

Abdelkrim Mokhtari

**Cohésion lexicale et automatisation:
Application à la désambiguïsation lexicale**

Thèse
présentée
à la Faculté des Études Supérieures
de l'Université Laval
pour l'obtention
du grade de Philosophiae Doctor (Ph. D.)

Département de Langues et Linguistique
Faculté des Lettres

UNIVERSITÉ LAVAL
QUÉBEC

Août 1998



**National Library
of Canada**

**Acquisitions and
Bibliographic Services**

395 Wellington Street
Ottawa ON K1A 0N4
Canada

**Bibliothèque nationale
du Canada**

**Acquisitions et
services bibliographiques**

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-36305-8

Canada

Résumé 1

Cette thèse, qui s'inscrit dans le cadre de la linguistique informatique, a pour objectif de traiter du problème de la levée de l'ambiguïté lexicale d'une manière automatique. Pour ce faire, l'hypothèse de travail retenue est d'exploiter quelques ressources de la cohésion lexicale, en particulier les collocations présentes dans le contexte. La thèse présente deux volets complémentaires: le premier est une analyse linguistique de la notion de la cohésion lexicale et des possibilités d'une exploitation de celle-ci dans un processus d'automatisation. Le second est d'ordre pratique: en effet, un système informatique composé de trois programmes a été développé, mettant en œuvre une stratégie particulière pour traiter le problème en question. La thèse comprend également des tests effectuées sur deux types de corpus: Le *Code de la Sécurité Routière* et le *Code Civil de Québec*. Les résultats encourageants de ces tests sont suivis de propositions pour une amélioration de la stratégie avancée.

Résumé 2

Cette thèse, qui s'inscrit dans le cadre de la linguistique informatique, a pour objectif de traiter du problème de la levée de l'ambiguïté lexicale d'une manière automatique. Pour ce faire, l'hypothèse de travail retenue est d'exploiter quelques ressources de la cohésion lexicale, en particulier les collocations présentes dans le contexte. La thèse présente deux volets complémentaires: le premier est une analyse linguistique de la notion de la cohésion lexicale et des possibilités d'une exploitation de celle-ci dans un processus d'automatisation. Le second est d'ordre pratique: en effet, un système informatique composé de trois programmes a été développé, mettant en œuvre une stratégie particulière pour traiter le problème en question.

Le premier programme, appelé MARQUEUR, permet d'annoter d'une manière semi-automatique les mots du texte en leur attribuant le lemme et la catégorie grammaticale. Le deuxième programme, développé en équipe et appelé COLLOCATEUR, utilise ces marques pour extraire du texte des séquences structurées, dites des collocations. Le troisième programme va tirer parti de l'apport des deux premiers pour offrir la possibilité de localiser et de lever l'ambiguïté aussi bien morphologique que sémantique des mots objets d'une requête. Dans notre stratégie, nous partons du constat que les collocations sont généralement inductrices de significations particulières, et nous présentons une façon simple mais productive d'emmagasiner ces données.

Au lieu de recourir à une analyse syntaxique et sémantique complexe, nous avons exploré une voie qui consiste à travailler à partir de données brutes extraites de textes attestés et soumises à des transformations mineures.

La thèse comprend également des tests effectués sur deux types de corpus: Le *Code de la Sécurité Routière* et le *Code Civil du Québec*. Le premier test a été effectué sur un texte dont les collocations ont été extraites au préalable. Le second sur un texte nouveau, sans autre traitement que le marquage. Les résultats encourageants de ces tests sont suivis de propositions pour une amélioration de la stratégie avancée.

Dédicace

A la mémoire de mes parents.

Avant-propos

Voici une tranche de vie qui s'achève. Un parcours des plus passionnants!

Le présent travail n'aurait sûrement pas vu le jour si de multiples bonnes volontés ne s'étaient manifestées à mon égard. A commencer par celle de mon directeur de thèse: Michael Mephram. Je ne le remercierai jamais assez tant son aide a été abondante, l'amitié dont il m'a honoré profonde et la confiance qu'il m'a accordée grande et sans détours. Pour l'avoir côtoyé durant plus de cinq ans, autant je sais qu'il ne cherche pas à être remercié quand il aide quelqu'un, autant lui sait que je ne suis pas loquace pour exprimer une reconnaissance. Qu'il trouve ici mes remerciements les plus chaleureux et ma profonde gratitude.

L'initiative de préparer cette thèse remonte à une époque où l'informatique ne connaissait pas encore l'explosion actuelle. Quelques circonstances heureuses ont favorisé ma décision. D'abord, une initiation fortuite à la programmation, ensuite un stage en linguistique informatique (financé par l'A.C.C.T.) à Tunis et enfin une discussion avec deux de mes collègues à Kénitra ont fini par me conduire au Québec.

En appuyant au bon moment ma demande de me libérer de mes responsabilités d'enseignant et en acceptant *ipso facto* de travailler plus, les collègues du département de langue et de littérature françaises (Faculté des Lettres, Université Ibn Tofail, Kénitra, Maroc) ont posé le geste inaugural de cette thèse. Qu'ils en soient remerciés.

Monsieur Larbi Mezzine, ex-Doyen de la Faculté des Lettres et des Sciences Humaines à Kénitra (Maroc), m'a longuement soutenu et encouragé dans mes efforts. Je lui dois beaucoup. Madame Fouzia Ghissassi, Doyenne de la Faculté des Lettres et des Sciences Humaines de Kénitra, m'a également apporté son aide. Qu'elle trouve ici l'expression de ma gratitude.

Je reste redevable au département de Langues et Linguistique (Université Laval) qui m'a accordé un fonds de soutien très précieux dans mes premières années à l'Université.

Beaucoup des collègues de ce département m'ont donné leur sympathie et leur aide: c'est le moment de leur dire merci. Je pense à Conrad Ouellon, Pierre Auger, Jacques Ouellet,

Marise Ouellet, Claude Paradis, Claude Rochette, Pierre Martin, Jacques Ladouceur, Denise Deshaies, Daniel Leflem, et à d'autres.

Le personnel du département m'a toujours offert un service appréciable. Je citerai Yasmine, Mme Paquet (partie à la retraite), mais beaucoup d'autres ont autant de mérite.

Je reste aussi redevable aux responsables de l'AUPELF qui m'ont accordé une bourse de trois mois (été 96), dans le cadre du projet d'amélioration du logiciel Marqueur.

Mes remerciements vont également aux responsables du CTI (SIT actuellement) qui m'ont confié le projet d'écrire le code du programme Collocateur. Mario Toupin, qui avait fait l'analyse informatique du projet, m'avait apporté un soutien logistique et didactique des plus fructueux. René Gignac, à titre informel, n'a pas ménagé ses efforts pour m'éclairer sur certains aspects de la programmation.

J'ai une dette envers quatre enseignants de la Faculté des Sciences, département d'informatique: Ali Kacem, Nadir BelKhiter, Guy Mineau, et Gérard Simian qui m'ont apporté une aide appréciable.

