

Auto-apprentissage de l'anglais langue étrangère sur micro-ordinateurs ayant accès à des dictionnaires informatisés

J. NOEL et J. JANSEN

1. Dictionnaires automatiques et intelligence naturelle

Le titre de ma communication mérite quelques commentaires, qui vont me permettre de situer les travaux de notre équipe liégeoise vis-à-vis de l'E.A.O. (enseignement assisté par ordinateur) et de la traduction assistée. Celle-ci fait d'ailleurs partie de notre système puisque, outre des *aides* en anglais, il offre des équivalents de traduction en français, la langue maternelle de nos étudiants.

1.1. Auto-apprentissage de l'anglais langue étrangère

En considérant celui qui apprend une langue étrangère comme un *auto-didacte*, j'évite à dessein de prendre parti sur les diverses options théoriques et pratiques – les progressions pédagogiques notamment – qui se présentent dans l'enseignement et l'apprentissage des langues – étrangères et maternelles. Quelle doit être, notamment, la part du travail d'*interprétation* du langage à effectuer par l'apprenant et celle assumée par le système d'enseignement, de *transmission* des connaissances préexistantes, plus ou moins précises, consignées dans les *textes* (bruts), *dictionnaires* et *grammaires* (Douglas Barnes, 1976, cf. Leech in Leech & Candlin, eds., 1986, pp. xii-xviii)? Dans notre système, nous ne prenons pas parti, notre seul

souci étant de fournir toute espèce d'*exercices* et d'*aides* à l'apprenant, considéré comme le seul maître de son action, de sa réflexion, de son emploi du temps.

Il pourra disposer non seulement de collections de textes pratiquement illimitées mais aussi de toutes les aides théoriques et pratiques, notamment de tous les exemples d'usages linguistiques nécessaires. Il pourra les faire défiler à l'écran autant de fois qu'il le faut pour les fixer dans sa mémoire. Une bonne raison de ne pas prendre parti dans le dilemme *interprétation/transmission* tient à la nature même des formes et contenus du langage, qui est mélange de donné et de construit. Tout acte langagier *innove*, au sens transitif d'«innover» dans la définition du *Petit Robert* :

INNOVER. Introduire dans une chose établie (qqch. de nouveau, d'encore inconnu). V. Changer. *Innover une mode, une coiffure.* V. Inventer, trouver. "*Ne rien innover, telle est la loi du pays*" (BALZ.).

fig. 0

Définition de "innover" dans le *Petit Robert*

Dans son état actuel, notre système se présente comme une *chaîne de production, automatique ou semi-automatique* (selon les buts poursuivis), *d'exercices d'anglais à partir d'un texte quelconque dans lequel nous pratiquons des trous*. Superficiellement, nos exercices ressemblent aux tests de closure, bien connus des spécialistes. Ils servent comme eux à exercer la *compréhension à la lecture et la production (rédaction dirigée, "guided composition" en anglais)*. La figure 1 donne un exemple de texte troué, accompagné de la liste correspondante des mots extraits de ce texte, rangés par ordre alphabétique et numérotés de 1 à 45. Ce format, d'un maximum de 50 trous par page, nous permet de recourir, si nous le souhaitons, à une *correction automatique*, par le truchement de feuilles spéciales sur lesquelles les étudiants encodent le numéro de leur réponse à l'aide d'un crayon ordinaire, et qui sont lues par un lecteur optique.

La figure 2 présente une des feuilles utilisées dans le système STEP (Système de Traitement Automatique d'Évaluations Pédagogiques), développé dans notre université (Debot et Leclercq, Laplanche) et employé dans la plupart de nos facultés pour traiter massivement de grands nombres de tests – notamment à choix multiple.

[124 TEXT E19]

<WINES to =(1)= a stranger>

To =(2)= wines or cordials for food =(3)= from Moussaka to Mangoes and Artichokes to Uglies poses something of a problem. On the other hand, the dishes =(4)= upon these unusual fruits and vegetables have a common characteristic. The main course dishes are strongly flavoured and the fruits have a pleasant astringency behind their sweetness. This =(5)= that one should =(6)= =(7)= wines that are very dry or of delicate flavour. Sparkling wines do not really =(8)= into this picture, but strong, virile wines do. A full-bodied wine for the main course, and a sweet wine for the fruits =(9)= be a good general rule.

One =(10)= do well to =(11)= the wines of Greece to =(12)= the moussaka. The special one, of course, is Retsina, a white wine matured in resinated casks. I =(13)= to =(14)= it from =(15)= it in the war years, but the fact must be =(16)=, it is an acquired taste. My wife =(17)= that it =(18)= of ping-pong balls. It might be a good thing to have a bottle or two for the initiated, and =(19)= to a white dry Samos for the majority. The price of both is about 8/0s. 6/0d. to 9/0s. per bottle. They should be =(20)= cold but not over-chilled.

A good red alternative =(21)= from Hungary - Egri Bikaver - the dark wine from Eger =(22)= as Bull's Blood. It =(23)= about 10/0s. 6/0d. per bottle. If you =(24)= something really strong, with a high alcoholic content, Yugoslavia =(25)= Dinjac, but it is not a wine for weak heads. It =(26)= about 8/0d. per bottle.

There are also Spanish and Portuguese wines that =(27)= well with strongly flavoured foods. The robust Spanish Chablis, the Rioja Burgundy, and the Portuguese Vila Real Tinto are examples, all =(28)= about 7/0s. 6/0d. per bottle. Russian wines are now on the market, and the red Mukuzani No. 4 =(29)= =(30)= this menu well. It =(31)= about the same as the Spanish wines.

Now for the wines to =(32)= with the fruit dishes. For those who =(33)= a really sweet wine there is what is often =(34)= the Sauterne of the Perigord; Monbazillac, =(35)= from 8/0s. to 10/0s. per bottle. Visitors to Vienna will perhaps =(36)= the ancient cellars at Gumpoldskirchner - the home of the Rotgipfler Auslese. This =(37)= about 15/0s. 6/0d. per bottle. Or to =(38)= to Greece, there is the famous Mavrodaphne - the Italians =(39)= a lot of it, and you may have =(40)= it in Vienna.

An alternative is to =(41)= the "straight" wines out of the planning and =(42)= for some kind of a cup. Pride of porto is a pleasant and simple one. To =(43)= about thirty glasses =(44)= one bottle of tawny port, half a gill of Orange Curaçao, one lemon, and a siphon of soda. Squeeze the juice of the lemon into a bowl, =(45)= the port

1. ADDING	13. COSTS	31. MEET
2. AVOID	17. DRINK	32. MET
3. BUILT	18. DRINKING	33. PRODUCE
4. CALLED	19. FACED	34. PRODUCES
5. CAME	20. FIT	35. RANGING
6. CHOOSE	20. FIT	36. REMEMBER
7. CHOOSING	22. GO	37. RETURN
8. COMES	22. GO	38. SERVED
9. CONSIDER	22. GO	39. STICK
10. CONTENTS	25. KNOWN	40. TAKE
11. COSTING	26. LEAVE	41. TASTES
11. COSTING	27. LIKE	42. WANT
13. COSTS	28. LOVE	43. WOULD
13. COSTS	29. MATCH	43. WOULD
13. COSTS	30. MEANS	43. WOULD

