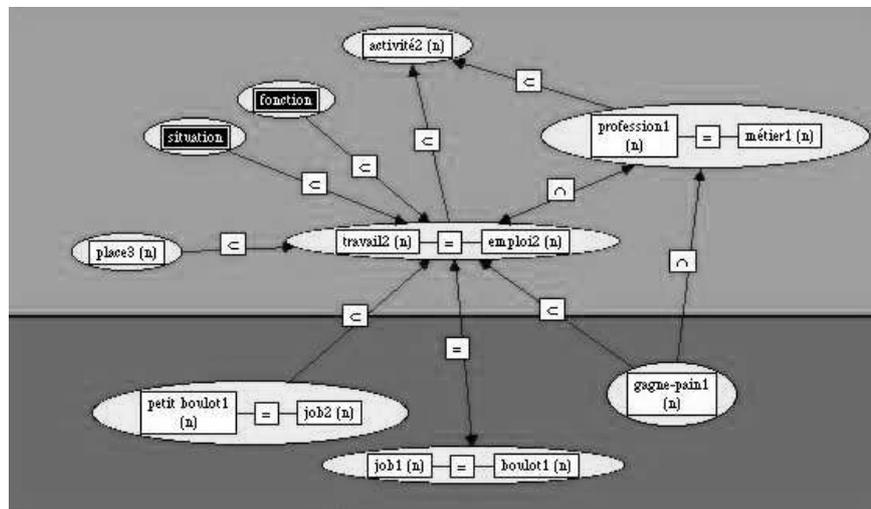


Figure 4. Réseau de synonymes pour emploi2

Nous ne discuterons pas dans cet article des problèmes concernant l'accès aux informations du dictionnaire ainsi que tout ce qui concerne l'aide à la compréhension et à la production (voir [SEL 98] et [SEL 99]). Ces aspects sont souvent négligés par les concepteurs de dictionnaires électroniques ([SEL 99], pp. 66-71) bien que plusieurs études sur l'utilisation des dictionnaires aient montré que ces derniers n'étaient pas toujours profitables aux apprenants dans des tâches de compréhension écrite.

La structure de la base de données simule l'organisation d'un « lexique mental de référence ». Ainsi, en plus des informations morphologiques, syntaxiques, sémantiques et pragmatiques associées aux vocables, les sens des mots sont reliés entre eux par des relations telles que la synonymie ou la dérivation. Le dictionnaire est donc un ensemble de réseaux lexicaux où l'accès alphabétique arbitraire n'est qu'un des moyens pour parvenir à extraire l'information recherchée.

4.3 Sélection des mots et dictionnaire personnel

A côté de ce dictionnaire de référence, l'environnement donne la possibilité à l'apprenant de constituer ses propres listes de mots. Les travaux de Goodfellow [GOO 95] ont montré l'importance d'un module de dictionnaire personnel visant à noter et à organiser le vocabulaire en partie connu. En effet, il ne suffit pas de chercher le sens d'un vocable dans un dictionnaire pour le retenir : c'est ce que montrent les travaux en didactique des langues ([TRE 96], [BOG 94]) ainsi que l'expérience avec GLOSSER [NER 98] où l'utilisateur consulte plusieurs fois les mêmes entrées de dictionnaire.

Nous avons élaboré et implémenté un dictionnaire personnalisé dans ALEXIA. Celui-ci permet de sélectionner des vocables dans un texte, de les regrouper suivant des caractéristiques communes (regroupement par sens (synonymes, antonymes), par forme (dérivés) ou par syntaxe (construction des verbes)) et de visualiser les groupes constitués. Cette visualisation permet, soit d'avoir une synthèse sur l'ensemble des mots étudiés en récapitulant l'ensemble des informations, ou bien d'avoir plus de détails en parcourant la fiche de chacun des mots et en la comparant avec d'autres si nécessaire. En outre, le module attribue un statut à chaque élément des groupes. Il correspond à la quantité de traitement effectué sur un item (calculé en fonction des actions effectuées autour de l'item, sélection dans les textes, consultation dans le dictionnaire, regroupement dans un ou plusieurs ensembles de mots, score dans les activités) et sera utilisé par les activités lexicales pour déterminer ceux sur lesquels elles porteront.

L'apprenant a également la possibilité de mettre des annotations sur les vocables qui ont été sélectionnés dans les textes : définitions personnelles, traductions en langue source ou toute information pertinente pour lui. Ces annotations sont sous forme de texte simple. Elles sont libres, c'est-à-dire qu'elles ne sont ni structurées, ni formalisées. De ce fait, le système ne peut que les reciter telles quelles sans

pouvoir les exploiter pour d'autres usages. Elles sont donc utiles principalement pour l'apprenant.

Par ces regroupements et ces annotations, l'apprenant a donc la possibilité de travailler en profondeur le sens et la forme des vocables qu'il a isolés des textes. Bien plus complet que la simple consultation de l'entrée dans le dictionnaire, ce travail de catégorisation permet de renforcer les liens faibles du lexique mental créés lors de la lecture et de l'exposition aux vocables.

Voyons maintenant comment peuvent être exploitées toutes les informations de ces modules pour pouvoir générer des activités lexicales.

5 Les activités lexicales

5.1 *Justification des activités lexicales*

En premier lieu, pourquoi intégrer des activités dans un environnement ? Leur apport est-il vraiment significatif ? C'est ce que prétendent Paribakht et Wesche [PAR 97] pour qui la présence d'activités lexicales dans l'apprentissage du vocabulaire se justifie par de meilleurs résultats des apprenants à des tests d'évaluation de compétences lexicales par rapport à d'autres processus d'apprentissage telle l'exposition à de nouveaux vocables par la lecture seule de textes. Les activités lexicales accéléreraient donc l'acquisition lexicale. Leurs travaux reprennent l'idée, déjà suggérée par Hulstijn [HUL 93], selon laquelle la rétention est favorisée par la quantité de travail effectué sur un vocable. Paribakht et Wesche ont donc mesuré les incidences pour l'apprentissage, d'une part, du processus de lecture seule, et d'autre part, du processus de lecture suivi d'activités lexicales. Le temps consacré aux activités dans le deuxième cas était compensé dans le premier par un supplément de lecture. Les activités proposées mettaient en œuvre plusieurs compétences lexicales, parmi lesquelles la reconnaissance des vocables dans des phrases, la capacité d'interprétation sémantique, de manipulation morphologique ou encore de production des vocables dans un nouveau contexte.