Deux amis ont beaucoup contribué à la réalisation de mon projet de thèse: Abdelkrim Jebbour et Mostafa Shoul. Pour tout ce qu'ils ont fait, un simple mot: merci. J'ai également plaisir à remercier mes amis Ahmed Alioua qui prenait à cœur le progrès de mon travail, et Lahcen Darhouani qui m'a toujours encouragé. Fouzia et Gery m'ont offert leur aide et leur amitié. Merci aussi à Rachid Kandri.

Mes remerciements vont également à mes ami(e)s Mustapha Bencheikh, Assia Belhabib, Hafida El Amrani, Ali Lemnouar. Les uns et les autres ont beaucoup fait pour moi.

Enfin, je n'oublie pas que les membres de ma famille ont suivi l'évolution de ce projet avec beaucoup d'intérêt, sinon d'émotion. Je les remercie tous.

Ceux et celles que je n'ai pas cité(e)s n'ont pas démerité. Ils (elles) savent que j'ai pour eux (elles) toute l'amitié qu'ils (elles) méritent.

Table des matières

RÉSUMÉ 1	I
RÉSUMÉ 2	II
DÉDICACE	III
AVANT-PROPOS	IV
TABLE DES MATIÈRES	VI
LISTE DES TABLEAUX	XI
LISTE DES FIGURES	XIII
0. INTRODUCTION GÉNÉRALE	1
1. CONSIDÉRATIONS THÉORIQUES	7
A. INTRODUCTION	7
B. LA NOTION DE COHÉSION	7
1. <i>Place de la notion de cohésion</i>	7
2. <i>Définition et caractéristiques</i>	11
a) Cohésion vs cohérence	11
b) Les ressources de la cohésion	13
(1) Les ressources grammaticales	13
(a) La référence	13
(b) La substitution	14
(c) L'ellipse	15
(d) La conjonction	15
(2) Les ressources lexicales	16
(a) La répétition	16
(b) La collocation	18
c) La collocation: un phénomène de cohésion lexicale	19
(1) Cohésion structurelle vs Cohésion non structurelle	19
(a) Collocation non structurée	20
(b) Collocation structurée	21
(2) Cohésion: relation anaphorique vs relation cataphorique	25
(3) Cohésion endophorique vs cohésion exophorique	26
(4) Cohésion: une question de degrés	26
C. L'AMBIGUÏTÉ LEXICALE	27
1. <i>L'ambiguïté, une réalité linguistique</i>	28
2. <i>Les différentes réalisations de l'ambiguïté</i>	28
a) L'ambiguïté "virtuelle"	28
b) L'ambiguïté "effective"	29
c) "L'équivoque réelle"	29
3. <i>Homonymie vs polysémie</i>	32
4. <i>Perception humaine de l'ambiguïté</i>	33

5. Perception automatique de l'ambiguïté.....	36
D. STRATÉGIES DE DÉSAMBIGUÏSATION AUTOMATIQUE	38
1. Linguistique informatique	38
2. Exploitation de la cohésion lexicale	39
a) De l'insuffisance de la syntaxe et de la sémantique.....	39
b) L'Intelligence Artificielle: "promesses et réalités"	40
(1) Les mots experts.....	41
(2) Les mots Polaroid	43
(3) Les préférences sémantiques	45
(4) Des sources d'information multiples.....	45
c) Pour un contexte littéral.....	47
E. CONCLUSION	48
2. COHÉSION LEXICALE ET DÉSAMBIGUÏSATION	
AUTOMATIQUE: APPLICATION.....	50
A. INTRODUCTION.....	50
B. STRATÉGIE GÉNÉRALE DE DÉSAMBIGUÏSATION.....	50
1. Le programme <i>MARQUEUR</i>	56
a) Marqueur au service de la désambiguïstation	57
b) Marqueur au service d'autres applications.....	65
2. Le programme <i>COLLOCATEUR</i>	66
a) Collocateur au service de <i>LOCALEX</i>	69
b) Collocateur au service d'autres applications	70
3. Le programme <i>LOCALEX</i>	71
4. Homogénéité et interdépendance des programmes	76
5. La base des connaissances lexicales.....	79
C. CONCLUSION	81
3. MISE A L'ÉPREUVE ET RÉSULTATS	83
A. INTRODUCTION.....	83
B. QUESTIONS DE MÉTHODE	83
1. Démarche générale	84
2. Paramètres à observer	86
a) L'adéquation de la forme.....	86
b) L'adéquation de la catégorie et du lemme.....	86
c) L'adéquation graphique des collocations	87
(1) Les différents types d'adéquation.....	87
(a) Adéquation continue	87
(b) Adéquation discontinue.....	88
(c) Adéquation disloquée ou pseudo-collocation	88
(2) L'inadéquation et ses causes	89
d) Les verdicts de <i>Localex</i>	90
(1) Heureuses	90
(2) Malheureuses.....	90
(3) Indécidables	91

C. TEST DE LA VALIDITÉ DE LA DÉMARCHE	93
1. <i>Corpus et base des connaissances</i>	93
a) Le corpus.....	93
b) Choix des mots objet de la recherche.....	93
c) Les acceptions	95
d) Les collocations.....	95
2. <i>Les réponses de Localex</i>	96
3. <i>Évaluation des réponses de Localex</i>	101
4. <i>Force et limites de la stratégie de désambiguïsation</i>	107
a) Adéquation continue / discontinue / disloquée	107
b) Carence et surenchère de collocations.....	108
D. TEST DE LA REPRÉSENTATIVITÉ DE LA BASE	109
1. <i>Présentation du test</i>	109
2. <i>Résultats et commentaires</i>	112
E. SOLUTIONS PROPOSÉES	116
F. CONCLUSION	117
4. CONCLUSION GÉNÉRALE	119
5. BIBLIOGRAPHIE	124
A. LA LINGUISTIQUE TEXTUELLE.....	124
B. LA LINGUISTIQUE INFORMATIQUE	126
C. L'INFORMATIQUE.....	129
D. LA RECHERCHE DOCUMENTAIRE.....	130
6. ANNEXES	131
A. GUIDE D'UTILISATION DE MARQUEUR	131
1. <i>PRÉAMBULE</i>	131
2. <i>FONCTIONS</i>	132
a) Description.....	132
b) Utilisateurs ciblés	134
3. <i>INSTALLATION</i>	134
4. <i>UTILISATION</i>	134
a) Démarrage.....	134
b) Création d'un projet	138
c) Ouverture d'un projet	140
d) Segmentation du texte	141
e) Marquage automatique	141
f) Marquage interactif.....	142
g) Modification de la structure du dictionnaire.....	146
h) Propriétés additionnelles.....	148
i) Consultation des données	148
j) Révision du marquage.....	149
5. <i>QUELQUES CONSEILS</i>	153