fig. 1

Trouage du texte sur les vins : LOB E19

MAT 86
ZZ 1260
MATRICULE : 1CA00126

Université de Liège

S T E P

0166LAM

TEST NO 26 FEUILLE : 1

<table border="0" style="width: 100%;"> <tr><td style="border: 1px solid black; padding: 2px;">1</td><td style="border: 1px solid black; padding: 2px;">21</td><td style="border: 1px solid black; padding: 2px;">36</td></tr> <tr><td style="border: 1px solid black; padding: 2px;">2</td><td style="border: 1px solid black; padding: 2px;">22</td><td style="border: 1px solid black; padding: 2px;">37</td></tr> <tr><td style="border: 1px solid black; padding: 2px;">3</td><td style="border: 1px solid black; padding: 2px;">23</td><td style="border: 1px solid black; padding: 2px;">38</td></tr> <tr><td style="border: 1px solid black; padding: 2px;">4</td><td style="border: 1px solid black; padding: 2px;">24</td><td style="border: 1px solid black; padding: 2px;">39</td></tr> <tr><td style="border: 1px solid black; padding: 2px;">5</td><td style="border: 1px solid black; padding: 2px;">25</td><td style="border: 1px solid black; padding: 2px;">40</td></tr> <tr><td style="border: 1px solid black; padding: 2px;">6</td><td style="border: 1px solid black; padding: 2px;">26</td><td style="border: 1px solid black; padding: 2px;">41</td></tr> <tr><td style="border: 1px solid black; padding: 2px;">7</td><td style="border: 1px solid black; padding: 2px;">27</td><td style="border: 1px solid black; padding: 2px;">42</td></tr> <tr><td style="border: 1px solid black; padding: 2px;">8</td><td style="border: 1px solid black; padding: 2px;">28</td><td style="border: 1px solid black; padding: 2px;">43</td></tr> <tr><td style="border: 1px solid black; padding: 2px;">9</td><td style="border: 1px solid black; padding: 2px;">29</td><td style="border: 1px solid black; padding: 2px;">44</td></tr> <tr><td style="border: 1px solid black; padding: 2px;">10</td><td style="border: 1px solid black; padding: 2px;">30</td><td style="border: 1px solid black; padding: 2px;">45</td></tr> <tr><td style="border: 1px solid black; padding: 2px;">11</td><td style="border: 1px solid black; padding: 2px;">31</td><td style="border: 1px solid black; padding: 2px;">46</td></tr> <tr><td style="border: 1px solid black; padding: 2px;">12</td><td style="border: 1px solid black; padding: 2px;">32</td><td style="border: 1px solid black; padding: 2px;">47</td></tr> <tr><td style="border: 1px solid black; padding: 2px;">13</td><td style="border: 1px solid black; padding: 2px;">33</td><td style="border: 1px solid black; padding: 2px;">48</td></tr> <tr><td style="border: 1px solid black; padding: 2px;">14</td><td style="border: 1px solid black; padding: 2px;">34</td><td style="border: 1px solid black; padding: 2px;">49</td></tr> <tr><td style="border: 1px solid black; padding: 2px;">15</td><td style="border: 1px solid black; padding: 2px;">35</td><td style="border: 1px solid black; padding: 2px;">50</td></tr> <tr><td style="border: 1px solid black; padding: 2px;">16</td><td></td><td></td></tr> <tr><td style="border: 1px solid black; padding: 2px;">17</td><td></td><td></td></tr> <tr><td style="border: 1px solid black; padding: 2px;">18</td><td></td><td></td></tr> <tr><td style="border: 1px solid black; padding: 2px;">19</td><td></td><td></td></tr> <tr><td style="border: 1px solid black; padding: 2px;">20</td><td></td><td></td></tr> </table>	1	21	36	2	22	37	3	23	38	4	24	39	5	25	40	6	26	41	7	27	42	8	28	43	9	29	44	10	30	45	11	31	46	12	32	47	13	33	48	14	34	49	15	35	50	16			17			18			19			20			<div style="border: 1px solid black; padding: 5px; height: 120px; margin-bottom: 10px;"> <!-- Large empty box for answer writing --> </div> <table border="0" style="width: 100%;"> <tr><td style="border: 1px solid black; padding: 2px;">21</td><td style="border: 1px solid black; padding: 2px;">36</td></tr> <tr><td style="border: 1px solid black; padding: 2px;">22</td><td style="border: 1px solid black; padding: 2px;">37</td></tr> <tr><td style="border: 1px solid black; padding: 2px;">23</td><td style="border: 1px solid black; padding: 2px;">38</td></tr> <tr><td style="border: 1px solid black; padding: 2px;">24</td><td style="border: 1px solid black; padding: 2px;">39</td></tr> <tr><td style="border: 1px solid black; padding: 2px;">25</td><td style="border: 1px solid black; padding: 2px;">40</td></tr> <tr><td style="border: 1px solid black; padding: 2px;">26</td><td style="border: 1px solid black; padding: 2px;">41</td></tr> <tr><td style="border: 1px solid black; padding: 2px;">27</td><td style="border: 1px solid black; padding: 2px;">42</td></tr> <tr><td style="border: 1px solid black; padding: 2px;">28</td><td style="border: 1px solid black; padding: 2px;">43</td></tr> <tr><td style="border: 1px solid black; padding: 2px;">29</td><td style="border: 1px solid black; padding: 2px;">44</td></tr> <tr><td style="border: 1px solid black; padding: 2px;">30</td><td style="border: 1px solid black; padding: 2px;">45</td></tr> <tr><td style="border: 1px solid black; padding: 2px;">31</td><td style="border: 1px solid black; padding: 2px;">46</td></tr> <tr><td style="border: 1px solid black; padding: 2px;">32</td><td style="border: 1px solid black; padding: 2px;">47</td></tr> <tr><td style="border: 1px solid black; padding: 2px;">33</td><td style="border: 1px solid black; padding: 2px;">48</td></tr> <tr><td style="border: 1px solid black; padding: 2px;">34</td><td style="border: 1px solid black; padding: 2px;">49</td></tr> <tr><td style="border: 1px solid black; padding: 2px;">35</td><td style="border: 1px solid black; padding: 2px;">50</td></tr> </table>	21	36	22	37	23	38	24	39	25	40	26	41	27	42	28	43	29	44	30	45	31	46	32	47	33	48	34	49	35	50	<table border="0" style="width: 100%;"> <tr><td style="border: 1px solid black; padding: 2px;">36</td><td style="border: 1px solid black; padding: 2px;">41</td></tr> <tr><td style="border: 1px solid black; padding: 2px;">37</td><td style="border: 1px solid black; padding: 2px;">42</td></tr> <tr><td style="border: 1px solid black; padding: 2px;">38</td><td style="border: 1px solid black; padding: 2px;">43</td></tr> <tr><td style="border: 1px solid black; padding: 2px;">39</td><td style="border: 1px solid black; padding: 2px;">44</td></tr> <tr><td style="border: 1px solid black; padding: 2px;">40</td><td style="border: 1px solid black; padding: 2px;">45</td></tr> <tr><td style="border: 1px solid black; padding: 2px;">41</td><td style="border: 1px solid black; padding: 2px;">46</td></tr> <tr><td style="border: 1px solid black; padding: 2px;">42</td><td style="border: 1px solid black; padding: 2px;">47</td></tr> <tr><td style="border: 1px solid black; padding: 2px;">43</td><td style="border: 1px solid black; padding: 2px;">48</td></tr> <tr><td style="border: 1px solid black; padding: 2px;">44</td><td style="border: 1px solid black; padding: 2px;">49</td></tr> <tr><td style="border: 1px solid black; padding: 2px;">45</td><td style="border: 1px solid black; padding: 2px;">50</td></tr> <tr><td style="border: 1px solid black; padding: 2px;">46</td><td></td></tr> <tr><td style="border: 1px solid black; padding: 2px;">47</td><td></td></tr> <tr><td style="border: 1px solid black; padding: 2px;">48</td><td></td></tr> <tr><td style="border: 1px solid black; padding: 2px;">49</td><td></td></tr> <tr><td style="border: 1px solid black; padding: 2px;">50</td><td></td></tr> </table>	36	41	37	42	38	43	39	44	40	45	41	46	42	47	43	48	44	49	45	50	46		47		48		49		50	
1	21	36																																																																																																																								
2	22	37																																																																																																																								
3	23	38																																																																																																																								
4	24	39																																																																																																																								
5	25	40																																																																																																																								
6	26	41																																																																																																																								
7	27	42																																																																																																																								
8	28	43																																																																																																																								
9	29	44																																																																																																																								
10	30	45																																																																																																																								
11	31	46																																																																																																																								
12	32	47																																																																																																																								
13	33	48																																																																																																																								
14	34	49																																																																																																																								
15	35	50																																																																																																																								
16																																																																																																																										
17																																																																																																																										
18																																																																																																																										
19																																																																																																																										
20																																																																																																																										
21	36																																																																																																																									
22	37																																																																																																																									
23	38																																																																																																																									
24	39																																																																																																																									
25	40																																																																																																																									
26	41																																																																																																																									
27	42																																																																																																																									
28	43																																																																																																																									
29	44																																																																																																																									
30	45																																																																																																																									
31	46																																																																																																																									
32	47																																																																																																																									
33	48																																																																																																																									
34	49																																																																																																																									
35	50																																																																																																																									
36	41																																																																																																																									
37	42																																																																																																																									
38	43																																																																																																																									
39	44																																																																																																																									
40	45																																																																																																																									
41	46																																																																																																																									
42	47																																																																																																																									
43	48																																																																																																																									
44	49																																																																																																																									
45	50																																																																																																																									
46																																																																																																																										
47																																																																																																																										
48																																																																																																																										
49																																																																																																																										
50																																																																																																																										

fig. 2
Feuille STEP

Jusqu'ici rien de neuf, en apparence, qui justifierait mon intervention dans cette séance de notre Congrès, consacrée aux *innovations*.

La nouveauté de notre démarche n'est pas visible à l'oeil nu dans notre exemple. Comme je vais l'expliquer, elle réside dans le rôle joué par divers *gros dictionnaires informatisés dans la production de nos exercices d'anglais selon des critères et clés d'accès contenus dans ces dictionnaires*. Notre objectif de recherche, qui est très ambitieux, est d'appliquer un dictionnaire quelconque à un texte quelconque.

1.2. A l'aide de micro-ordinateurs

Comme je viens de l'indiquer, notre prototype, dans son état actuel, n'offre pas un système de logiciels et didacticiels intégré et interactif, comme il en existe sur de petits micros bas de gamme pour le CALL - *Computer-Assisted Language Learning* - dans le domaine de l'anglais (Higgings & Johns, 1984; Michiels, 1986; Leech & Candlin, *op. cit.*). Nous ne disposons d'ailleurs pour l'instant que d'un seul micro-ordinateur avec 40 MB de mémoire à accès rapide, en liaison permanente avec un système central, où nous disposons aussi d'une capacité de mémoire équivalente. Pour les principales opérations de recherches sur nos gros dictionnaires et la plupart des traitements de leurs (sous-)produits évoqués ici, nous restons tributaires du système central, et notamment de son logiciel d'interrogation (le STAIRS). Le transfert d'un de nos dictionnaires (le *Longman*) sur notre micro-ordinateur (sous DBase III) est en cours. Pour ce qui est des divers traitements que nous faisons subir aux textes, par contre, nous avons opté d'emblée pour des programmes (en TURBOPASCAL) exécutés sur le micro-ordinateur (un *IBM PC AT*). Il en va de même pour l'application à des textes de divers sous-fichiers extraits de nos dictionnaires.