Les tests ont montré que les résultats étaient meilleurs lorsque le groupe d'apprenants pratiquait les activités lexicales. Ils ont permis de vérifier les hypothèses suivantes :

- Les étudiants possèdent une meilleure connaissance des vocables-cible après la séance de lecture suivie d'exercices mais aussi après la séance de lecture seule (ceci pour vérifier que la lecture était utile à l'apprentissage).
- Pour un temps donné et égal dans les deux cas, les gains en apprentissage étaient plus grands pour la lecture suivie d'exercices que pour la lecture seule.
- Les gains en vocabulaire étaient à la fois quantitatifs (plus de vocables connus à la fin) et qualitatifs (meilleure connaissance des vocables, mesurée par l'application d'une échelle de connaissance spécifique)

- Les gains dans le cas de la lecture avec exercices concernaient davantage les mots pleins (verbes, noms) que les mots grammaticaux.

L'expérimentation a aussi montré que les étudiants avaient une opinion favorable des activités, pensant que celles-ci amélioreraient leurs compétences lexicales.

5.2 Caractéristiques des activités informatisées

L'intérêt d'activités lexicales dans un environnement informatique est leur caractère computationnel, c'est-à-dire qu'elles doivent être conçues de manière à pouvoir exploiter le potentiel de l'ordinateur. La programmation et l'intégration d'activités élaborées à l'origine pour le papier est bien entendu tout à fait réalisable (c'est d'ailleurs ce que l'on retrouve le plus souvent, par exemple sur les sites Internet de La Passerelle [PAS 00]), mais l'on se priverait dans ce cas de la généralité et de la puissance de traitement de la machine.

Les activités informatisées doivent donc présenter quatre caractéristiques :

- Elles doivent être pertinentes en termes d'apprentissage, c'est-à-dire que l'apprenant doit fortifier ou valider son acquisition grâce à elles.
- Elles doivent être reproductibles, c'est-à-dire que l'apprenant pourra avoir de nouvelles épreuves à volonté. D'autre part, elles pourront potentiellement se déclencher sur l'ensemble du lexique et ne seront pas réservées à certains vocables.
- Elles doivent se prêter à la mise en place d'un système d'aide qui confortera le rôle pédagogique de l'environnement, c'est-à-dire que le système ne se contentera pas de répondre vrai ou faux, mais aiguillera l'apprenant pour trouver la solution.
- Elles doivent être réalisables en termes informatiques, c'est-à-dire que leur mise au point ne doit pas révéler une explosion combinatoire au-delà d'un certain nombre d'items considérés, et le temps de réponse du système, tant au niveau de la préparation des épreuves que dans l'aide qu'il apporte, doit être convenable et acceptable pour que l'activité puisse se dérouler normalement.

5.3 Type d'activité

Il existe deux grands types d'activités : les activités hors-contexte (ou décontextualisées) et les activités en contexte. Le premier type fait intervenir les mots seuls, isolés de la phrase et du texte. Dans ce cas, le sens du mot n'est pas déterminé, s'il est polysémique, et l'on ne distingue pas non plus les homonymes.

En raison de ces problèmes d'ambiguïtés et désirant développer les capacités d'inférence de l'apprenant, nous nous sommes tournés vers les activités en contexte. Celles-ci s'appuient sur un corpus de textes dont il est possible d'extraire des sous-parties (texte, paragraphe, phrase) mettant en évidence tel ou tel vocable. Replacé

dans son contexte, le sens du vocable est donc plus ou moins déterminé par ceux qui l'environnent, par la phrase, ou, plus occasionnellement, par le paragraphe. Ces activités ne portent pas seulement sur la connaissance du sens du vocable mais montrent aussi comment celui-ci s'emploie et fonctionne dans la chaîne discursive. Les différentes propriétés du vocable sont ainsi illustrées de manière plus naturelle et plus complète que dans le cas des activités hors-contexte. En passant du vocable à ses lexies, il est possible de concevoir des activités mettant en œuvre les différentes relations linguistiques vues plus haut, sans oublier les propriétés morpho-syntaxiques, collocationnelles ou pragmatiques.

Les activités en contexte permettent de plus à l'apprenant de développer ses capacités d'inférence. En effet, l'information manquante peut, dans certains cas, être déduite des éléments du contexte. Nous avons vu plus haut que l'inférence était un procédé utile à l'apprentissage en favorisant la connaissance de nouveaux vocables par une exploitation des indices contextuels.

Explorons donc ce deuxième type d'activités.

L'exercice le plus classique, très répandu, est le test de closure, ou texte à trous. Il consiste à « supprimer des vocables d'un texte et à inviter les sujets à restituer les vocables manquants » [MOT 75]. Il existe plusieurs critères pour supprimer les vocables que l'expérimentateur veut faire deviner. Ceux-ci peuvent être tout simplement retirés à intervalle fixe (tous les cinq à sept vocables, cas du test dit « classique » ou « aléatoire », [HUG 89] et [TRE 96]). Comme le retrait peut concerner n'importe quelle catégorie grammaticale (des verbes, des noms, mais aussi des articles, des pronoms, etc.), une variante consiste, par un étiquetage syntaxique préalable qui détermine la catégorie grammaticale de chaque vocable, à éviter les catégories les moins intéressantes comme les articles par exemple. Un autre critère concerne la fréquence des vocables à retirer, partant du principe qu'il y a une corrélation entre les compétences langagières d'un apprenant et la taille de son vocabulaire. Les vocables à retrouver sont donc ceux qui sont les moins fréquents. Coniam ([CON 96] et [CON 97]), dans son évaluation de ces trois variantes, conclut que le critère de fréquence conduit à des meilleurs tests.

Quant aux vocables à trouver, on peut soit les donner dans le désordre, soit, plus difficile, ne pas les donner, soit préparer une version à choix multiple (pour chaque vocable manquant, le sujet doit choisir parmi quatre réponses).