B. GUIDE D'UTILISATION DE COLLOCATEUR	156
1. <i>PRÉAMBULE</i>	156
2. <i>FONCTIONS</i>	156
a) Description	156
b) Utilisateurs ciblés	157
3. <i>INSTALLATION</i>	158
4. <i>UTILISATION</i>	158
a) Démarrage	158
b) Création d'un projet	159
c) Ouverture d'un projet	160
d) Extraction des collocations	161
(1) Type de repérage	162
(a) Repérage simple	162
(b) Repérage multiple	163
(2) Enregistrement des collocations	163
(3) Paramètres de base	163
(a) Règles d'admissibilité	164
(b) Règles par défaut	165
(4) Paramètres complémentaires	166
(a) Longueur de l'extrait	166
(b) Longueur de la clé grammaticale	166
(c) Extraction par défaut	167
(d) Ponctuation	167
e) Consultation des données	167
C. GUIDE D'UTILISATION DE LOCALEX	171
1. <i>PRÉAMBULE</i>	171
2. <i>FONCTIONS</i>	171
a) Description	171
b) Utilisateurs ciblés	172
3. <i>INSTALLATION</i>	172
4. <i>UTILISATION</i>	172
a) Démarrage	172
b) Création d'un projet	173
c) Ouverture d'un projet	175
d) Localisation automatique	175
(1) Paramètres de la recherche	176
(2) Recherche conditionnée	178
(3) Base des connaissances	181
e) Consultation des fichiers	185
D. MOTS AMBIGUS ET LEUR CONTEXTE	190
1. <i>Corpus Sécurité Routière du Québec</i>	190
a) Le mot droit (s)	190
b) Condition	203
c) Examen	213
d) Remettre	219
e) Objet	226

f) Application	233
g) Avis	237
h) Remise	241
i) Pièce	243
j) Charge	246
k) Porter	248
2. <i>Corpus: Code Civil du Québec</i>	250
a) Droit	250
b) Condition	260
E. STRUCTURE DES FICHIERS	262

Liste des tableaux

TABLEAU 1 DIFFÉRENTES RÉALISATIONS DE LA COLLOCATION.	23
TABLEAU 2 STRUCTURE GLOBALE DE LA BASE DE DONNÉES.	53
TABLEAU 3 RÈGLES À SUIVRE POUR FORMER UN LEMME.	58
TABLEAU 4 DIFFÉRENCES DE SIGNIFICATION LIÉES À LA CATÉGORIE GRAMMATICALE.	59
TABLEAU 5 EXEMPLE DE MOTS LEMMATISÉS.	60
TABLEAU 6 EXEMPLE DE LEMMATISATION DE MOTS HOMONYMES.	61
TABLEAU 7 DÉCOUPAGE D'UN MOT COMPOSÉ.	61
TABLEAU 8 LES DIFFÉRENTS CAS DE FIGURE DANS LE MARQUAGE.	64
TABLEAU 9 RELATION DES FORMES ET DES LEMMES.	64
TABLEAU 10 UTILISATION DU FICHIER DES LIENS.	65
TABLEAU 11 RÈGLES D'ADMISSIBILITÉ DANS COLLOCATEUR.	68
TABLEAU 12 EXEMPLE DE COLLOCATIONS EXTRAITES À L'AIDE DE COLLOCATEUR	70
TABLEAU 13 RELATIONS ENTRE LES DIFFÉRENTES INFORMATIONS.	71
TABLEAU 14 CODES ADOPTÉS PAR LOCALEX.	72
TABLEAU 15 REPRÉSENTATION DE COLLOCATION DANS LES CONTEXTES GAUCHE ET DROIT.	75
TABLEAU 16 EXTENSIONS DES FICHIERS ET LEURS SIGNIFICATIONS.	78
TABLEAU 17 MODES D'UTILISATION DU SYSTÈME DE DÉSAMBIGUÏSATION LEXICALE.	79
TABLEAU 18 EXEMPLES D'ADÉQUATION CONTINUE.	88
TABLEAU 19 EXEMPLES D'ADÉQUATION DISCONTINUE.	88
TABLEAU 20 EXEMPLE DE LA PSEUDO-COLLOCATION.	89
TABLEAU 21 SIGNIFICATIONS DU MOT CONDITION.	91
TABLEAU 22 LES DIFFÉRENTS PARAMÈTRES À OBSERVER.	92
TABLEAU 23 LES COLLOCATIONS EXTRAITES AUTOMATIQUEMENT OU MANUELLEMENT.	95
TABLEAU 24 LES DIFFÉRENTS MOTS RETENUS ET LEUR FRÉQUENCE (TEST 1).	96
TABLEAU 25 MOTS RECHERCHÉS DANS LE CORPUS ET LEURS SIGNIFICATIONS RESPECTIVES. ...	99
TABLEAU 26 LES RÉPONSES DE LOCALEX.	100
TABLEAU 27 ÉVALUATION DES RÉPONSES DE LOCALEX.	101
TABLEAU 28 LES SUCCÈS ET LES ÉCHECS DE LOCALEX.	106
TABLEAU 29 COLLOCATIONS CONTINUES, DISCONTINUES, PSEUDO-COLLOCATIONS.	108
TABLEAU 30 SIGNIFICATIONS DU MOT "DROIT".	111
TABLEAU 31 COLLOCATIONS EMMAGASINÉES POUR LA SIGNIFICATION 1 DU MOT "DROIT".	111
TABLEAU 32 SIGNIFICATIONS DU MOT "CONDITION".	112
TABLEAU 33 COLLOCATIONS RELATIVES À LA SIGNIFICATION 1 DU MOT "CONDITION".	112
TABLEAU 34 LES DIFFÉRENTS MOTS RETENUS ET LEUR FRÉQUENCE (TEST 2).	113
TABLEAU 35 LES RÉPONSES DE LOCALEX (TEST 2).	113
TABLEAU 36 ÉVALUATION DES RÉPONSES DE LOCALEX (TEST 2).	114
TABLEAU 37 LES SUCCÈS ET LES ÉCHECS DE LOCALEX (TEST 2).	115
TABLEAU 38 LÉGENDE DES COULEURS IMPOSÉES PAR LOCALEX	180
TABLEAU 39 <i>LE MOT DROIT</i>	202
TABLEAU 40 <i>LE MOT CONDITION</i>	213
TABLEAU 41 LE MOT EXAMEN.	218

TABLEAU 42 LE MOT REMETTRE.....	226
TABLEAU 43 <i>LE MOT OBJET</i>	232
TABLEAU 44 <i>LE MOT APPLICATION</i>	236
TABLEAU 45 <i>LE MOT AVIS</i>	241
TABLEAU 46 LE MOT REMISE.....	243
TABLEAU 47 <i>LE MOT PIÈCE(S)</i>	245
TABLEAU 48 LE MOT CHARGE.....	248
TABLEAU 49 LE MOT PORTER.....	249
TABLEAU 50 LE MOT DROIT (TEST 2).....	259
TABLEAU 51 LE MOT CONDITION (TEST 2).....	261
TABLEAU 52 FORMES.DIC.....	262
TABLEAU 53 LIENS.LIE.....	262
TABLEAU 54 LEMMES.DC2.....	262
TABLEAU 55 ACCEPT.ACC.....	262
TABLEAU 56 COLLOCAT.COL.....	262
TABLEAU 57 SAAQ_6T.DBF.....	263
TABLEAU 58 CAT_G.CTG.....	263
TABLEAU 59 SAAQ.CLC.....	263