Le principal fichier de textes que nous utilisons, qui comprend plus d'un million de mots, est le recueil ou corpus LOB (Lancaster-Oslo-Bergen). Cet échantillonnage de textes *authentiques*, bien connu des anglicistes est l'équivalent, pour l'anglais britannique, du célèbre corpus BROWN de l'anglais américain; premier grand recueil de textes anglais informatisés à avoir été conçu en tant qu'échantillonnage représentatif des divers types de

textes existant dans cette langue, le BROWN Corpus est aussi très largement diffusé, au prix d'un livre de diffusion moyenne (de l'ordre de 100 \$).

1.3. *Ayant accès à des dictionnaires informatisés*

Les deux dictionnaires informatisés dont il sera principalement question ici sont le *Longman (Longman Dictionary of Contemporary English, 1978)*, monolingue anglais, et le *Collins & Robert Dictionnaire français-anglais / English-French Dictionary*, également paru en 1978 dans sa version imprimée. Nous disposons des fichiers informatiques de ces dictionnaires en vertu de contrats de recherche que nous avons signés avec les maisons d'édition (respectivement en 1979 et en 1985).

Avant d'entrer dans le détail, bornons-nous ici à souligner d'emblée la grande richesse des informations contenues dans la plupart des gros dictionnaires imprimés traditionnels; les ouvrages que je viens de citer n'en sont que deux bons exemples. C'est cette richesse qui explique à la fois les difficultés des exploitations autres que l'impression d'un livre – destination première de ces fichiers – et la variété de ces exploitations dans pratiquement tous les domaines qui touchent au traitement des langues (Amstler, 1984). Pour rendre ces exploitations possibles, on doit se livrer, si possible par programme, à divers travaux longs et fastidieux de nettoyage, de réécriture et de restructuration (entrepris au sein de notre équipe par notre collègue Jacques Jansen).

Je me bornerai dans ce qui suit à présenter le principal résultat de ces travaux : nos bases de données *Longman* et *Collins & Robert*. Contrairement aux dictionnaires imprimés, qui rangent l'information dans le seul ordre alphabétique des entrées, nos bases de données nous donnent, grâce à un logiciel d'interrogation (STAIRS, DBase), un *accès sélectif et instantané* à n'importe quelle information contenue dans nos fichiers dictionnaires, c'est-à-dire à n'importe quel mot ou signe, avec le contexte choisi, qui répondent à une interrogation donnée. Dans nos fichiers STAIRS, ce sont les définitions de dictionnaires, avec toutes les informations qui y sont associées (y compris, bien sûr, les mots-vedettes correspondants), qui sont considérés comme les "documents" dont le texte intégral ou une partie peuvent répondre à une question posée.

Comme dans tout système d'interrogation, il importe absolument – pour éviter le "silence" (l'absence de réponse) comme le "bruit" (trop de réponses non pertinentes) – de *structurer* l'information de nos dictionnaires selon les différents *types d'informations* à distinguer, qui déterminent les divers critères de *sélection* que l'on veut se donner. Dans la suite de cet exposé, nous allons présenter ces aspects fondamentaux de notre projet sous les quatre rubriques suivantes :

1. les *types d'informations*, extraites de nos bases de données lexicales,
2. la *sélection des textes*,
3. la *sélection des trous*,
4. la *sélection des aides*.

2. Types d'informations

Comme nos exploitations du *Collins & Robert* peuvent être plus utilement envisagées au chapitre des *aides*, consacrons-nous ici au *Longman*, dont voici quelques caractéristiques générales :

- C'est un dictionnaire anglais-anglais de taille moyenne, en un volume, contenant quelque :
 - ♦ 40 000 entrées,
 - ♦ 75 000 définitions,
 - ♦ 46 000 exemples.
- Il est spécialement destiné à l'apprenant, et d'ailleurs sous-titré "*A Learner's Dictionary*".
- Comme d'autres dictionnaires, il associe aux mots et locutions de l'anglais divers codes et étiquettes linguistiques que je me borne à énumérer : codes morphologiques, codes sémantiques (ceux-ci n'apparaissent pas dans le dictionnaire imprimé), étiquettes pragmatiques (géographiques, niveaux de langue, attitudes : péjoratif, etc.) et, bien sûr, les parties du discours (nom, verbe, etc.)

Mais c'est à quatre principaux types d'informations que je voudrais m'attacher parce qu'ils constituent l'originalité de ce dictionnaire. Nous examinerons successivement :

- a) les *exemples et définitions*,
- b) les *codes grammaticaux*,

- c) les *codes matières* qui, comme les codes sémantiques, n'apparaissent pas dans le dictionnaire imprimé mais qui sont présents dans le fichier informatique. L'intérêt de ces types d'informations n'est pas seulement intrinsèque mais réside évidemment aussi dans la possibilité de *combiner* les divers types d'informations associées dans les entrées lexicales pour des recherches et *sélections* à ce niveau ou sur l'ensemble du fichier.

2.1. Exemples et définitions

Les rédacteurs du dictionnaire se sont imposé un *vocabulaire contrôlé* de quelque 2 000 mots, préfixes et suffixes, pour écrire *toutes* les définitions et *tous* les exemples. Les exceptions sont clairement repérables : mots-vedettes n'appartenant pas au vocabulaire contrôlé dans les exemples, rares mots en majuscules, dans les définitions, renvoyant à l'entrée qui les décrit.

L'emploi généralisé d'un vocabulaire contrôlé fait du fichier *Longman* une des *bases de connaissance* à mes yeux les plus intéressantes jamais réalisées pour une langue quelconque et en particulier pour l'anglais, en ce que ce vocabulaire contrôlé omniprésent établit entre tous les éléments de la langue et tous les codes associés des réseaux de relation d'une richesse unique *tout en facilitant l'accès à toute l'information, dans une base de données, par sa nature même de langage d'interrogation limité et explicitement formulé*. En un mot, ce vocabulaire minimal permet un accès maximal et sélectif. Précisons maintenant les avantages que le vocabulaire contrôlé présente pour *l'apprenant*.

Les avantages *linguistiques*, tout d'abord. En premier lieu, il permet de multiplier considérablement le nombre d'*exemples* d'emplois du vocabulaire de base, puisqu'il apparaît dans tous les exemples du dictionnaire. Nous n'avons pas manqué, à partir de recherches STAIRS sur certains mots de base, d'extraire tous les exemples où figurent ces mots afin de les présenter à nos étudiants. Les figures 3 et 4 montrent comment nous avons illustré l'emploi d'un mot de base (*seem*) et un point de grammaire traditionnelle (les auxiliaires) par des concordances de type KWIC et des listes d'exemples produites à partir de recherches STAIRS sur les exemples du *Longman*.

HEADNUM	DEF	ND+NI=NT	EXA	HEADWORD	POS	WORD
=====	===	=====	===	=====	=====	=====
A0011600	D02	03+01=04	E01	abstract	adj	SEEM
Your ideas on this matter seem a little abstract.						
A0025900	D03	07+00=07	E01	act	v	SEEMS
I can't take her seriously because she always seems to be acting						
A0158800	D08	09+08=17	E01	as	conj	SEEMS
Improbable as it seems, it's true						
A0162500	D03	04+01=05	E01	as if	conj	SEEMED
It seemed as if the night would never end.						
A0175600	D03	14+01=15	E08	at	prep	SEEM
(infml) I'm leaving you, dearest, because#how shall I put it ?						
#You always seem to be talking at me rather than to me.						
A0175700	D01	02+01=03	E02	at all	adv	SEEM
He doesn't seem at all interested in my plan						
A0196400	D00	01+00=01	E01	aversive	adj	SEEM
aversive training to get rid of bad habits by making them seem unpleasant						
B0000700	D04	04+00=04	E01	babble	n	SEEM
What's that babble of running water I seem to hear ? #compare BURBLE						
B0025100	D04	04+00=04	E01	bankrupt	adj	SEEM
They seem to be bankrupt of all kind feelings						
B0060000	D02	06+09=15	E01	bed	n	SEEM
What's wrong with young people these days ?						
They seem to think of nothing but bed !						
B0066800	D05	05+01=06	E01	beg	v	SEEMS
Your plan is interesting, but it seems to beg the real difficulties						
B0080350	D00	01+00=01	E01	bent on	adj	SEEMS
Jim seems bent on becoming a musician						
B0087400	D04	04+00=04	E01	betray	v	SEEM
He tried to seem angry, but his smile betrayed him/ but he betrayed himself by smiling						
B0093500	I00	00+01=01	E01	bide	v	SEEMS
He seems to be doing nothing, but really he's just biding his time						
B0117200	D02	02+00=02	E02	blank	adj	SEEMED
Every day seemed blank and meaningless						
B0122000	D03	04+00=04	E01	blend	v	SEEM
These houses seem to blend well with the trees and the COUNTRYSIDE.						
B0122000	D04	04+00=04	E01	blend	v	SEEM
These houses seem to blend into the COUNTRYSIDE.						
B0122000	D04	04+00=04	E02	blend	v	SEEM
The COUNTRYSIDE and the houses seem to blend (into each other)						
B0124100	D00	01+00=01	E02	blind alley	n	SEEMED
(fig.) We tried one idea after another but they all seemed to be blind alleys						
B0140000	I02	01+03=04	E01	blush	n	SEEMED
It seemed a good idea at first blush, but there were several faults in it						
B0164900	D02	03+01=04	E01	borrow	v	SEEMS
Somebody seems to have borrowed my watch when I wasn't looking						