Les tests de closure présentent l'avantage d'être faciles à générer automatiquement à partir de n'importe quel texte. En termes de performances, l'évaluation de Coniam indique que les meilleurs tests (ceux produits à partir du critère de fréquence) mènent à la génération d'un tiers d'items non acceptables (c'est-à-dire trop faciles ou trop difficiles à trouver ou bien pas assez discriminatoires). Ce taux chute à moins de la moitié d'items acceptables dans le cas des tests de closure à intervalle fixe. Il semble donc malgré tout qu'une correction humaine soit nécessaire, même si la machine a fait le plus gros du travail.

Ces tests ont néanmoins montré leur efficacité et sont assez répandus. Toutefois ils présentent deux faiblesses. Premièrement, il n'y a pas forcément de relations entre les vocables à deviner. Même si ce n'est pas nécessaire, il semble plus

intéressant de faire travailler l'apprenant sur un champ sémantique déterminé en l'invitant à retrouver par exemple un ensemble de synonymes ou d'antonymes donné, ou bien des vocables appartenant à la même famille lexicale (dérivés morphologique comme travail, travailler, travailleur). Tout en exerçant les capacités d'inférence propres à tout test de closure, l'apprenant travaille et renforce les liens entre les réseaux de son lexique mental.

Deuxièmement, il n'y a pas de lien entre les vocables à deviner et ceux faisant partie de son dictionnaire personnel. Autrement dit, il ne connaîtra pas forcément les vocables qu'il doit retrouver, et aura d'autant plus de mal à réussir son épreuve. Il ne semble en effet pas pertinent de déclencher des activités lexicales sur des vocables inconnus. Le phénomène peut être limité en générant les tests à partir des textes lus, mais d'une part cela restreint le choix des textes et d'autre part la réussite peut s'expliquer par un simple effet de remémorisation, sans entraîner de réflexion sur le vocable lui-même.

L'une des activités lexicales que nous préconisons pour remédier à ces deux faiblesses consiste en un exercice de recontextualisation de vocables (figure 5) auxquels l'apprenant a déjà été exposé [TRE 96]. Il s'agit de redonner un contexte, le plus souvent une phrase, à des vocables donnés, de préférence dans le dictionnaire personnel. Comme le disent ces auteurs (pp. 116-117), «L'association de ces vocables avec les contextes appropriés met en œuvre des stratégies d'inférence qui s'exercent à partir des vocables eux-mêmes. Ceux-ci, étant partiellement connus des apprenants, permettent d'établir des liens (sémantiques ou morpho-syntaxiques) avec certains éléments de leur entourage et de tenter une interprétation du sens des phrases ».

A partir des données du dictionnaire personnel, le système définit une quantité de vocables (par exemple entre cinq et dix) parmi ceux suffisamment travaillés mais non encore maîtrisés par l'apprenant (voir 7.1 pour les critères de sélection), recherche dans le corpus de textes les phrases qui contiennent ces vocables et les présente à l'apprenant tout en retirant les vocables en question (figure 5). Tout comme dans le test de closure, l'apprenant doit deviner les vocables manquants.

L'exercice peut être orienté soit en compréhension, si l'on donne dans l'énoncé les vocables à placer dans les trous, soit en production, en laissant l'apprenant les deviner. Cependant, si aucune condition ne pèse sur les vocables, les phrases étant disjointes, cet exercice peut s'avérer dans la plupart des cas, très difficile à résoudre, et donc rebutant. Pourtant, nous allons voir qu'en liant les vocables à trouver par certaines relations linguistiques, l'exercice devient abordable et pertinent d'un point de vue productif.

Utilisez les mots suivants pour compléter les phrases (les mots sont présentés sous la forme qui convient au contexte approprié) :

salaires – patron - usines - licencierait – dirigerai – condition – emploi - impôts

1) Une entreprise qui _____ massivement aurait beaucoup de mal à continuer à attirer les jeunes diplômés.

2) Il éprouve de plus en plus de difficultés à se "vendre" à un employeur qui lui préférera souvent un non-chômeur cherchant à changer d'_____.

3) Le rachat du constructeur tchèque par Volkswagen, a entraîné un accroissement de la productivité, et aussi des _____ ...

4) Pourtant les _____ tournent au ralenti avec un chômage partiel important.

5) Les revenus nécessaires pour les financer sont levés principalement sous forme d'_____ et de charges sociales.

6) Il en est ainsi parce que se former n'est qu'une _____ nécessaire à l'emploi.

7) Le _____ d'une P.M.E., c'est un travailleur comme un autre.

8) À l'avenir, je _____ aussi mes recherches du côté des collectivités locales.

Figure 5. *Activité de recontextualisation*

Cet exercice peut être complexifié en donnant les vocables à placer uniquement sous leur forme canonique. La tâche est alors un peu plus difficile pour l'apprenant qui ne peut plus compter sur la flexion des vocables pour les placer dans les différentes phrases, mais doit aussi les fléchir pour les intégrer correctement.

Pour Tréville et Duquette [TRE 96] pp. 116-121, cet exercice sous cette forme (concordances à partir de mots non liés) est plutôt destiné aux apprenants de niveau élémentaire. Les activités s'adressant à des niveaux plus élevés (intermédiaire à avancé) doivent mettre en jeu des relations linguistiques qui permettent d'explorer le lexique de manière plus systématique. Il s'agit des relations sémantiques (synonymie, antonymie, hyperonymie, actance), des relations de dérivation (pour passer d'une catégorie grammaticale à une autre), des rapports des co-occurrences ou collocationnels et enfin des valeurs stylistiques et pragmatiques.

Il semble donc judicieux de reprendre l'exercice vu plus haut et de l'améliorer de manière à faire intervenir ces relations. Le principe de recontextualisation étant toujours le même, les relations interviennent dans le choix des vocables à retrouver qui ne sont plus pris indépendamment les uns des autres. L'extraction des vocables reliés se fait par l'exploration du dictionnaire général.