Liste des figures

FIGURE 1 L'AMBIGUÏTÉ EN CONTEXTE.....	32
FIGURE 2 COOPÉRATION ENTRE LES TROIS PROGRAMMES ET TRANSFERT DES FICHIERS.....	77
FIGURE 3 FENÊTRE D'ACCUEIL DE MARQUEUR	135
FIGURE 4 MENU PRINCIPAL	136
FIGURE 5 CRÉATION DE PROJET	139
FIGURE 6 OUVERTURE DE PROJET	140
FIGURE 7 MARQUAGE AUTOMATIQUE.....	142
FIGURE 8 MARQUAGE INTERACTIF.....	143
FIGURE 9 MODIFICATION DE LA STRUCTURE DU DICTIONNAIRE	147
FIGURE 10 SAISIE DE PROPRIÉTÉS ADDITIONNELLES	148
FIGURE 11 CONSULTATION DES DONNÉES.....	149
FIGURE 12 RÉVISION DE MARQUAGE, OPTIONS: TEXTE, UNIQUE.	150
FIGURE 13 RÉVISION, OPTIONS: TEXTE, MULTIPLE, <u>SANS</u> VISUALISATION DU CONTEXTE.....	151
FIGURE 14 RÉVISION, OPTIONS: TEXTE, MULTIPLE, <u>AVEC</u> VISUALISATION DU CONTEXTE	152
FIGURE 15 RÉVISION: OPTIONS: DICTIONNAIRE.	153
FIGURE 16 BUREAU DE MARQUEUR (MENU PRINCIPAL ET PARAMÈTRES ACTIFS)	159
FIGURE 17 CRÉATION DE PROJET	160
FIGURE 18 OUVERTURE DE PROJET	161
FIGURE 19 PARAMÈTRES.....	162
FIGURE 20 RÈGLES D'ADMISSIBILITÉ.....	164
FIGURE 21 RÈGLES PAR DÉFAUT.....	165
FIGURE 22 CONSULTATION DES DONNÉES: LE FICHIER MARQUÉ, EXTENSION DBF	168
FIGURE 23 FICHIER DES COLLOCATIONS EXTRAITES.	169
FIGURE 24 LISTE DES COLLOCATIONS CONTENANT LE MOT <u>DROIT</u>	170
FIGURE 25 MENU PRINCIPAL	173
FIGURE 26 CRÉATION DE PROJET	174
FIGURE 27 OUVERTURE DE PROJET	175
FIGURE 28 PARAMÈTRES.....	176
FIGURE 29 SÉLECTION DU MOT À RECHERCHER.....	177
FIGURE 30 SÉLECTION DE L'ACCEPTION DÉSIRÉE.	178
FIGURE 31 RÉCAPITULATION DES PARAMÈTRES DE RECHERCHE DANS LA GRILLE DE REQUÊTE.....	179
FIGURE 32 AFFICHAGE DU TEXTE ET DES OCCURRENCES REPÉRÉES.....	180
FIGURE 33 SÉLECTION DU LEMME À TRAITER.....	182
FIGURE 34. AJOUT D'ACCEPTIONS ET SÉLECTION DE L'ACCEPTION À TRAITER	183
FIGURE 35 SAISIE, AJOUT OU MODIFICATION DES COLLOCATIONS	184
FIGURE 36 LISTE DE COLLOCATIONS BRUTES PROVENANT DE COLLOCATEUR.	185
FIGURE 37 TEXTE MARQUÉ.	186
FIGURE 38 DICTIONNAIRE DES FORMES ET DICTIONNAIRE DES LEMMES RELIÉS.....	187
FIGURE 39 FICHIER DES ACCEPTIONS	188
FIGURE 40 FICHIER DES COLLOCATIONS.....	189

INTRODUCTION GÉNÉRALE

0. INTRODUCTION GÉNÉRALE

La cohésion d'un texte peut être vue comme l'ensemble des liens qui sous-tendent son unité et contribuent à sa soudure. Ces liens, qui sont de nature variée¹, sont assurés par les moyens lexico-grammaticaux du système linguistique. Certains de ces moyens, comme la référence anaphorique, sont prévisibles et peuvent faire l'objet d'une grammaire qui pourra en inventorier les marques et en systématiser l'emploi. Une démarche algorithmique pourra par la suite se servir de cette grammaire pour des visées d'automatisation. Nous ne préjugeons cependant en rien de la complexité de cette grammaire². D'autres ressources de la cohésion sont moins systématisables; c'est le cas des ressources lexicales. Elles sont par conséquent plus rétives à l'automatisation. Néanmoins, étant reconnaissables et suffisamment présentes dans les textes, elles offrent une matière dont nous sommes tenté de tirer profit.

Le présent travail se propose justement d'étudier les possibilités d'exploiter certaines ressources de la cohésion des textes dans une perspective d'automatisation. Notre attention s'arrêtera sur les ressources dites de *cohésion lexicale*, en particulier sur le phénomène de la collocation. Il s'agira de voir comment ces dernières peuvent être mises au service de l'automatisation de certaines tâches de traitement et d'analyse de textes. Ces tâches peuvent être nombreuses, mais notre étude s'occupera essentiellement de la désambiguïsation³ automatique des mots d'un texte donné. Désambiguïsation catégorielle, mais aussi sémantique. Notre réflexion sera concrétisée dans la réalisation de trois programmes qui, ensemble, montreront le degré de faisabilité d'un tel projet.

Automatiser certaines tâches concernant un texte donné présuppose qu'un certain nombre de conditions soient réunies. Outre la nécessité que ce texte soit disponible sous forme

¹ Ces liens peuvent être de nature grammaticale ou sémantique.

² Cf. B. Combettes (1975) et G. Hirst (1981).

³ Le terme *désambiguïsation* est considéré ici comme l'équivalent de *levée d'ambiguïté*. Bien que le premier terme puisse être jugé comme étant moins normatif, nous l'avons utilisé plus fréquemment car à notre sens il est plus commode.

d'enregistrement électronique⁴, il faudrait supposer également qu'il contienne suffisamment de marques régulières et que cette régularité puisse être cernée par une méthode ou par une autre.

Le point de départ est d'être en mesure de reconnaître dans un texte des formes en tant qu'unités discrètes. Grâce à un algorithme approprié, cette étape peut être franchie, mais non sans quelques compromis. La tâche devient cependant plus complexe dès l'instant où l'on se propose d'identifier automatiquement la catégorie ou la classe à laquelle appartient chacune des formes reconnues. Si le sujet humain peut reconnaître plus ou moins intuitivement ce qu'est, par exemple, un *nom* ou un *verbe*, la machine⁵, qui ne peut se contenter d'approximations, a besoin qu'on lui inculque, au moyen d'instructions très précises, les critères rigoureux⁶ qu'elle doit mettre en œuvre chaque fois qu'elle devra décider si une forme donnée constitue ou ne constitue pas un *nom*. C'est là un premier niveau de difficulté auquel nous apporterons une solution à notre manière; c'est l'étape du marquage des mots du texte. En fait, les problèmes qui se posent à ce niveau peuvent être, dans un premier temps, tellement complexes qu'ils finissent par être réhébilitaires. Au lieu de verser dans le découragement, nous croyons possible, voire légitime, dans notre cas, de régler provisoirement ces problèmes par une intervention humaine.