fig. 3

Extrait d'un KWIC d'exemples sur des variantes du verbe *seem*, tiré du LDOCE

- MOD - They say that it will be good weather tomorrow
 MOD - "'Shall we go?'" "'Yes, Let's.'"
 MOD - We needn't go yet; the show doesn't start for an hour
 (=we don't have to go yet).
 MOD - Old people used to say "'Children must be seen and not heard'.
 MOD - Would that we had seen her before she died
 MOD - I might have known he'd do something silly; he's been acting strangely all
 week
 MOD - May there never be another world war!
 MOD - The person you mentioned will be the father of the person with the same
 name, is that right?
 MOD - I should think she must be at least 40.
 MOD - Must you shout so loudly!
 MOD - That would be in 1967, I think
 MOD - I wrote down his telephone number so that I could remember it
 MOD - Can you hold on a minute, please?
 MOD - Why shouldn't I say what I think?
 MOD - It shall be done as you wish
 MOD - He might come or he might not
 MOD - The effect of the tax should be felt in higher prices.
 MOD - I'll see what can be done.
 MOD - I wonder whether I oughtn't to speak to him
 MOD - If he had told me I must have helped him
 MOD - I must leave at 6 (compare I had to leave at 6 yesterday).
 MOD - They said it would be fine (actual words "'It will be fine'")
 MOD - I couldn't get the tickets yesterday
 MOD - He oughtn't to have said that (but he did).
 MOD - The teacher says they mustn't talk during lessons.
 MOD - The prisoner had hopes that he might be set free
 MOD - May I leave this with you?
 MOD - If people study, they will learn (=If people study, they learn)
 MOD - 20 degrees? I should have thought it was colder than that (if you hadn't
 told me)
 MOD - We looked and looked, but the ring was nowhere to be found
 MOD - I must look funny in this hat!
 MOD - Must I drink this nasty soup?
 MOD - You might have known she'd refuse.
 MOD - Would you please lend me your pencil?
 MOD - I may say I find your questions rather rude (=it is my opinion that they
 are rude)
 MOD - Children will often be (=usually are) full of life when their parents are
 tired

fig. 4

Exemples d'emplois d'auxiliaires modaux, tirés du LDOCE

Un autre élément intéressant que nous pouvons calculer à partir de notre base STAIRS consiste en différents comptages: nombre de définitions, nombre de codes grammaticaux par mot vedette. Ceci permet de calculer des *indices de complexité* des mots (fig. 5).

```

*****
A0009600 D00 absent v A0009600 D00 absent v P-001 P-001
*****
01 T1 from
P-001 ** 01 .. ** E01 He absented himself from the meeting P-001
*****
01 T0
02 T1
A0014500 D01 accept v P-001 ** 02 .. ** E01 I cannot accept your gift. P-001
A0014500 D01 accept v P-001 ** 03 .. ** E02 He asked her to marry him and she accepted (him) P-001
-----
01 T1
02 T5
*****
fm l;
03 wv5
A0014500 D02 accept v P-001 ** 04 .. ** E01 I accept your reasons for being late.
A0014500 D02 accept v P-001 ** 05 .. ** E02 (fm l) I accept that the aircraft has no choice but to crash into the sea.
A0014500 D02 accept v P-001 ** 06 .. ** E03 accepted principles of behaviour P-001
-----
01 T1
A0014500 D03 accept v P-001 ** 07 .. ** E01 I'll accept the bill only if you send it to my home address #see REFUSE#
*****
(USAGE)
01 X1
02 X7
*****
A0018500 D00 account v A0018500 D00 account v P-001 P-001
*****
P-001 ** 08 .. ** E01 He was accounted a wise man.
A0018500 D00 account v P-001 ** 09 .. ** E02 He accounted himself lucky to be alive P-001
*****
A0020600 D00 accustom v A0020600 D00 accustom v P-001 P-001
*****
01 T1 to
A0020600 D00 accustom v P-001 ** 10 .. ** E01 He had to accustom himself to the cold weather of his new country P-001
*****
01 I0
A0021350 D01 ache v P-001 ** 11 .. ** E01 I ache all over.
A0021350 D01 ache v P-001 ** 12 .. ** E02 My head aches P-001
-----
01 I0 for
02 I0 with
03 I3
A0021350 D02 ache v A0021350 D02 ache v P-001 P-001
*****
P-001 ** 13 .. ** E01 He was aching to go.
A0021350 D02 ache v P-001 ** 14 .. ** E02 aching with desire to go

```

```

*****
01
02
A0025900 b01 act v P.001 ** 15 .. ** E01 Olivier is acting ('theLol') tonight I0
----- T1 A0025900 b02 act v P.001 L1 A0025900 b03 act v P.001
A0025900 b02 act v P.001 ** 16 .. ** E01 He is always acting the experienced man who has seen everything I0
----- I0 A0025900 b04 act v P.001 L9 A0025900 b05 act v P.001
A0025900 b03 act v P.001 ** 17 .. ** E01 I can't take her seriously because she always seems to be acting I0
----- L9 A0025900 b04 act v P.001
A0025900 b04 act v P.001 ** 18 .. ** E01 The play acts well I0
----- I0 for I0 on
A0025900 b05 act v P.001 ** 19 .. ** E01 Think before you act! I0
----- I0 on
A0025900 b05 act v P.001 ** 20 .. ** E02 She acted favourably on our suggestion. I0
----- I0 on
A0025900 b05 act v P.001 ** 21 .. ** E03 I, this case I'm acting for my friend Mr. Smith I0
----- L9 A0025900 b06 act v P.001
A0025900 b06 act v P.001 ** 22 .. ** E01 to act foolishly/bravely and quickly I0
----- I0 on I0 upon
A0025900 b07 act v P.001 ** 23 .. ** E01 Does the drug take long to act (on the pain)? #see also ACT AS, ACT OUT,
----- ACT UP, ACT UP TO

```

fig. 5

Complexité syntaxique des verbes : mise en évidence par un extrait d'un document d'appariement "codes grammaticaux/exemples du LDOCE"

Les *définitions* en vocabulaire contrôlé permettent elles aussi des exploitations particulièrement utiles à l'apprenant, compte tenu du fait que, en décrivant le sens des mots, les définitions (même dans un dictionnaire de langue comme le *Longman* qui n'a aucune prétention encyclopédique) font l'*interface* entre les connaissances générales ("connaissances du monde") et linguistiques ("connaissance de la langue"), c'est-à-dire présentent cette *double compétence* (J.C. Gardin) indispensable à toute compréhension ou production langagières. Voici trois séries d'illustrations.

La première (fig. 6) concerne la facilité d'accès à un savoir encyclopédique, en l'occurrence les noms de papillons. Dans les dictionnaires qui utilisent un vocabulaire définitoire libre, l'accès à ces nomenclatures est généralement très difficile, les termes savants étant généralement définis par des mots tout aussi sinon plus inaccessibles à l'apprenant. (Voir *ugli*, fig. 1, introuvable

SEARCH - QUERY
00001 BUTTERFL\$.DEFITEXT.

HEADWORD caterpillar
DEFITEXT a type of small long many-legged wormlike creature (LARVA of the BUTTERFLY and other insects) which feeds on the leaves of plants #see picture at INSECT

HEADWORD copper
DEFITEXT a type of small BUTTERFLY, usu. with reddish-brown wings

HEADWORD moth
DEFITEXT any of several types of quite large winged insects, related to the BUTTERFLY but not (so) brightly coloured, which fly mainly at night and are attracted by flames or lights #see picture at INSECT

HEADWORD peacock
DEFITEXT also peacock butterfly #/'@@ "000/# an insect (BUTTERFLY) with large wings which have patterns on them like those on the tail of this bird

HEADWORD proboscis
DEFITEXT a long tubelike part of the mouth of some insects (esp. the BUTTERFLY) and worms, used for drawing in liquid, making holes, etc. #see picture at INSECT

HEADWORD red admiral
DEFITEXT a kind of BUTTERFLY with bright red bands on its black wings, common in Europe and America

HEADWORD tortoiseshell
DEFITEXT a type of BUTTERFLY with brownish markings

fig. 6

Mots tirés du LDOCE dont la définition contient "*butterfly*"

dans le *Longman* et d'autres dictionnaires, traduit dans *Robert & Collins E/F* par *tangelo*, lui-même inconnu du LDOCE et du *Robert & Collins F/E*, ...).

La deuxième illustration (fig. 7) montre comment les définitions facilitent l'accès à des informations de nature plus linguistique. Comme l'a indiqué Amsler, dans une thèse de doctorat non publiée, la tête sémantique des définitions (le *genus* aristotélicien, ou terme générique) est commune à tous les termes synonymes; dans une base de données comme la nôtre, ceci permet d'engendrer, à partir du dictionnaire alphabétique, un dictionnaire analogique, beaucoup plus révélateur qu'un thesaurus imprimé ou un dictionnaire de synonymes équivalent comme il en existe couramment incorporés aux systèmes de traitement de texte. Pourquoi plus révélateur? Pour une raison encore mal étudiée mais qui saute aux yeux: la "queue" de nos définitions (grossièrement assimilable aux différences spécifiques aristotéliciennes; cf. les travaux de Fillmore sur la différence entre "critique" et d'autres synonymes, par exemple) offre une source d'analyse considérable du *non dit*. Avec les ressemblances formelles (*critique* rime avec *pique*), c'est ce non dit caché dans les différences spécifiques entre des mots et locutions synonymes (*conceptuellement* et donc *grammaticalement équivalents*: *dire du mal de*) qui constitue sans doute l'essentiel de ce que nous exprimons *implicitement*, pour laisser entendre ou insinuer quelque chose: depuis les jeux de mots jusqu'aux manifestations les plus caractéristiques d'une sensibilité, d'un milieu, d'une civilisation et selon les goûts, de tel ou tel humanisme, qui se traduit dans les belles lettres ou ailleurs.