La figure 6 montre un exercice de recontextualisation sur une relation d'actance (les actants sont les participants de l'action exprimée, voir énoncé de la figure 6). A partir du vocable de départ *étudier* (choisi dans le dictionnaire personnel pour son niveau de traitement), le système calcule les actants déclarés à cette entrée dans la base de données lexicales, recherche les concordances contenant ces vocables et génère l'activité (version avec formes canoniques) :

Utilisez les mots suivants, en les adaptant au contexte, pour compléter les phrases :

discipline – étude – étudiant – étudier – matière - université

1) Soit qu'ils _____, soit qu'ils soient au chômage.

2) Cette opération impose des _____ qui sont en cours et une validation scientifique incontestable.

3) Philippe, un _____ de 22 ans, ne gardera pas un très bon souvenir de son séjour en Angleterre.

4) Je me lève tôt, je m'habille correctement : il faut se donner une _____, structurer sa vie.

5) Diplômé de Supélec, je n'avais aucune compétence en _____ fiscale, marketing ou encore commerciale.

6) La persévérance de Vincent Decloitre, jeune diplômé de l'_____ en Sciences Eco, a payé.

Figure 6. Relation d'actance dans une activité de recontextualisation

Dans cet exercice, l'apprenant renforce les liens dans son lexique mental entre les vocables « satellites » (les actants) d'*étudier* en réfléchissant (travail en profondeur) sur les relations et les sens propres de ces vocables.

Cette activité peut se décliner en d'autres versions en changeant la relation lexicale. Ainsi, une version peut faire intervenir une relation de dérivation (l'énoncé indique le vocable, l'apprenant doit retrouver les mots de la même famille lexicale et les replacer dans les phrases), ou bien une relation de synonymie (les vocables à replacer sont synonymes les uns des autres).

6 Préparation des textes

Pour pouvoir proposer les activités décrites ci-dessus, le système doit être capable de générer des concordances qui contiennent les vocables à retrouver. Ceci n'est pas possible à partir des textes bruts, comme le font la plupart des concordanceurs, car il est nécessaire de pouvoir repérer les vocables dans les textes indépendamment de leur flexion. Il faut donc procéder au préalable à une préparation des textes.

Celle-ci comporte trois étapes :

- L'étiquetage syntaxique des textes : cette opération, qui consiste à assigner une catégorie grammaticale à chaque suite de caractères reconnu comme un mot, permet de lever un certain nombre d'ambiguïtés de manière automatique. L'état de l'art correspond à un taux d'erreur de 5 %, ce qui n'est malgré tout pas négligeable, car cela signifie qu'un mot sur vingt est mal étiqueté (soit en moyenne un par phrase). Toutefois les erreurs peuvent parfois provenir de cas difficiles (ambiguïté entre un participe passé et un adjectif) et comme il n'est pas sûr qu'un étiquetage manuel produise moins d'erreur, ce taux reste satisfaisant dans le cadre des activités. L'étiquetage d'un corpus n'est pas une fin en soi, mais facilite les traitements futurs et les rend plus efficaces. Ainsi, le système ne travaille plus sur des suites de caractères, mais sur des vocables, reconnus en tant que tels. Par ailleurs, l'étiquetage permet de lemmatiser les textes, ce qui assure la correspondance avec les entrées du dictionnaire général et du dictionnaire personnalisé.
- Le balisage et l'indexation des textes : les textes n'étant qu'une suite de caractères, il convient d'en isoler les phrases, voire les paragraphes. Le découpage en phrases est effectué par l'étiqueteur sur la base des ponctuations et des majuscules, tandis que la distinction des paragraphes se fait en fonction des retours chariot. Pour plus d'efficacité dans la production de concordances, il faut ensuite indexer les mots des textes, c'est-à-dire conserver leur position et celle des phrases qui les contiennent dans une base de données, à la manière des moteurs de recherche sur Internet.
- L'étiquetage sémantique des textes : pour pouvoir exploiter les relations sémantiques du dictionnaire, il est nécessaire de pouvoir déterminer les lexies dans les textes. En effet, les réseaux du dictionnaire relient les mots de sens à sens, de lexie à lexie (par exemple emploi sens 2 et travail sens 2a, tous deux relatifs à l'activité qui permet de gagner sa vie, sont déclarés synonymes), et il faut donc s'assurer de la lexie d'un vocable avant de proposer un synonyme ou un actant. C'est là un des points faibles du système, car contrairement à l'étiquetage syntaxique, la désambiguïsation sémantique ne peut se faire automatiquement. Les programmes qui lèvent les ambiguïtés ne sont pas assez fiables. Dès lors, le travail doit être fait à la main, ce qui est extrêmement coûteux. Cependant, une particularité du corpus rend le travail moins fastidieux. Du fait que les textes ont été

sélectionnés dans un thème donné, la distribution des lexies est différente de celle d'un corpus plus général. Les occurrences des vocables polysémiques caractéristiques du domaine auront presque toujours le sens correspondant. C'est le cas de cadre qui est presque toujours dans notre corpus un dirigeant, ou encore de boîte qui est la plupart du temps une entreprise. Néanmoins, si cette approximation n'est pas tolérable, la solution consiste à s'acheminer vers une génération semi-automatique des activités, dans laquelle on isole, par décision pédagogique, les vocables à apprendre, et on prépare un certain nombre de concordances (par exemple une vingtaine) pour chacun d'entre eux.

7 Les activités dans ALEXIA

Voyons à présent les caractéristiques informatiques des activités de recontextualisation telles qu'elles ont été définies plus haut, et les problèmes que leur implémentation soulève.

En effet, contrairement aux exercices conçus sur le papier, la génération automatique de ces activités à partir de matériaux bruts et authentiques, les textes du corpus, amène à les spécifier de manière précise. L'enseignant rédigeant les épreuves fait abstraction de beaucoup d'éléments que nous ne pouvons nous empêcher de considérer. L'ALAO impose donc une vue plus générale et plus complète de la conception d'activités.