Mais le problème se pose avec plus d'acuité quand on désire chercher un mot (ou une séquence de mots) employé avec une signification précise à l'exclusion de toute autre signification. Le problème ici est qu'un mot peut avoir, dans l'absolu, plusieurs significations; et qu'il n'est pas évident de déterminer de façon automatique laquelle est attestée dans le contexte. Qui plus est, un mot peut aussi avoir des homonymes.

⁴ On connaît l'impact des techniques de numérisation, des banques de données déjà disponibles, sous différents supports.

⁵ L'ordinateur, quand il n'est pas pré-programmé à l'origine comme c'est le cas de certains calculateurs, ne travaillera qu'en fonction de ce qu'on va lui *inculquer*.

⁶ Dans une logique procédurale, les moindres détails sont prévus, et l'algorithme est un guide très pointilleux. Dans une logique déclarative, il y a une large part de décisions imprévisibles. *A fortiori*, une démarche connexioniste invoquerait un raisonnement proche de celui du sujet humain. Mais toutes

Pour être en mesure de repérer dans un texte libre⁷ un mot obéissant à une contrainte de sens, un système informatique fondé sur une analyse linguistique se doit de tenir compte de l'ambiguïté des mots. Nous croyons justement que la prise en compte des moyens linguistiques qui servent à assurer la cohésion lexicale du texte peut contribuer à la désambiguïstation des mots ambigus. C'est là le second niveau de difficulté que nous essaierons de résoudre.

Comme on peut s'y attendre, on a besoin de constituer une base de données qui recueillerait des informations lexicales. Cette tâche gagnerait elle-même à être automatisée, et c'est pourquoi nous avons considéré que c'est un troisième niveau de difficulté qui devrait être traité.

D'une manière générale, le présent travail vise l'étude des possibilités d'exploitation du phénomène de la collocation en tant que source de cohésion lexicale dans des traitements automatiques de textes. Plus spécifiquement, nous voulons montrer de quelle façon on peut exploiter ces moyens linguistiques dans un processus de désambiguïstation lexicale automatique. Notre problème sera, en premier lieu, de fournir le moyen (analyse linguistique et implémentation informatique) qui permettrait de repérer uniquement les occurrences d'un mot dont le sens correspond à la requête. On peut vouloir, par exemple, chercher les occurrences du mot «canard» avec le sens d'oiseau à l'exclusion de toutes les autres occurrences signifiant *journal* ou tout autre sens. Cet objectif inclut l'architecture de ce système et la réalisation d'un prototype de ce système, mais également le développement d'outils susceptibles d'aider d'autres chercheurs dans des travaux plus ou moins proches de notre préoccupation. Cela dit, la réalisation d'un tel projet ne va pas sans l'analyse linguistique de ces moyens linguistiques.

ces démarches se caractérisent par de la rigueur, la différence est dans l'interprétation qu'on donne à ce concept.

⁷ Par *texte libre*, nous désignons un texte continu non-compartmenté dans des champs fixes comme c'est le cas des données des questionnaires de fichiers par exemple. Le format exigé est le fichier texte. Nous n'avons pas engagé de solution pour les hypertextes ni pour les hyperdocuments (documents comprenant des objets multimédia).

Le système envisagé, qui se situe dans un cadre plus général de la compréhension automatique du langage, tiendra compte essentiellement des applications possibles dans le domaine des sciences de l'information et de la recherche documentaire, mais également de l'Analyse du Discours et de nombreux domaines concernés par le texte.

L'intérêt de la rencontre qui s'est faite ces dernières années entre la linguistique et l'informatique n'est plus à démontrer. Nous constatons en effet un engouement pour cette nouvelle discipline sans doute plus prononcé que celui que la linguistique théorique a connu il y a deux décennies. Cet enthousiasme s'est manifesté dans deux directions complémentaires. La première, qui relève de la *recherche fondamentale*, vise à trouver des solutions radicales aux problèmes posés par le traitement des langues naturelles et s'appuie sur une analyse de fond et sur des modèles linguistiques à portée générale. La seconde, qualifiée généralement d'*industries de la langue* ou de *génie linguistique*, sans dénigrer la rigueur théorique, ne s'interdit pas pour autant de recourir à des stratégies ponctuelles, adaptées au besoin. Le présent travail s'inscrit plutôt dans cette dernière position.

L'ordinateur, qui manipule des symboles malgré sa réputation d'opérer primordialement sur des chiffres, devait nécessairement un jour s'occuper des langues —elles-mêmes faites de symboles. En effet, alors même que les progrès technologiques dans ce domaine n'en étaient qu'à leur début, l'ambition des chercheurs était d'exiger de la machine les mêmes compétences observées chez les humains quant à l'utilisation du langage. Ces espoirs généreux, fondés sans doute sur une perception trop simpliste de la nature de la langue, devaient amener les chercheurs à poser le problème d'abord du côté linguistique. Pour ne citer qu'un aspect, et non des moindres, la linguistique informatique se heurte à la double difficulté qui consiste à trouver un modèle formel et une façon de représenter les connaissances. Non que les modèles informatiques soient pauvres, mais ce sont les connaissances qui échappent à la représentation, soit parce qu'elles sont incommensurables, soit parce que le réseau des relations qui leur est sous-jacent est inextricablement complexe.

Face à cette difficulté de recourir à un formalisme rigoureux ou d'imaginer une solution pour procéder à une représentation des connaissances d'une manière massive, une voie intermédiaire reste cependant possible. Nous suggérerons une façon relativement simple

d'associer des connaissances textuelles à des unités linguistiques de telle sorte que le système qui en résulte puisse accomplir certaines tâches sans passer par une analyse de fond du réseau des relations sous-jacentes aux connaissances manipulées.

Notre stratégie, qui pourrait s'inscrire dans la lignée des propositions de F. Rastier et alii (1994), consistera donc à répertorier des tranches de contexte qui constituent des occurrences particulières et de les utiliser comme des sources d'information discriminatoires. Nous ne prévoyons pas d'analyse de fond de la phrase, encore moins du texte. De ce point de vue, notre point de départ est le *discours*, qui est en quelque sorte l'aboutissement ultime du fonctionnement du système linguistique. En partant du discours pour aller vers la signification, c'est comme si on abordait le problème en se plaçant en amont du système linguistique. En fait, ce vers quoi nous nous orientons est une *théorie du contexte*⁸.