SEARCH - QUERY
00002 AVOID\$.DEFITEXT.

HEADWORD anchorite
DEFITEXT a person who lives alone and avoids company in order to reach union with God; hermit (1)

HEADWORD arm
DEFITEXT to avoid being friendly with somebody

HEADWORD around
DEFITEXT so as to avoid or get past; round
EXAMPLES @E01 Let's go around the town, not through it.
@E02 (fig.) How can we get around the new taxes and keep some more money for ourselves? #see round (USAGE)

HEADWORD aversive
DEFITEXT tending or causing to avoid something that is unpleasant or painful
EXAMPLES @E01 aversive training to get rid of bad habits by making them seem unpleasant

HEADWORD	avert
DEFITEXT	to prevent happening; avoid
EXAMPLES	ⒶE01 Accidents can be averted by careful driving
HEADWORD	avoidance
DEFITEXT	the act of avoiding
EXAMPLES	ⒶE01 avoidance of danger
HEADWORD	back
DEFITEXT	often derog to avoid; go away from (esp. when one should stay)
EXAMPLES	ⒶE01 He's always been kind to me#I can't just turn my back on him now he's ill and poor
HEADWORD	beg
DEFITEXT	to avoid
EXAMPLES	ⒶE01 Your plan is interesting, but it seems to beg the real difficulties
HEADWORD	blacklist
DEFITEXT	to put on a BLACKLIST# and to avoid, not give help or work to, not trade with, etc.
EXAMPLES	ⒶE01 blacklisted for non-payment of debts/for political reasons
HEADWORD	blush
DEFITEXT	infml to avoid making someone BLUSH# (1)
EXAMPLES	ⒶE01 You shouldn't say all these nice things about me; spare my blushes!
HEADWORD	Bob
DEFITEXT	(used for showing satisfaction that a way of avoiding or preventing some difficult state of affairs has been found)
EXAMPLES	ⒶE01 I needed to get to town but I haven't got a car: my friend has a car but he doesn't know the way: and so we agreed to travel together#Bob's your uncle!

fig. 7

Mots tirés du LDOCE (extrait) dont la définition contient "avoid"

Dans une troisième série d'illustrations, j'emprunte à Charles Fillmore (1982) son exemple d'interaction entre les *savoirs linguistiques* et *généraux* dans la compréhension d'un énoncé difficile, même si – par une aberration pédagogique – il a servi de test de compréhension dans des écoles américaines. Les extraits qui suivent de notre base *Longman* montrent l'intérêt qu'il y aurait à mettre une telle base de données dans les mains de l'apprenant, même anglophone. Voici donc le passage en question (fig. 8)

If a bronco buster wants to win a rodeo contest he must observe the contest rules.

Pour gagner dans un rodéo, un cowboy doit en observer les règles.

fig. 8

Texte extrait de Fillmore (1982), avec ma traduction

avec ma traduction, suivi de cinq informations que donne le *Longman* pour des mots difficiles (voir fig. 9-13).

SEARCH - QUERY
00004 BRONCO.HEADWORD.

HEADWORD bronco
DEFITEXT a wild or half-wild horse of the western US; MUSTANG

fig. 9
Définition de "*bronco*", tirée du LDOCE.

SEARCH - QUERY
00005 BUSTER.HEADWORD.

HEADWORD buster
DEFITEXT fellow
EXAMPLES @E01 Come here, buster!

HEADWORD -buster
DEFITEXT a person who BUSTs# or breaks
EXAMPLES @E01 Our crimebusters will catch those criminals

fig. 10
Définitions de "*(-)buster*", tirée du LDOCE

SEARCH - QUERY
00006 CONTEST.HEADWORD.

HEADWORD contest
DEFITEXT to compete for; fight for
EXAMPLES @E01 How many people are contesting this seat on the town council?

HEADWORD contest
DEFITEXT to question the truth or rightness of (something)
EXAMPLES @E01 I intend to contest the judge's decision in another court

HEADWORD contest
DEFITEXT a struggle or fight in which 2 or more people compete for victory
EXAMPLES @E01 a contest of skill
@E02 A contest developed for the position of minister of foreign affairs

HEADWORD contest
DEFITEXT a competition, esp. one judged by a PANEL (=a group) of specially chosen judges
EXAMPLES @E01 a beauty contest
@E02 a dancing contest

fig. 11
Recherche STAIRS sur "*contest*" dans le LDOCE

SEARCH - QUERY
00007 RODEO.HEADWORD

HEADWORD rodeo
DEFITEXT a gathering together of cattle

HEADWORD rodeo
DEFITEXT a public performance at which COWBOYS ride wild horses, catch cattle with ropes, etc.

fig. 12
Recherche STAIRS sur "rodeo" dans le LDOCE

SELECT - QUERY
00010 9 HEADNUM=B0187600

HEADWORD break
DEFITEXT to bring under control
EXAMPLES @E01 to break a horse/a child's spirit

fig. 13
Recherche sur "break a horse" dans le LDOCE

2.2. Codes grammaticaux : le Longman, un dictionnaire/grammaire

Grâce à des programmes de réécriture très complexes – écrits par l'ingénieur de notre équipe, Jacques Jansen – nous avons accès aux codes grammaticaux, l'autre principale richesse de la base *Longman*. Unique à cet égard, le *Longman* associe des codes grammaticaux très précis à tous les sens pertinents, c'est-à-dire à ses définitions (fig. 14). Nous allons bientôt publier un tableau statistique donnant le nombre d'occurrences de chaque code grammatical (Jansen, Noël, Vanandroye : sous presse) et notre équipe prépare aussi une communication sur un codage grammatical des exemples du *Longman*, effectué *manuellement* par nos étudiants, faute de ne l'avoir été par Longman dans le fichier d'origine (travail fait sous la direction de J.P. Mergeai et J. Vanandroye grâce à des outils développés par J. Jansen). Ces informations grammaticales viendront donc enrichir considérablement le fichier tel qu'il nous a été prêté par la firme Longman. Limitons-nous ici à trois brefs exemples et à une illustration. Le premier exemple concerne les structures du type Nom + Préposition avec "that" (*that-clause*). Un mémoire d'une de nos étudiantes (N. Fraipont) montre l'insuffisance des descriptions tant lexicographiques que grammaticales de ces

Table of Codes	<div style="display: flex; justify-content: space-between; padding: 0 5px;"> <i>n or adj used before a noun</i> <i>ordinary adj</i> <i>count noun</i> <i>ditransitive v with 2 different objects</i> <i>adj, adv, or n used after a noun</i> <i>adj or adv used after a verb</i> <i>group countable n</i> <i>group uncountable n (or adj)</i> <i>adv used with preposition and other adverbs</i> <i>intransitive v with no object</i> <i>linking v with complement</i> </div>											
	A	B	C	D	E	F	GC	GU	H	I	L	
nouns with no letter are [C], adjectives with no letter [B]	[A] the main difficulty General Smith	[B] a happy man The man is happy. She made him happy	[C] a lone dog boxes		[E] the president elect 3 years ago Saadi Pasha	[F] She was asleep. The meeting was yesterday	[GC] the committee is are the committees are	[GU] the Admiralty is are the accused is are	[H] right through splash into the water			
0 need not be followed by anything										[I0] We paused		
followed by one or more nouns or pronouns 1				[D1] Give the boy a book. Buy him a book							[L1] She became Queen. It cost £6	
followed by the infinitive without to 2										[I2] I can fly		
followed by the infinitive with to 3		[B3] an easy person to please	[C3] an attempt to climb the mountain			[F3] John is eager to please				[I3] He lived to be 90	[L3] The difficulty is to know what to do	
followed by the -ing form 4										[I4] She came running	[L4] She ended up dancing on the table	
followed by a that- clause 5			[C5] a desire that she (should) go	[D5] He warned her (that) he would come		[F5] He was sure (that) she knew				[I5] It appears (that) she will win	[L5] The trouble is that you know too much	
followed by a wh- word 6			[C6] the reason why he came	[D6] Tell me who should go where to go		[F6] I'm not sure where to go where I should go				[I6] It appears as if she will win	[L6] It's as if we'd never even started	
followed by an adjective 7											[L7] She became famous	
followed by a past participle 8										[I8] Smoking is not permitted	[L8] He got trapped	
needs a descriptive word or phrase 9		[B9] politically aware	[C9] a fitful tennis buff			[F9] located in Florida		[GU9] his posterity			[L9] She lives here	