Les points à étudier et les problèmes à résoudre s'articulent autour de cinq points :

- le choix des vocables/lexies à partir desquels les concordances vont être générées : s'agit-il en effet de vocables ou de lexies, comment ceux-ci ou celles-ci sont-elles déterminées par le système ?
- l'affichage des concordances : les concordances doivent-elles être des phrases ou des contextes ? Que faire si la phrase est trop courte ou trop longue, si le contexte n'est pas suffisant ?
- l'aide proposée : quel type d'aide ? Sera-t-elle la même en fonction de toutes les variantes ?
- l'acceptation des réponses : demande-t-on, dans le cas d'exercice à trous, de fléchir les formes à retrouver en fonction du contexte ? Si oui, a-t-on les moyens de vérifier automatiquement leur correction ou leur semi-correction ?
- la notation : une réponse est-elle juste ou fausse ? Ou peut-on délimiter des réponses partiellement justes ?

Nous nous proposons maintenant de développer et d'étudier les points ci-dessus.

7.1 Détermination des vocables pour les activités

Ces vocables sont retenus en fonction de leur statut dans le dictionnaire personnel. On peut discerner trois groupes. Il y a, d'une part, les vocables sur lesquels peu de travail a été effectué. En ce sens, il est légitime de penser que l'apprenant ne connaît pas suffisamment d'informations sur eux et qu'il n'est donc pas nécessaire de déclencher des activités. Cela peut signifier aussi que l'apprenant connaît déjà bien le vocable et qu'il n'a pas besoin des connaissances des ressources lexicales. Cependant, il est peu probable que ces vocables soient sélectionnés dans les textes, car ils présentent peu d'intérêt du point de vue de l'apprentissage. Nous considérons donc qu'un faible travail implique une connaissance insuffisante et nous écarterons dès lors ces vocables pour les activités. D'autre part, à l'opposé, il y a ceux sur lesquels beaucoup de travail a été effectué et qui ont reçu une bonne note lors d'une activité précédente. Le système se doit donc de les écarter en les considérant comme acquis. Entre ces deux groupes se trouvent les vocables avec un statut intermédiaire, suffisamment travaillés, mais pas encore maîtrisés. Ce sont ceux-là que le système retient et sur lesquels porteront les activités lexicales.

Concernant les activités portant sur les lexies (activités de recontextualisation avec relation de synonymie, d'actance ou de dérivation), le dictionnaire personnalisé ne comportant que des vocables (il n'y a pas de lien avec les lexies du dictionnaire), il appartient au système de choisir la lexie du vocable sur laquelle portera l'activité. Ce choix se fait en comparant les membres du groupe auxquels appartient le vocable en question avec les informations qui lui sont associées dans la base lexicale. Si parmi les mots associés à emploi dans le dictionnaire personnel, on retrouve travail, le système saura, par examen dans la base lexicale des mots associés aux lexies d'emploi, qu'il faut faire porter l'activité sur la lexie exprimant le sens de gagner sa vie.

7.2 L'affichage des concordances

A partir de l'ensemble des lexies/vocables (suivant le cas) sélectionnés, le système affiche les concordances et occulte les éléments à retrouver. L'apprenant doit taper la bonne forme dans le champ situé à droite des phrases (figure 7). Lorsqu'il valide sa réponse en tapant sur la touche *retour-chariot*, le système vérifie qu'elle fait bien partie des items de départ et enlève celle-ci de la liste. En cas de modification, ou bien de faute de frappe, l'utilisateur a la possibilité de corriger sa réponse directement dans le champ s'il n'a pas validé, ou bien en cliquant dans le champ avec la touche *option* enfoncée s'il a validé. Dans ce dernier cas, l'item est replacé dans la liste de départ. Si trop de modifications sont nécessaires, l'utilisateur peut remettre tous les champs à vide en appuyant sur le bouton « Tout effacer ». En appuyant sur le bouton figurant une ampoule éclairée, l'apprenant peut recevoir de l'aide. Lorsqu'il estime avoir trouvé toutes les réponses, il les valide en totalité avec le bouton « Valider ». Le système met alors en gras celles qui sont justes, établit une

note et expose les résultats. S'il y a des erreurs, l'apprenant peut essayer de les rectifier, jusqu'au sans-faute, mais la note calculée à la première validation reste la même.

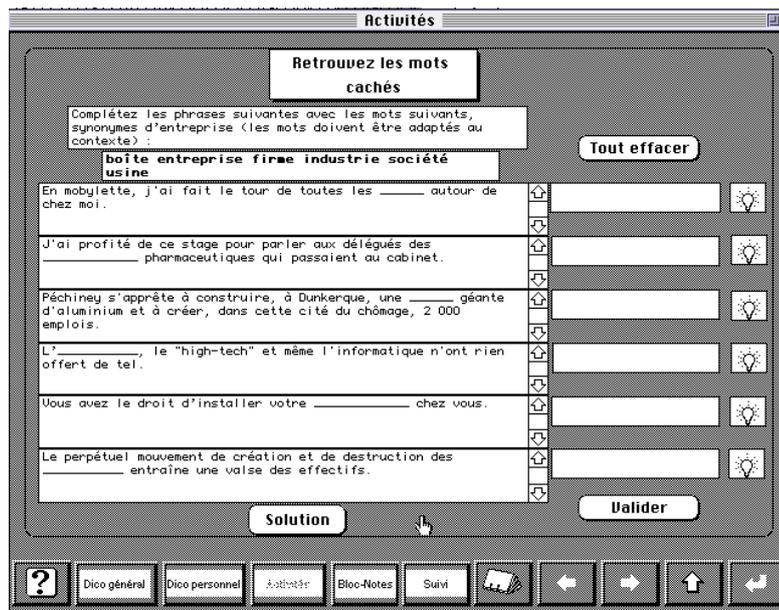


Figure 7. Affichage des concordances dans un exercice de recontextualisation (synonymie)

L'occultation des lexies est facilitée du fait de l'indexation de la forme fléchie avec sa forme canonique dans le fichier des concordances. Il est aussi possible, grâce aux indications de positionnement de l'occurrence dans la phrase, de s'affranchir de divers petits traitements préalables pour une analyse de chaque vocable avec l'analyseur morphologique (élimination puis remplacement des apostrophes ou des signes de ponctuation). Néanmoins, le problème principal ne se résume pas à ces petites tracasseries informatiques, mais concerne le contexte entourant la lexie à retrouver.