Dans un premier chapitre, nous donnons quelques jalons théoriques sur la notion de cohésion lexicale. Ce survol de la littérature sera pour nous l'occasion d'avancer quelques idées pour exploiter les ressources lexicales dans un processus de désambiguïsation. Dans un deuxième chapitre, nous exposerons notre stratégie personnelle pour répondre aux objectifs tracés. En fait, il s'agira d'un système composé de trois programmes. Nous montrerons comment ces derniers mettent en œuvre notre stratégie de désambiguïsation. Un troisième chapitre contiendra l'exposé de la mise à l'épreuve du système et les résultats des tests de validation. En fait, nous aurons à effectuer deux types de test: le premier aura pour objet un extrait du *Code de la Sécurité Routière*, le second s'appliquera à un extrait du *Code Civil du Québec*. Enfin, la conclusion devrait nous permettre d'évaluer notre travail, et de présenter les perspectives qui s'ouvrent à l'horizon.

⁸ Cf. F. Rastier (1994)

CHAPITRE PREMIER
CONSIDÉRATIONS THÉORIQUES

1. CONSIDÉRATIONS THÉORIQUES

A. INTRODUCTION

Ce chapitre comportera trois sections. Dans la première, intitulée "la notion de cohésion", nous tenterons de situer cette notion dans un cadre général; nous essaierons ensuite de la définir (en la distinguant de la notion de cohérence), de montrer ses différentes ressources, en particulier celles relatives au lexique, et enfin de discuter certaines de ses caractéristiques.

Dans la deuxième section, intitulée "l'ambiguïté lexicale", nous aborderons les différentes réalisations de l'ambiguïté lexicale, comment l'ambiguïté est perçue par le sujet humain et comment, en contraste, l'ordinateur "perçoit" l'ambiguïté. Nous parlerons également du problème de la polysémie et de l'homonymie.

La troisième section, intitulée "stratégies de désambiguïsation automatique", nous permettra d'évoquer quelques solutions au problème de la désambiguïsation retenues ou préconisées par d'autres chercheurs. Nous montrerons plus particulièrement les possibilités d'exploitation de la cohésion lexicale dans un processus de désambiguïsation. Tout en présentant ces propositions, nous essaierons d'en montrer les enjeux théoriques. Cette section sera aussi l'occasion d'introduire notre démarche personnelle. (en attendant de la développer en détail dans le deuxième chapitre).

B. LA NOTION DE COHÉSION

1. Place de la notion de cohésion

L'intérêt des chercheurs pour les notions de cohésion et de cohérence ne pouvait se manifester tant que la linguistique ne s'était pas décidée à reconnaître la nécessité de dépasser le cadre de la phrase. En effet, la linguistique est restée longtemps confinée dans la dimension de la phrase, le texte ne semblant pas constituer un objet digne d'une analyse rigoureuse. D'autre part, comme le note D. Maingueneau, une approche ouverte sur le texte devait rencontrer une difficulté majeure: il est difficile de faire une distinction entre un texte cohérent

et un texte incohérent dans la mesure où il est souvent possible de trouver un contexte qui justifierait presque n'importe quel texte et le rendrait par conséquent cohérent. Cette difficulté aurait détourné les chercheurs vers la détermination des procédures utilisées pour rendre un texte cohérent, au lieu d'essayer de cerner la compétence textuelle des sujets.

Une fois ces difficultés dépassées, deux domaines d'étude, au demeurant complémentaires, devaient voir le jour: l'Analyse du Discours et la Linguistique Textuelle.

Pour le premier, il a fallu d'une part, comme le mentionne D. Maingueneau (1976), que l'on remette en cause la dichotomie saussurienne *langue / parole*. Sachant que la langue est une réalité sociale et la parole une réalité individuelle, il ne devait pas y avoir de place pour le discours, "concept qui vise à déposséder le sujet parlant de son rôle central pour l'intégrer au fonctionnement d'énoncés, de textes dont les conditions de possibilité sont systématiquement articulés sur des formations idéologiques". Cf. D. Maingueneau (1976).

D'autre part, la linguistique du début du siècle s'était construite autour de la notion de système. Il fallait appliquer cette notion à l'étude de textes et c'est ce que les formalistes russes ont fait en appliquant le principe d'immanence.

Les formalistes russes (V. Propp, R. Jakobson) ont permis de dégager une logique des enchaînements transphrastiques dans le domaine du conte folklorique. Leur contribution a permis d'asseoir les fondements de l'analyse structurale des formes littéraires. Propp, par exemple, dans *Morphologie du conte* a montré qu'il était possible de construire un modèle élémentaire duquel on dérivait une infinité de contes folkloriques différents dans leur apparence.

Les travaux de cette école ont préparé la prise en compte par la linguistique de ce qu'on allait appeler le *discours*. Ils ont permis d'opérer une coupure avec "l'approche impressionniste et philologique" des oeuvres littéraires. Cf. D. Maingueneau (1976).

Mais l'Analyse du Discours n'a commencé véritablement à se constituer qu'à partir des années cinquante. Z. S. Harris en 1952 a proposé une extension des procédures de la linguistique distributionnelle américaine à des énoncés qui dépassent le cadre de la phrase.

D'un autre côté, dans les travaux de Roman Jakobson et d'Émile Benveniste sur l'énonciation, la langue n'est plus une entité univoque qui se réduit à une simple combinaison des signes. L'étude de la langue consiste maintenant à voir comment le sujet parlant s'inscrit dans les énoncés qu'il émet.

L'intérêt de l'Analyse du Discours pour la cohésion ou pour la cohérence des textes (la différence entre les deux notions n'est pas toujours explicitée) est venu tardivement. Selon le témoignage de D. Maingueneau, il y avait au départ une certaine "méfiance à l'égard des formes de continuité discursive". L'Analyse du Discours a, jusqu'à récemment, accordé la priorité à différentes approches (syntaxiques, lexicologiques et lexicographiques) mais des approches qui ne tenaient pas compte de l'unité cohésive du texte.

Le second domaine qui a émergé à la suite du dépassement de la dimension de la phrase est celui de la Linguistique Textuelle. Dans les premiers travaux de V. Dijk, par exemple, l'attitude adoptée face au texte est similaire à celle que les générativistes ont face à la phrase. Le postulat avancé est qu'il existe chez le locuteur une compétence qui lui permettrait de décider si une suite de phrases formerait un texte (à l'instar du locuteur qui jugerait d'une suite de mots comme formant ou ne formant pas une phrase).

La Linguistique Textuelle s'est également occupée du problème de la cohésion des textes. En témoignent les travaux de M.A.K. Halliday et R. Hasan dont en particulier leur livre : **Cohesion in English** publié en 1976. T. V. Dijk a écrit en 1972 un texte de base sur la grammaire textuelle: **Some aspects of text-Grammars** suivi de **Text and Context**; il faut citer aussi un livre important publié d'abord en allemand (1972) puis traduit en anglais (1981): **Introduction to Text Linguistics** de R. de Beaugrande et W. Dressler; un article célèbre de I. Bellert (1970), «*On a condition of the coherence of Texts*»; et enfin (sans que la liste soit exhaustive) un travail original de P. Blumenthal qui a consisté à déceler dans le texte ce que l'auteur appelle les **schémas de la cohésion**, sortes de régularités de certains patrons observables dans les textes.