fig. 14

Tableau des codes grammaticaux du LDOCE

vocative n used in direct address	plural n	n that is a name or name-like	singular n	transitive v with one object	uncountable n	v with one object + verb form	adjective and adverb	noun	pronoun	verb	v with one object + something else
N	P	R	S	T	U	V	Wa	Wn	Wp	Wv	X
[N] Good-bye doctor!	[P] The police are here. Do the dead return?	[R] God the Earth	[S] a think a babble of voices		[U] sugar is sweet						
				[T1] She kicked the boy. She blew up the bridge			[Wa1] takes etc., etc.: Alice, Alice, Alice!	[Wn1] plural sometimes like sing.: several Boas (lions)	[Wp1] personal pronouns: see table	[Wv1] the verb form: see table	[X1] They considered him their enemy
			[T2] I helped clean the windows		[V2] I saw the man leave	[Wa2] sometimes takes -er, -est; secure (+r/+s) or more, most secure	[Wn2] plural usu like sing.: several quail (quails)	[Wp2] other pronouns: see table	[Wv2] AUXILIARY verbs: see table		
	[P3] qual- ifications to do the job		[S3] a yen to be alone	[T3] I want to go	[U3] There is some reason to believe it	[V3] I want him to go	[Wa3] no (s) before an ending: simply /simply/	[Wn3] plural same as sing.: several grouse		[Wv3] no (s) before -ing, -er ending: coupling /kʌplɪŋ/	
				[T4] I enjoyed singing their singing		[V4] He watched mother cooking the dinner	[Wa4] -ly pro- nounced /li/; histori- cally /hi- storkli/			[Wv4] used in -ing form as adj.: flying feet	
			[S5] a feeling that he'll come	[T5] I know that he'll come	[U5] Is there proof that he is here?		[Wa5] no COMPARA- TIVE OR SUPERLA- TIVE: stompe bomb			[Wv5] used in -ed form as adj.: plaited hair	
				[T6] He decided where to go/who should go						[Wv6] not used in -ing form: He sees me now	
						[V8] to have a house built					[X7] They considered him dead
	[P9] powers of memory	[R9] the third inter- national	[S9] a fine heli- geoce		[U9] jurisdic- tion over us all						[X9] Put it in the box

Note: the numbers in the codes in this box have a different meaning

fig. 14
Tableau des codes grammaticaux du LDOCE

constructions, sans doute plus productives en anglais qu'en français : *The joy of (the book) / (the region) is that ... ???* *La joie de ... est que ...* Un autre mémoire en préparation (D. Hasoppe) donne, sur base de la célèbre *Grammaire Longman* (Quirk et al., 1985), des démonstrations précises de l'insuffisance notoire des descriptions des adverbes (*disjuncts*: comme *demonstrably*, *démontrablement) dans notre dictionnaire *Longman*. Tout ceci n'est nullement une critique mais bien au contraire un effort d'amélioration de ce dictionnaire et des grammaires de l'anglais. Comme aides à nos étudiants, nous présentons actuellement des listes de mots, comme celle des noms non comptables (*Uncountable*) mal connus des francophones; la figure 15 donne un sous-ensemble des noms qui peuvent régir une proposition avec *that* (code U5, *Uncountable*, non comptable + *that*).

SEARCH - QUERY
00011 %U5\$.GRAMCODE.

HEADWORD	anticipation	HEADWORD	fear
HEADWORD	anxiety	HEADWORD	hope
HEADWORD	assertion	HEADWORD	hope
HEADWORD	certitude	HEADWORD	hope
HEADWORD	conception	HEADWORD	hope
HEADWORD	confirmation	HEADWORD	hope
HEADWORD	consciousness	HEADWORD	hope
HEADWORD	conviction	HEADWORD	hope
HEADWORD	desire	HEADWORD	indication
HEADWORD	discovery	HEADWORD	possibility
HEADWORD	dissatisfaction	HEADWORD	probability
HEADWORD	doubt	HEADWORD	proof
HEADWORD	entreaty	HEADWORD	proposat
HEADWORD	evidence	HEADWORD	realization, -sation
HEADWORD	evidence	HEADWORD	reservation
HEADWORD	evidence	HEADWORD	risk
HEADWORD	evidence	HEADWORD	speculation
HEADWORD	evidence	HEADWORD	supposition
HEADWORD	external evidence	HEADWORD	testimony
HEADWORD	faith		

fig. 15

Recherche STAIRS : noms non comptables + proposition avec "that" (U5),
tirés du LDOCE

2.3. Codes matières

Je passe maintenant à la sélection des textes selon des critères et informations du *Longman*, et je donne comme exemple l'application des codes matières à la sélection d'un texte sur les vins de

quelque 2 000 mots (fig. 1) dans le corpus LOB comprenant plus d'un million de mots. Il va de soi, cependant, que la démonstration n'est nullement limitée à cette application: comme on va le voir, le trouage de la figure 1 (selon les critères: verbe + vocabulaire contrôlé) montre la possibilité de sélectionner un texte selon toute espèce de critère, ou combinaison de critères, afin de calculer la population de tel ou tel type de mot dans un texte ou dans un ensemble de textes.

3. Sélection des textes

Grâce à des programmes rédigés par Jacques Jansen utilisant un logiciel de calcul statistique bien connu (SAS), nous avons réalisé une *indexation automatique* dont le principe repose sur l'application des codes matières du *Longman* à un texte quelconque afin de représenter le sujet dont il traite.

La méthode est sommaire puisqu'elle repose sur un comptage de fréquence de mots isolés. Mais elle est performante et économique, ce qui la recommande pour l'analyse du contenu dans des fichiers très volumineux de textes (application d'Amsler, non publiée, aux fichiers d'un quotidien, le *New York Times*).

Dans le fichier *Longman*, plus de 32 000 définitions sont munies de codes matières de quatre caractères. Les deux premiers représentent de grands domaines du savoir, et les deux derniers soit un autre grand domaine, soit une subdivision du premier, introduite par un Z en troisième position et symbolisée par le quatrième caractère.

Nous avons appliqué ce codage matière au texte de la figure 1, qui fait partie d'une série de trois textes tirés d'un magazine féminin (E19 dans le LOB) sur l'art de choisir ses vins. Les programmes opèrent en deux étapes: un premier comptage statistique sur tous les codes matières associés aux mots d'un texte, un second sur les codes matières les plus fréquents ainsi repérés. Appliqué à notre texte, ce programme nous a produit une liste de codes matières, rangés en ordre de fréquence décroissante, où les codes FO (Food) et BV (Beverages) apparaissent en tête comme les codes de loin les plus fréquents (fig. 16).

SAS

OBS	COUNT	PERCENT	MAT1	LIBEMAT
1	76	23.3129	FO	FOOD
2	30	9.2025	BV	BEVERAGES
3	26	7.9755	MD	MEDICINE & BIOLOGY
4	21	6.4417	BZ	BUSINESS
5	16	4.9080	LW	LAW
6	13	3.9877	LN	LING. AND GRAMMAR
7	13	3.9877	SP	SPORTS
8	9	2.7607	PL	POLITICAL SCIENCE
9	8	2.4540	RL	RELIGION
10	7	2.1472	AG	AGRICULTURE
11	7	2.1472	MS	MEASURES
12	6	1.8405	CO	COLOUR
13	6	1.8405	MH	MATHEMATICS
14	6	1.8405	MI	MILITARY
15	5	1.5337	NB	NUMBERS
16	4	1.2270	BW	BOWLING AND BOLS
17	4	1.2270	EC	ECONOMICS
18	4	1.2270	EG	ENGINEERING
19	4	1.2270	ML	METEOROLOGY
20	4	1.2270	SI	SCIENCE
21	3	0.9202	AF	ART
22	3	0.9202	CL	CLOTHING
23	3	0.9202	GB	GAMBLING
24	3	0.9202	GO	GEOLOGY AND GEOGRAPHY
25	3	0.9202	MU	MUSIC
26	2	0.6135	AH	ANIMAL HUSBANDRY
27	2	0.6135	AM	ANIMAL NAMES
28	2	0.6135	AP	ANTHROPOLOGY
29	2	0.6135	BD	BUILDING
30	2	0.6135	EN	
31	2	0.6135	GA	GAMES
32	2	0.6135	HH	HOUSEHOLD
33	2	0.6135	NA	NAUTICAL
34	2	0.6135	NT	NET GAMES
35	2	0.6135	OR	ORDERS
36	2	0.6135	PM	PLANTS NAMES
37	2	0.6135	PT	PRINTING AND PUBLISHING
38	2	0.6135	SO	SOCIOLOGY
39	1	0.3067	AS	ASTRONOMY
40	1	0.3067	BL	BIBLE & APOCRYPHA
41	1	0.3067	CA	CALENDAR
42	1	0.3067	CK	
43	1	0.3067	DE	DENTISTRY
44	1	0.3067	ES	
45	1	0.3067	HI	HISTORY
46	1	0.3067	HR	HOROLOGY
47	1	0.3067	JE	
48	1	0.3067	MP	MOTION PICTURES
49	1	0.3067	NU	NUMISMATICS
50	1	0.3067	OC	OCCULT
51	1	0.3067	OZ	ZOOLOGY
52	1	0.3067	PH	PHILOSOPHY
53	1	0.3067	SN	SOUNDS
54	1	0.3067	YT	

fig. 16
Statistiques sur codes matières appliqués au texte des vins

4. Sélection des trous

Comme le sont aussi la plupart des travaux sur le LOB (p. ex. : *LET*; voir toutefois les travaux de Leech sur l'analyse syntaxique, in Leech & Candlin, *op. cit.*), notre programme de trouage est actuellement encore limité au mot.