En effet, les concordances affichées sont sélectionnées aléatoirement parmi celles du vocable disponible. Dès lors, le système peut ne pas avoir la main heureuse en retenant des phrases trop courtes, possédant un contexte trop pauvre (comme le choix d'un titre tel que « Le chômage en France »), ou au contraire, des phrases trop longues, dans lesquelles la lexie joue un rôle minime. L'apprenant est alors confronté à une masse de texte inutile. Inutile ? Peut-être pas autant qu'on pourrait le croire. Car c'est peut-être ce contexte qui va lui permettre de déduire le vocable à

retrouver. Dans les deux cas, on s'aperçoit qu'il faut donner à l'apprenant la possibilité de pouvoir gérer le contexte à afficher, en lui permettant de le réduire ou au contraire de l'augmenter, voire carrément de changer la concordance si la phrase s'avère inexploitable. On touche ici le problème crucial de devoir travailler à partir de données réelles et non contrôlées, par rapport à des phrases qui auraient été pensées par un enseignant.

Concrètement, le système affiche un contexte avant et après le vocable de 70 caractères, en s'arrêtant aux limites de la phrases si celles-ci sont atteintes. Au-delà de 70 caractères, la phrase est coupée par défaut. En cliquant sur le champ contenant la concordance, si la phrase initiale est trop longue, le système alterne la phrase affichée par défaut avec la phrase entière. En cliquant sur le champ avec la touche *option* enfoncée, une nouvelle concordance est générée avec le même vocable.

7.3 L'aide

L'aide consiste à donner des éléments d'information partiels mais suffisamment significatifs sur la forme ou le sens d'un vocable, de manière à faciliter les inférences et à aiguiller l'apprenant vers la bonne solution. Ces informations sont tirées des connaissances que possède le système, connaissances disséminées en différents endroits : synonymes et définitions (avec le vocable occulté) dans le dictionnaire général, nom du groupe ainsi que les autres éléments en faisant partie pour le dictionnaire personnel, premières lettres ou longueur du vocable. Le registre peut également servir d'aide dans le cas où l'apprenant aurait trouvé le sens effectif de l'item, mais au registre près. Par contre, du fait de leur caractère libre et non contrôlé, les annotations ne peuvent pas servir d'aide : l'apprenant pourrait y trouver des informations, comme une traduction, pouvant le mettre sur la voie immédiatement, sans avoir besoin de faire des inférences.

Les aides doivent être adaptées à l'activité, car il est inutile par exemple de donner les premières lettres d'un vocable à retrouver si celui-ci figure dans l'énoncé (ce serait trop facile) ou bien il est superflu de donner un synonyme dans un exercice portant sur la synonymie (car ils figureront aussi dans l'énoncé). Dès lors, il faut examiner chacune des variantes des activités de recontextualisation et décider quelle aide est appropriée ou non.

Les différentes informations sont accessibles dans une fenêtre supplémentaire associée à chacune des concordances où l'utilisateur peut demander l'élément d'aide voulue en fonction de ceux disponibles (figure 8).



Figure 8. Indices d'aide pour l'item 1 de la phrase 1 (boîte, figure 7)

Les boutons d'éléments d'aide non disponibles sont désactivés (grisés). Concernant les synonymes, le système en fournit un seul par demande (clic sur le bouton « Syn. »). La priorité sera donné aux quasi-synonymes, puis, s'il n'y en a pas, aux hyperonymes, aux hyponymes et aux synonymes intersectifs. La relation de synonymie est indiquée.

7.4 L'acceptation

Un des objectifs de ce module d'activités lexicales est de pouvoir adapter son diagnostic en fonction des réponses de l'apprenant. Contrairement aux premiers programmes limités techniquement qui ne pouvaient répondre que par vrai ou faux, les environnements sont maintenant capables de donner des avis plus nuancés. En effet, une réponse n'est pas toujours entièrement fautive et peut se décomposer en plusieurs éléments dont certains s'avèrent justes. C'est le cas par exemple des réponses dont le vocable (ou la lexie) est proche sémantiquement sans être toutefois celui (celle) attendu(e) par le système.

Les réponses partiellement justes sont fonction des énoncés et donc des activités. Celles-ci peuvent comporter des problèmes de flexion s'il est demandé d'adapter les vocables aux contextes des différentes phrases. Dans ce cas, une réponse où le vocable est correct mais mal fléchi ne peut être considérée comme entièrement fausse.

En outre, l'activité sur les synonymes présente quelques spécificités propres à la nature de la relation sémantique en jeu. En effet, les vocables de départ peuvent être quasi-synonymes et donc s'interchanger entre les phrases. De ce fait, une réponse partiellement juste peut être une réponse :

- où le vocable proposé est sémantiquement proche de celui à deviner (société à la place d'entreprise).
- où le sens est correct mais pas le registre (virer à la place de licencier).

Pour ces deux derniers cas, il est possible, grâce à l'analyseur morphologique intégré dans le système et aux réseaux lexicaux du dictionnaire général de savoir si une réponse est proche sémantiquement de celle demandée, en d'autres termes si l'apprenant a répondu par un synonyme à la place de la solution. Il suffit de vérifier, après lemmatisation, que réponse de l'apprenant et solution sont synonymes égaux à ceci près que, l'utilisateur ne pouvant proposer que des vocables et non des lexies, la vérification portera sur les synonymes de toutes les lexies de sa réponse pour voir si parmi elles se trouve la solution (par exemple, il faudra vérifier si, parmi tous les sens du vocable société, l'un d'eux a pour synonyme entreprise).

La problématique du registre est un cas particulier de la proximité sémantique vue ci-dessus. Il faudra cette fois vérifier si la réponse de l'apprenant possède un synonyme égal, mais avec un registre différent, qui se trouve être la bonne solution.

Le problème de la bonne (ou non) flexion est plus difficile à trancher, surtout en ce qui concerne les verbes. En l'absence d'analyseur syntaxique dans le système nous ne pouvons compter que sur quatre informations : la forme fléchie et canonique de la bonne réponse et la forme fléchie et canonique (par le biais de l'analyseur morphologique) de la réponse de l'apprenant. Le problème est que même si les deux formes canoniques sont identiques, ce qui indique que l'apprenant a trouvé le bon vocable, mais que sa réponse est fautive (mauvaise flexion), il faut absolument le lui signaler, car il est pédagogiquement peu recommandable de laisser faire des fautes sans au moins les signaler. Or, comment savoir si le vocable est bien fléchi en l'absence d'analyse syntaxique ?