Mais il faut noter l'émergence d'une seconde voie. Celle-ci consiste à s'arrêter sur les connexions interphrastiques. Toutes sortes de connecteurs font alors l'objet de ce type

d'études: connecteurs grammaticaux ou lexicaux, éléments anaphoriques, marques de temps, etc. On peut voir à ce sujet, par exemple, les travaux de O. Ducrot⁹. C'est dans cette approche qu'on pourrait placer le travail de M. A. K. Halliday et R. Hasan. Bien entendu, on ne doit pas croire que dans cette approche, on s'interdit de postuler qu'il y a dans le texte des contraintes qui s'expliqueraient par la macrostructure du texte. Ainsi, l'on doit dire que certains connecteurs sont sélectionnés en fonction du type de discours utilisé.

Pour compléter cette mise au point, il convient de situer la notion de cohésion dans les travaux qui ont trait à des traitements automatiques de la langue. Pour être plus exact, ce qui nous intéresse ici, c'est de voir si la notion de cohésion lexicale a recueilli l'attention d'autres chercheurs quant à une exploitation de cette ressource dans des tâches automatiques.

A la différence de la notion de cohérence qui semble échapper au contrôle de toute modélisation et *a fortiori* à une exploitation dans un programme, la notion de cohésion lexicale semble se prêter à différents objectifs liés à un traitement automatique. Certains chercheurs comme J. Morris (1988) et G. Hirst (1991) ont essayé de tirer parti de la cohésion lexicale pour dégager la structure du texte, voire une information sur sa cohérence. En revanche, les travaux faisant appel à cette notion pour les besoins de la désambiguïsation lexicale sont plutôt rares. G. Hirst (1991) en a fait la suggestion. Par ailleurs, plus nombreux sont les chercheurs intéressés par la levée de l'ambiguïté lexicale et qui recourent au concept de collocation. Ce dernier concept renvoie en fait à deux réalités différentes (soit au sens d'une association de mots disséminés à travers le texte et appartenant à un même champ notionnel, soit au sens d'expressions plus ou moins figées), mais rares sont ceux qui ont essayé de tirer profit de la notion de collocation au premier sens pour des objectifs de désambiguïsation¹⁰. En même temps, ceux qui ont traité de la question n'envisageaient pas le phénomène de la collocation au deuxième sens comme une notion qui participe à la cohésion lexicale, F. Smadja (1993) excepté. Dans le présent travail, nous utiliserons également le concept de

⁹ Entre autres, le livre **Les Mots du Discours** est assez représentatif de cette tendance. (Ducrot 1985).

¹⁰ La raison est compréhensible. Dans ce domaine, on est à la recherche de l'efficacité des programmes. Or, l'exploitation de la collocation au sens d'association de mots est visiblement moins efficace. Elle sera par conséquent beaucoup plus utilisée comme complémentaire à d'autres sources d'informations.

collocation, fondamentalement dans le deuxième sens du terme (expressions plus ou moins figées) mais nous l'aborderons comme un phénomène intégré au texte, et donc participant, à un certain niveau, à sa cohésion. Accessoirement, nous allons tenir compte de l'apport de la collocation prise au premier sens. Cet apport est conçu comme complémentaire.

Nous venons de donner quelques éléments qui aideront à situer le problème dans un contexte large. Dans ce qui suit, nous reviendrons progressivement sur les différentes questions que notre propos a soulevées.

2. Définition et caractéristiques

a) Cohésion vs cohérence

Les deux notions cohésion et cohérence sont parfois utilisées d'une manière confuse dans la littérature. C'est pourquoi il convient de préciser ce que nous entendons par l'une et par l'autre.

La *cohésion* est entendue ici au sens de l'ensemble des relations instaurées entre les différentes unités qui forment le texte par le biais de moyens lexico-grammaticaux. On dira qu'il y a relation cohésive entre deux unités dès lors que l'interprétation de l'une est tributaire de l'interprétation de l'autre. Il y a cohésion quand un élément du texte réfère à un autre (relation phorique) ou quand deux éléments renvoient l'un à l'autre en vertu d'une relation sémantique qui existe entre eux. La cohésion concerne l'unité du texte en tant que charpente de références, de liens, tous azimuts à l'oeuvre dans ou à travers le texte. Elle permet donc au texte d'être soudé ("stick together") et de fonctionner comme un seul bloc.

Il convient donc de dire —avec M.A.K. Halliday et R. Hasan (1976)— que la cohésion est une relation qui se fait entre un terme d'une phrase et ce qui précède ou suit). Ce type de relation est appelé simplement un *lien*. Tout le raisonnement sur la notion de cohésion est fondé sur ce concept de lien. Le pronom personnel, par exemple, n'est pas interprété en lui-même, mais dans sa relation avec le terme auquel il renvoie. Il y a, bien sûr, plusieurs types de

liens et chacun a son fonctionnement propre dans le texte. Ces liens constituent dans leur variété la forme lexico-grammaticale de la cohésion.

La cohésion concerne les relations de construction du texte, de sa 'texture'. Elle est constituée de l'ensemble des marques linguistiques représentées explicitement dans le texte. (le cas de l'ellipse peut être interprété comme la forme non marquée de la cohésion). C'est pourquoi leur exploitation informatique est possible.

La cohérence, en revanche, concerne la propriété de ce qui est relaté dans le texte d'être en harmonie avec notre conception des choses de ce monde, du bon sens, du topos¹¹.

La cohérence se présente comme l'ensemble des relations entre les différents concepts évoqués par le texte. Il s'agit donc des représentations cognitives produites dans le texte. On parlera de cohérence quand il y a une congruence sémantique entre deux séquences d'un discours (voir T. J. M. Sanders et al. 1992). Il est donc compréhensible qu'il n'y ait pas de mécanisme automatique pour relever ces relations. D'ailleurs, dans certains cas, leur identification échappe même aux sujets humains. Pour le comprendre, regardons l'exemple emprunté par J. Morris et G. Hirst (1991) à J. R. Hobbs (1978):

(1) *John can open Bill's safe.*

(2) *He knows the combination.*

La relation entre les deux énoncés peut être vue, en fonction du contexte, de la connaissance du monde et des croyances, comme une élaboration ou comme une explication.

La cohérence a attiré l'attention de nombreux chercheurs. Comme le mentionnent T. J. M. Sanders, W. P. M. Spooren et L. G. M. Noordman (1992), deux attitudes sont adoptées. La première met l'accent sur la cohérence du point de vue du contenu, ce type de

¹¹ Un des arguments qui nous fait pencher vers la distinction des deux notions cohésion et cohérence, c'est la tendance du lecteur à trouver un sens à des textes dont la forme est déstructurée. Généralement, dans de telles situations, on essaie de retrouver une certaine cohérence aux choses présentées pêle-mêle.

cohérence étant appelé continuité référentielle ou encore continuité thématique (cf. par exemple T. Givon 1983). On trouve dans la littérature trois types de continuité: (1) une coréférence au même ensemble d'entités (cf. par exemple V. Dijk, 1978), (2) une congruence sémantique entre deux unités de discours (Polanyi, 1986), (3) l'existence d'un patron de situations stéréotypées, comme par exemple la visite d'un restaurant (R. C. Schank et R. P. Abelson, 1977). La seconde approche met l'accent sur la relation qui existe entre deux segments ou plus. (cf. par exemple J. R. Hobbs 1983).