Bornons-nous à reprendre la figure 1, c'est-à-dire le texte qui a été sélectionné par un calcul sur les codes matières. Dans ce texte, nous avons effectué un trouage *semi-automatique* selon les deux critères de sélection suivants :

- (i) le premier consiste à repérer dans le texte tous les mots et variantes morphologiques qui *peuvent* être des *verbes* du vocabulaire contrôlé;
- (ii) le second consiste à choisir manuellement ceux de ces mots qui sont *effectivement* utilisés comme verbes.

Nous pouvons utiliser deux types de *stoplists* (liste de mots sautés dans le trouage). Outre une liste utilisée dans notre illustration et constituée par nos soins d'après une liste publiée des fréquences des mots du LOB (*LET*), nous disposons aussi d'une liste de 326 mots vides (*CS = Closed System = mots non lexicaux*: cette liste va de *'em* à *yet*). On peut aussi concevoir, pour les débutants, un trouage sur ces mots vides ou sur un sous-ensemble plus sélectif de ceux-ci: par exemple, les prépositions.

Une sélection automatique de tous les *homographes* qui peuvent être *verbes* en anglais donnerait des résultats curieux, comme le savent les spécialistes de *l'informatique linguistique*. Voici mon exemple favori en français, dont j'ai oublié la source. Tous les mots de ce bref énoncé (pourtant non équivoque pour le francophone moyen) sont *syntactiquement ambigus* pour un automate qui n'a accès à aucun savoir, sinon syntaxique (fig. 17).

Le car entre par la porte centrale de la gare de l'est.

fig. 17
Exemple français d'ambiguïté syntaxique

5. Sélection des aides

Quelles aides peut-on présenter sur l'écran, dans une petite fenêtre, dans l'E.A.O. de demain – ou sur papier, comme nous l'avons fait jusqu'à présent? J'ai déjà en partie répondu à la question en décrivant les divers types d'informations accessibles dans nos dictionnaires automatiques. Une fois encore, soulignons qu'il resterait à mettre en oeuvre des algorithmes d'analyse et de synthèse qui "comprennent l'anglais" pour qu'un système comme le nôtre tende vers une application *automatique* d'un dictionnaire quelconque à un texte quelconque (dans son doctorat, mon collègue A. Michiels a conçu des algorithmes d'analyse basés sur les informations syntaxiques et autres du *Longman*). En attendant, nos sélections d'aides restent *semi-automatiques*. Reprenons ici les principaux types d'aides et esquissons, en guise de conclusion, quelques prolongements importants :

a) les *aides grammaticales*.

Elles se limitent actuellement à distribuer à nos étudiants des listes de mots portant un ou plusieurs codes grammaticaux (comme U5, fig. 5) ou des concordances et exemples tirés du *Longman* (fig. 3 et 4). Le codage manuel des exemples, actuellement en cours, enrichira considérablement ce type d'exploitations.

b) les *synonymes* (voir illustration ci-après).

Nous envisageons ici l'exploitation d'un minidictionnaire de poche dont nous avons également obtenu le fichier par contrat (le *Collins GEM English Dictionary*) spécialement pour les verbes et adjectifs. Beaucoup plus courtes que celles du *Longman*, les définitions du *GEM* sont équivalentes à des listes de synonymes et ceux-ci nous paraissent meilleurs que ceux des dictionnaires de synonymes de taille équivalente ou même supérieure, y compris ceux qu'actionnent certains traitements de texte.

c) les *expressions idiomatiques et locutions*.

Outre celles du *Longman* et du *Collins & Robert*, nous envisageons d'exploiter les expressions idiomatiques de deux autres dictionnaires de locutions, dont nous avons également obtenu les fichiers par contrats : LDOEI, ODCIE (vol. 1 & 2). Il va de soi que, dans le domaine des expressions idiomatiques – sinon pour ce qui concerne tous les aspects formulaïques des langues – les dictionnaires sont pratiquement notre *seule* source d'information.

d) la *production de la parole parlée* (aides à la lecture, production orale).

Comme l'ont encore fait les auteurs de la célèbre grammaire de Quirk et al. dans la dernière version de cet ouvrage (Quirk et al.,

1986), les linguistes insistent à juste titre sur la primauté de la parole parlée. Pour l'instant, notre équipe ne dispose pas d'un synthétiseur. Mais nous pouvons faire deux choses dans le domaine de la parole. Nous pouvons utiliser les notations phonétiques et les représentations de l'accent tonique ou de l'intonation, qui sont contenues dans les dictionnaires et les corpus. Voici un exemple de mot de plus d'une syllabe extrait du fichier du *Collins GEM*.

```
#3ADD #6v. #5join; increase by; say further #laddi#ition #6n.
#5 #laddi#itional #6a. #5 #ladd#itive #6n. #5something added,
#6esp. #5to foodstuffs<

#3AL#!COHOL #6n. #5intoxicating fermented liquor; class of organic
chemical substances #lalcohol#!ic #6a. #5 #6n. #5one addicted
to alcoholic drink #lal#!coholism #6n. #5disease, alcohol
poisoning<

#3AL#!TERNATE #5$(awl#!-$) #6v. #5occur or cause to
#5 occur by turns #lalter#!inate #6a.
#5#5one after the other, by turns #lalter#!nately #6adv.
#5 #lalterna#ition #6n. #5 #lalter#!natively #6n. #5one of two
choices #6a. #5replacing #lalter#!natively #6adv.

#3AR#!TICHOKE #6n. #5thistle-like perennial,
#5 edible flower #1Jerusalem artichoke #5sunflower with edible
tubers like potato<
```

fig. 18
Extrait du Collins GEM : 4 mots en A

Nous pouvons utiliser le corpus d'anglais parlé de Svartvik et Quirk (London-Lund). Ce corpus, dont nous disposons également, donne des textes parlés authentiques avec des transcriptions de l'intonation phrastique (fig. 19).

Enfin :

e) les *aides en traduction* (traduction assistée) extraites du *Collins & Robert*.

Voici quatre mots pleins en A dans le texte de la figure 1 : figures 20-23.

L'exposition de cet excellent fichier aurait pu faire l'objet, à elle seule, d'un exposé comme celui qui doit se terminer ici.

1 1 1 10 1 1 B 11 ((of ^Spanish)) . graph\ology /
1 1 1 20 1 1 A 11 ^w=ell# /
1 1 1 30 1 1 A 11 ((if)) did ^y/ou _set _that# - /
1 1 1 40 1 1 B 11 ^well [J\oe and _I# /
1 1 1 50 1 1 B 11 ^set it betw\een _us# /
1 1 1 60 1 1 B 11 ^actually |Joe 'set the :p\aper# /
1 1 1 70 1 1 B 20 and *((3 to 4 sylls))* /
1 1 1 80 1 1 A 11 *^w=ell# . /
1 1 1 90 1 1 A 11 ""^m/\ay* I _ask# /
1 1 1 100 1 1 A 11 ^what goes |v\into that paper n/ow# /
1 1 1 110 1 1 A 11 be^cause I |have to adv=ise# . /
1 1 1 120 1 1 A 21 ((a)) ^couple of people who are |d\oing [dhi: @] /
1 1 1 130 1 1 B 11 well ^what you :d\o# /
1 1 1 140 1 2 B 12 ^is to -- this is sort of betw\een the :tw\o of /
1 1 1 140 1 1 B 12 us# /
1 1 1 150 1 1 B 11 ^what *you* :d\o# /
1 1 1 160 2 1 B 23 is to ^make sure that your 'own . |c\andidate /
1 1 1 170 1 1 A 11 *^[\m]#* /
1 1 1 160 1 2(B 13 is . *.* ^that your . there's ^something that your /
1 1 1 160 1 1(B 13 :own candidate can :h\andle# - - /
1 1 1 180 2 1 B 21 ((I ^won't)) /
1 1 2 190 1 1 A 11 *((^y\eah))* /
1 1 2 180 1 1(B 11 ((be a m/inute# - - /
1 1 2 200 1 1 B 20 3 to 4 sylls)) - - - /
1 1 2 210 1 1 A 11 ^\ah# - /
1 1 2 220 1 1 A 11 [@:] you ^mean that [dhi dhi] (the |p\apers) ""\are# /
1 1 2 230 1 1 A 11 ^more or less :set ad |h\ominem# /
1 1 2 240 1 1 A 11 ^/are _they# - /
1 1 2 250 1 1 B 11 [@:h] - - they ^sh\ouldn't b/e# - - /
1 1 2 260 1 1 B 12 ^but [@h] - I ^mean /\one# /
1 1 2 270 1 1 B 11 ^sets - - \one question# /
1 1 2 280 1 2 B 11 ^now I |mean |this fellow's doing ((the)) language /
1 1 2 280 1 1 B 11 of :\advertising# *.* /
1 1 2 290 1 1 B 11 ((so ^v\erv)) w\ell# /
1 1 2 300 1 1 A 11 *^y\eah*# /
1 1 2 310 1 1(B 21 *((^give him one on))* /
1 1 2 320 1 1 A 11 *is ^this a spare p\aper*# /
1 1 2 330 1 1 B 11 (([@] ^y\eah# /
1 1 2 340 1 1 B 11 I actually ^got it for y/ou#)) /
1 1 2 350 1 1 A 11 ^thank you very m/uch# /
1 1 2 360 1 1 A 11 ((1 syll)) *^thanks very much in:d\eed*# /
1 1 3 370 1 1 B 11 *I've ^probably got some :m\ore*# /
1 1 3 380 1 1 B 11 [f@:] ^further b\ack# . /
1 1 3 390 1 1 A 12 be^cause you see ((I)) - [d] ^some of our p\eople# /
1 1 3 400 1 1 A 11 (. clears throat) ((who)) are ^d\oing LEs# - - /
1 1 3 410 1 2 A 11 [d:m] ^have to consider which |p\aper) . to /
1 1 3 410 1 1 A 11 "d\o# . /
1 1 3 420 1 2 A 11 and I ^wondered . whether . [dhi] . gra:phology /
1 1 3 420 1 1 A 11 p\aper# . /
1 1 3 430 1 2 A 12 is in ^fact . [d:m] . ^whether it |tends to be a /
1 1 3 430 1 1 A 12 com:p\arative gra_phology _paper# /
1 1 3 440 1 1 A 11 ^or - a his|t\orical gra_phology _paper# - /
1 1 3 450 2 1 A 21 ((^whether it's)) *like* /
1 1 3 460 1 1 B 21 *((^well you))* /
1 1 4 450 1 1(A 11 |old and Middle :English graph/ology# /
1 1 4 470 1 1 A 11 ((or ^something *like |th\at# /
1 1 4 480 1 1 A 11 ^you s/ee#))* /
1 1 4 490 1 1 B 11 *^w=ell# /
1 1 4 500 1 1 B 11 ^you give* them the :l\ot {^you s=ee#}# *.* /
1 1 4 510 1 1 B 11 ^that's the **p\oint**# /