On pourrait penser que la comparaison des deux formes fléchies peut suffire à résoudre ce problème. Ce n'est pas forcément le cas comme le montre cette phrase :

4) Mais sur la totalité des entreprises existant en France, seule une sur deux _____ plusieurs salariés.

où le vocable occulté est emploierait dans le texte d'où est extraite cette concordance. Or, en l'absence de contexte plus large, qui n'est d'ailleurs pas traitable par le système, emploie est une réponse tout à fait correcte. De même on peut trouver quantité de phrases où le temps du verbe, ou bien le mode, n'est pas

clairement déterminé par les vocables environnants, ce qui autorise plusieurs réponses possibles, toutes plus justes les unes que les autres. La solution adoptée est donc de vérifier que la personne et le nombre coïncident. Dans cet exemple, il suffirait que la solution soit le vocable *employer* et qu'il soit conjugué à la troisième personne du singulier. Il faut cependant garder à l'esprit que cette vérification n'est pas toujours suffisante. En effet, on peut se trouver confronté à des phénomènes de concordances de temps qui ne permettent pas des flexions à des temps ou des modes approximatifs. Nous espérons cependant que ces cas ne se produiront pas trop souvent. Si ce n'était pas le cas, il faudrait alors se résoudre soit à ne pas demander d'adapter les verbes au contexte en les fléchissant (pour les autres catégories grammaticales, la flexion ne pose pas les mêmes problèmes, puisqu'à chaque fois, une seule solution est possible), soit, l'incorporation d'un analyseur syntaxique n'étant pas suffisante pour un phénomène aussi précis que la concordance des temps, à signaler que le système n'est pas sûr de son diagnostic concernant le temps. Il constate que la personne et le nombre sont corrects, mais, les temps ne coïncidant pas, ne peut déterminer si la réponse est juste sur ce point précis.

Pour terminer sur ces questions de correction, nous abordons les problèmes de fautes de frappe. Il serait dommage d'indiquer comme totalement fausse une réponse contenant une faute de frappe (par exemple *boullit* à la place de *boulot*). Dès lors, celles-ci seront indiquées lors de la validation individuelle de chaque item si elles entrent dans les deux cas suivants : la réponse et la solution diffèrent soit par la permutation de deux lettres ou bien par le remplacement d'une lettre par une autre. Le cas des redoublements de consonnes est plus délicat à traiter, car il peut s'agir soit d'une coquille, soit de l'ignorance de la bonne orthographe. Nous prenons la décision arbitraire de considérer une réponse concernée par cela comme fausse.

7.5 La notation

Le score obtenu aux activités lexicales doit évidemment tenir compte des réponses partiellement justes suivant les critères déterminés ci-dessus. Il n'est donc pas question de comptabiliser par 1 ou 0 les réponses obtenues mais d'établir un barème en fonction de l'éloignement avec la solution et des possibilités de reconnaissance du système.

La notation est fonction de l'objectif pédagogique de chaque activité. Par exemple pour l'activité de recontextualisation avec relation d'actance, on peut donner 3 points si la réponse est bonne, 1 point si le vocable est trouvé mais mal fléchi, 0 si le vocable n'est pas trouvé.

L'utilisation de l'aide rentre également en compte pour la notation. Chaque indice n'a pas la même importance. Ainsi, une définition apparaît comme l'élément pouvant le plus aider. De ce fait, on peut proposer -0,5 point à chaque fois que l'aide est demandée et -1 point s'il s'agit d'une définition.

Les notes concernent chaque vocable. Elles s'ajoutent à son score dans le module activité lexicale du dictionnaire personnel. Ce barème est pour l'instant indicatif et

doit bien sûr être confirmé ou affiné par des données expérimentales (à ce jour, seul le dictionnaire de l'environnement a été expérimenté, ceci dans le cadre d'une évaluation formative, vu le faible nombre de participants), qui permettront entre autres de déterminer quelle type d'aide est le plus efficace en fonction de l'activité et quel est celui le plus demandé.

8 Conclusion

Nous avons vu dans cet article comment les ressources lexicales d'ALEXIA et les informations qu'elles comprennent pouvaient être exploitées pour la génération automatique d'activités lexicales à l'aide d'outils spécifiques au traitement automatique du langage naturel. Elles sont à la fois le but de l'environnement et en assurent la cohérence. Nous avons choisi des activités de recontextualisation produites à partir de phrases d'un corpus centré sur un domaine déterminé, tirant parti des relations lexicales contenues dans le dictionnaire du système et déclenchées à partir des informations du dictionnaire personnel, c'est-à-dire en fonction des vocables que le système estime suffisamment travaillés mais non encore maîtrisés. Les activités lexicales ainsi générées sont renouvelables et tiennent compte du travail de l'apprenant déjà accompli. Les limites se situent dans la reconnaissance automatique des lexies dans les textes, nécessaire pour pouvoir extraire des concordances en rapport avec le sens voulu (une solution consiste alors à se diriger vers une génération semi-automatique dans un cadre défini), et, dans certains cas peu fréquents, dans l'échec du diagnostic de correction. Les principes pédagogiques issus des travaux en psycholinguistique et en didactique des langues ont grandement influé sur la conception de l'environnement, notamment l'organisation du lexique mental qui a décidé de la structure du dictionnaire et l'importance dans l'apprentissage du processus de révision et de la stratégie d'inférence. Enfin, nous avons pu à nouveau mesurer combien l'autonomie, dans le cadre des environnements d'aide à l'apprentissage lexical de deuxième génération, était favorisée par l'incorporation de ressources lexicales adaptées.

9 Bibliographie

- [AIT 87] AITCHISON J., *Words in the mind*, Oxford, Blackwell.
- [BOG 94] BOGAARDS P., *Le Vocabulaire dans l'Apprentissage des Langues Étrangères*, Langues et Apprentissage des Langues, CREDIF, ENS St-Cloud, Hatier/Didier.
- [BOG 95] BOGAARDS P., « Dictionnaires et compréhension écrite », *Cahiers de Lexicologie* 67, 1995-2, pp. 37-53.
- [BOG 98] BOGAARDS P., « Des dictionnaires au service de l'apprentissage du français langue étrangère », *Cahiers de Lexicologie* 72, 1998-1, pp. 127-167.

- [CHA 95] CHANIER T., FOUQUERE C., ISSAC F., « AlexiA : Un environnement d'aide à l'apprentissage lexical du français langue seconde », Conférence *Environnements Interactifs d'Apprentissage avec Ordinateur* (EIAO'95), pp. 79-90, Paris, Eyrolles.
- [CHA 98] CHANIER T., SELVA T., « The ALEXIA system: The Use of Visual Representaion to Enhance Vocabulary Learning », *Computer Assisted Language Learning* 11, 5, pp. 498-521.
- [CLA 92] CLARKE M., « Vocabulary learning with and without computers - Some thoughts on a way forward », *Computer Assisted Language Learning* 5, 3, pp. 139-146.
- [CON 96] CONIAM D., « Using corpus word frequency data in the automatic generation of english language cloze test », Actes de *Teaching and Language Corpora '96*, Lancaster, pp. 29-43.
- [CON 97] CONIAM D., « A preliminary inquiry into using corpus word frequency data in the automatic generation of english language close tests », *CALICO* 14, 2-4, pp. 15-33.
- [DEC 93] DECOO W., « Lexical composition and morpho-syntactic variation in language textbooks - Computer-based approaches as another dimension of CALL, *Computer Assisted Language Learning* 6, 2, pp 123-144.
- [GOO 94] GOODFELLOW R., « A computer-based strategy for foreign language vocabulary learning », Unpublished PhD thesis, Institute of Educational Technology, Open University.
- [GOO 95] GOODFELLOW R., « A Review of Types of Programs for Vocabulary Instruction », *Computer Assisted Language Learning* 8, 2-3, pp. 205-226.
- [HIG 84] HIGGINS J., JOHNS T., *Computers in language learning*, Collins Educational.
- [HUG 89] HUGHES A., *Testing for language teachers*, Cambridge, Cambridge University Press.
- [HUL 93] HULSTIJN J. H., « When do foreign-language readers look up the meaning of unfamiliar words? The influence of task and learner variables », *The Modern Language Journal* 77 (2), pp. 139-147.
- [KEN 90] KENNING M.-M., « Computer assisted Language Learning », *Language Teaching*, Cambridge, Cambridge University Press, pp. 67-76.
- [KUK 88] KUKULSKA-HULME A., « A computerized interactive vocabulary development system for advanced learners », *System* 16 (2), pp 163-170.
- [MEL 92] MEL'CUK I., *Dictionnaire Explicatif et Combinatoire du français contemporain*, Recherche lexico-sémantique III, Montréal, Les Presses de l'Université de Montréal.
- [MOT 75] MOTHE J.-C., *L'évaluation par les tests dans la classe de français*, Paris, Hachette/Larousse.
- [NER 98] NERBONNE J., DOKTER D., SMIT P., « Morphological Processing and Computer-Assisted Language Learning », *Computer-Assisted Language Learning* 11, 5, pp. 543-559.
- [PAR 97] PARIBAKHT T. S., WESCHE M., « Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition », J. Coady, T. Huckin (Eds),

Second Language Vocabulary: a rationale for pedagogy, Cambridge, Cambridge University Press, pp 174-200.

[SEL 98] SELVA T., CHANIER T., « Apport de l'informatique pour l'accès lexical dans les dictionnaires pour apprenants : projet Alexia », *EUROpean Association for Lexicography* (EURALEX'98), Liège, Belgique, pp. 631-642.

[SEL 99] SELVA T., « Ressources et activités pédagogiques dans un environnement informatique d'aide à l'apprentissage lexical du français langue seconde », Thèse d'Université, Université de Franche-Comté, octobre 1999, 210 pages. <http://lifc.univ-fcomte.fr/RECHERCHE/P7/pub/theseselva.pdf>

[SUS 94] SUSSEX R., CUMMING G., Cropp S., « A tools-based environment for discovery-oriented CALL - Cognitive, pedagogical and ergonomic issues for interactive learning », *Computer Assisted Language Learning* 7, 2, pp. 133-149.

[TRE 96] TREVILLE M.-C., DUQUETTE L., *Enseigner le vocabulaire en classe de langue*, Paris, Hachette.

Sites Internet

[PAS 00] LA PASSERELLE, *The Half-Baked Interactive Tests*, consulté en juin 2000 : <http://www.lapasserelle.com/lm/pagespeciales/half.baked/halfbakedtests.index.html>

[SIL 00] SILFIDE, Serveur Interactif pour la Langue Française, son Identité, sa Diffusion et son Etude, consulté en juin 2000 : <http://www.loria.fr/Projet/Silfide/>

Thierry Selva est post-doctorant au sein du GRELEP (Groupe de Recherche En Lexicographie Pédagogique, www.kuleuven.ac.be/grelep) à l'Université Catholique de Leuven (Louvain) depuis novembre 1999, peu après la soutenance de son doctorat en informatique à l'université de Franche-Comté [SEL 99].

Il travaille actuellement sur le projet DAFLES (Dictionnaire d'Apprentissage du Français Langue Étrangère ou Seconde), un dictionnaire électronique interactif en ligne pour apprenants du français. La deuxième année post-doctorale sera consacrée à la conception d'un environnement d'apprentissage lexical en ligne.

Thierry Chanier est professeur à l'université de Franche-Comté où il anime l'équipe de recherche SICAH (systèmes d'information et de communication pour l'aide à l'apprentissage humain) du laboratoire d'informatique LIFC (lifc.univ-fcomte.fr).

Après des études de doctorat en traitement automatique du langage, il a concentré ses recherches dans les systèmes d'aide à l'apprentissage des langues. Dans un passé récent, il a organisé le congrès européen du domaine (EUROCALL'99) et est rédacteur en chef de la revue ALSIC (apprentissage des langues et systèmes d'information et de communication : alsic.org). Sa thématique principale de recherche est aujourd'hui tournée vers l'apprentissage collaboratif dans les systèmes de formation à distance.