fig. 19
Extrait du London-Lund, corpus de textes parlés

SEARCH - QUERY

00001 ADD.HEADWORD.

HOMOGR : 0

HEADNUM : 303

DEFINUM : 0A

HEADWORD add

MEANING vt ajoutez (to a). add some more pepper ajoutez encore or rajoutez un peu de poivre; @ to add insult to injury porter l'insulte a son comble; add ed to which ... ajoutez a cela que ...

HOMOGR : 0

HEADNUM : 303

DEFINUM : 0B

HEADWORD add

MEANING vt (Math) figures additionner. add ing machine calculatrice f machine f a calculer.

HOMOGR : 0

HEADNUM : 303

DEFINUM : 0C

HEADWORD add

MEANING vt (say besides) ajouter (that que). @ there is nothing to add c'est tout dire, il n'y a rien a ajouter.

HOMOGR : 0

HEADNUM : 303A

DEFINUM : 0

HEADWORD add in

MEANING vt sep details inclure, ajouter; considerations faire entrer en ligne de compte.

HOMOGR : 0

HEADNUM : 303B

DEFINUM : 0

HEADWORD add to

MEANING vt fus augmenter, accroitre, ajouter a. @ this only adds to our anxiety ceci ne fait qu'ajouter a or qu'accroitre notre inquietude.

HOMOGR : 0

HEADNUM : 303C

DEFINUM : 0

HEADWORD add together

MEANING vt sep figures, advantages, drawbacks additionner.

HOMOGR : 0

HEADNUM : 303D

DEFINUM : 0A

HEADWORD add up 1

MEANING vi (Math) @ these figures don't add up (right) or @ won't add up ces chiffres ne font pas le compte (exact); (fig) @ it all adds up>u89< tout cela concorde, tout s'explique; (fig) @ it doesn't add up>u89< cela ne rime a rien, il y a quelque chose qui cloche. 2 vt sep figures additionner. @ to add up a column of figures totaliser une colonne de chiffres.

HOMOGRA : 0
 HEADNUM : 303D
 DEFINUM : 0B
 HEADWORD add up 1
 MEANING _vi (Math) @ these figures don't add up (right) _or @ won't add up
 ces chiffres ne font pas le compte (exact); (fig) @ it all adds up>u89<
 tout cela concorde, tout s'explique; (fig) @ it doesn't add up>u89< cela
 ne rime a rien, il y a quelque chose qui cloche. 2 _vt sep (fig)
 _advantages, reasons _faire la somme de.

HOMOGRA : 0
 HEADNUM : 303E
 DEFINUM : 0
 HEADWORD add up to
 MEANING _vt fus (figures) _s'elever a, se monter a; (>u89< _fig: mean _)
 signifier, se resumer a.

fig. 20

Recherche sur "add"

Extraits de *Collins & Robert, French/English English/French Dictionary*,
 Copyright 1978 William Collins Sons & Co Ltd & La Société du Nouveau
 Littre Dictionnaire Le Robert.

SEARCH - QUERY
 00006 ALCOHOLIC.HEADWORD.

HOMOGRA : 0
 HEADNUM : 645
 DEFINUM : 1
 HEADWORD alcoholic
 MEANING _adj person _alcoolique; _drink _alcoolise, alcoolique.

HOMOGRA : 0
 HEADNUM : 645
 DEFINUM : 2
 HEADWORD alcoholic
 MEANING _n _alcoolique _mf. @ A alcoholic s Anonymous _societe _f _d'entraide
 des alcooliques, alcooliques _mpl _anonymes.

fig. 21

Recherche sur "alcoholic" (extraits de *Collins & Robert*)

SEARCH - QUERY
 00002 ARTICHOKE.HEADWORD.

HOMOGRA : 0
 HEADNUM : 1453
 DEFINUM : 0
 HEADWORD artichoke
 MEANING _n _artichaut _m _; _v @ globe, Jerusalem.

fig. 23

Recherche sur "artichoke" (extraits de *Collins & Robert*)

SEARCH - QUERY

00005 ALTERNATIVE.HEADWORD.

HOMOGR : 0
 HEADNUM : 768
 DEFINUM : 1
 HEADWORD alternative
 MEANING adj possibility, answer _autre; (Philos) _proposition _alternatif;
 (Mil) _position _de repli; (Tech) de rechange. alternative proposal
 _contre-proposition _f _; @ the only alternative method _la seule autre
 methode, la seule methode de rechange; (Aut) alternative route _itineraire
 _m _de delestage.

HOMOGR : 0
 HEADNUM : 768
 DEFINUM : 2
 HEADWORD alternative
 MEANING _n (choice) (between two) alternative _f _; choix _m _; (between
 several) choix; (solution) (only one) alternative, seule _autre solution,
 solution unique de rechange; (one of several) autre solution, solution de
 rechange; (Philos) terme _m _d'une alternative _or _d'un dilemme. @
 she had no alternative but to _accept _elle n'avaït pas _d'autre solution
 que d'accepter, force lui a ete d'accepter.

fig. 22

Recherche sur "alternative" (extraits de Collins & Robert)

Bibliographie

- Amsler, R.A., (1980). *The Structure of the Merriam - Webster Pocket Dictionary*, Austin TX, University of Texas, Ph.D Thesis. Available from : University Microfilms, Ann Arbor, MI.
- Amsler, R.A., (1984). *"Machine-Readable Dictionaries"*, *Annual Review of Information Science and Technology (ARIST)*, Vol. 19, Martha E. Williams ed.
- Barnes, D., (1976). *From Communication to Curriculum*, Harmondsworth, Penguin Education.
- Collins *Gem English Dictionary*, (1982). London & Glasgow, Collins.
- Debot, F. & Leclercq, D., (1978). *Système de Traitement Automatique d'Evaluations Pédagogiques*, Université de Liège.
- Higgins, J. & Johns, T., (1984). *Computers in Language Learning*, Glasgow, Collins.

- Fillmore, Ch.J., (1982). "Monitoring the Reading Process", *Linguistics in the Morning Calm*, ed. by The Linguistic Society of Korea, Seoul, Hanshin, pp. 329-348.
- Johansson, S. (in collaboration with Leech, G.N., Goodluck, H.), (1978). *Manual of Information* (to accompany the Lancaster-Oslo-Bergen corpus of British English, for use with digital computers), University of Oslo, Dept. of English.
- Laplanche, G., (1977). *Guide technique d'utilisation du STEP*, Université de Liège, Centre de Calcul, Service de Documentation.
- LDOCE (Longman Dictionary of Contemporary English)*, (1978). London, Longman.
- LDOEI (Longman Dictionary of English Idioms)*, (1979). London, Longman.
- Leech, G.N. & Candlin, C.N., eds., (1986). *Computers in English Language and Research*, London, Longman.
- Leech, G.N., (1986). "Automatic grammatical analysis and its educational applications", in Leech & Candlin.
- LET*, Engels, L.K. et al., (1981). *L.E.T. Vocabulary List*, K.U.L., Leuven, ACRO.
- Michiels, A., (1982). *Exploiting a large dictionary database*, Liège, Université de Liège, Thèse de doctorat.
- Michiels, A., (1986). Paper to be read at Düsseldorf University.
- ODCIE (Oxford Dictionary of Current Idiomatic English)*, London, OUP.
- Cowie, A.P. & Mackin, R., (1975). Vol. 1, *Verbs with Prepositions and Particles*.
- Cowie P., Mackin, R. & McCaig, I.R., (1983). Vol. 2, *Phrase, Clause and Sentence Idioms*.
- Quirk, R., Greenbaum, S., Leech, G.N. & Svartvik, J., (1985). *A Comprehensive Grammar of the English Language*, London, Longman.
- Svartvik, J. & Quirk, R., (1980). *A Corpus of English Conversation*, Lund, Lund Studies in English 56, Liber/Gleerups.