Un texte peut être cohésif sans être cohérent. Il suffit qu'on ait du mal à voir une relation entre deux propositions (au sens logique). On peut, à l'inverse, penser à des textes non-cohésifs et pourtant montrant une large part de cohérence. Ainsi, un enfant qui avoue: "*moi chocolat aime*" se fait comprendre aisément, même si ses parents peuvent se sentir obligés de lui inculquer quelques notions de cohésion.

b) Les ressources de la cohésion

La cohésion est assurée par un arsenal de ressources assez étendu. M. A. K. Halliday et R. Hasan les ont cataloguées sous cinq catégories. Dans notre présentation, nous plaçons les quatre premières sous l'étiquette: *Les ressources grammaticales* et la cinquième sous l'étiquette: *les ressources lexicales*.

(1) Les ressources grammaticales

(a) La référence

La cohésion est une relation établie entre une unité linguistique et une autre. Cette relation peut être assurée par certains mots grammaticaux qui sont en eux-mêmes vides de sens. Il s'agit ici des personnels, des démonstratifs et des comparatifs. Ces différentes catégories sont regroupées par M. A. K. Halliday et R. Hasan, à juste titre, sous l'étiquette *référence*. En effet, ces différentes parties du discours ont pour fonction essentielle de *référer* à autre chose. Le pronom personnel "il", une fois placé dans le discours, se met à indiquer que le référent dont il est question devrait être cherché dans l'environnement. L'exemple suivant illustre ce type de cohésion:

- *La nouvelle Citroën XM inaugure une nouvelle conduite. Elle anticipe les irrégularités de la route.*

L'exemple suivant est intéressant parce que le lecteur ne sait pas tout de suite à quoi réfère le pronom *elle*. En fait, il s'agit ici d'une cataphore.

- *Bien sûr, elle en a vu d'autres. Bien sûr, elle en a connu de pires: plus graves, plus sanglantes. Bien sûr, elle s'en remettra. Il n'empêche: c'est une bien sale journée que s'apprête à vivre la République.*¹²

L'environnement peut être de deux sortes: contextuel ou extra-textuel. S'il est contextuel, la référence est dite endophorique dans la mesure où la relation s'établit avec une unité linguistique qui se trouve dans le texte. Le pronom personnel indique qu'il s'agit en l'occurrence du même référent. La nature de la relation cohésive est la référence spécifique. M.A. K. Halliday et R. Hasan considèrent, avec raison, que c'est uniquement avec cette condition que la cohésion est assurée.

Dans le cas de l'extra-textuel, on parle de relation exophorique. La relation est établie avec un référent extérieur au texte. Sachant que la cohésion est une notion textuelle, il n'y a pas à proprement parler de relation cohésive dans ce cas.

Considérés hors contexte, les pronoms sont neutres. En effet, rien ne dit dans l'absolu qu'un pronom (ou toute autre particule assurant la référence) est endophorique ou exophorique. On devine la difficulté de les distinguer d'une manière automatique.

(b) La substitution

La substitution est une autre manière de relier dans un texte des unités linguistiques entre elles. La deuxième unité renvoie à la première.

¹² Exemple extrait du journal Le Monde du 21 mars 1998, page 34. Nous avons apporté de légères modifications au texte.

Il s'agit ici de mots qui ont cette propriété de référer à un autre mot en prenant sa place. Il y a par conséquent une conformité au niveau de la nature et de la fonction grammaticale.

Contrairement à la référence qui est une relation sémantique, la substitution est une relation d'ordre grammatical. L'exemple suivant illustre le cas de la substitution:

- Deux scénarios s'offrent à cet homme: partir ou rester. Il devrait choisir le premier.

La séquence *le premier* se substitue au verbe *partir*.

(c) L'ellipse

L'ellipse est aussi une forme par laquelle la cohésion se réalise. En vérité, comme le remarquent M. A. K. Halliday et R. Hasan, l'ellipse constitue avec la substitution le même phénomène. Leurs mécanismes diffèrent mais le processus est le même. Dans les deux cas, on interprète une marque textuelle en la rattachant à une autre qui précède. Dans le cas de l'ellipse, la marque en question est absente, mais précisément cette absence est l'indice du renvoi à quelque chose. Considérons l'exemple suivant:

Il avait menacé de démissionner; sa femme aussi.

La séquence *sa femme aussi* est normalement interprétée comme: *sa femme aussi a menacé de démissionner*.

(d) La conjonction

La catégorie de la conjonction peut également assurer une cohésion entre les mots du texte, mais uniquement d'une manière indirecte. Contrairement aux autres catégories déjà citées, la conjonction n'a pas d'effet anaphorique (ni cataphorique). Si elle est malgré cela retenue comme relation de cohésion, c'est parce que, disent les auteurs, elle constitue une indication qu'un contenu sémantique est à relier à un autre contenu sémantique. Illustrons cette relation par un exemple:

On sait que le premier tour n'est qu'un tour de chauffe, mais déjà l'ambiance est lourde.

Ici, la séquence *déjà l'ambiance est lourde* doit être interprétée en opposition à la première tranche de l'énoncé.

Cette relation sémantique peut d'ailleurs s'accommoder de différentes tournures langagières. Ainsi, dans une relation chronologique, on peut exprimer le même rapport d'antériorité en plaçant la proposition temporelle en tête de phrase ou à la fin de celle-ci.

(2) Les ressources lexicales

La cohésion est une relation complexe et multiforme. Cette complexité trouve son écho dans la diversité des ressources qui l'assurent. En plus des ressources grammaticales, il faudrait recenser celles qui sont fondées sur le lexique de la langue. Il s'agit là d'une liste ouverte et par définition infinie. On dira dans un premier temps qu'il y a cohésion lexicale dès lors qu'un mot réfère à un autre par le biais de la répétition, de la synonymie ou encore de l'hyponymie. Dans un cas comme dans l'autre, la relation de cohésion est évidente, d'autant plus qu'il peut y avoir co-référence entre les deux termes en question.

Mais cette coréférence n'est pas nécessaire à la cohésion. Sans avoir à référer au même objet, deux mots peuvent entretenir une relation qui contribue à leur unité. Il suffit parfois qu'une relation sémantique, même ténue, soit établie entre deux unités pour que l'interprétation de l'une soit tributaire de l'interprétation de l'autre.

Il s'ensuit que la cohésion lexicale est une relation très large dans laquelle on peut distinguer deux classes. La première est dite, selon la dénomination de M. A. K. Halliday et R. Hasan, *réitération*. La seconde, connue dans la littérature, est reprise par M. A. K. Halliday et R. Hasan sous le nom de *collocation*.

(a) La réitération

La réitération consiste dans la répétition dans le texte d'unités linguistiques. Cette répétition peut prendre plusieurs formes. La plus évidente est la répétition pure et simple. La

synonymie (ou proche synonymie), l'hyponymie, le mot générique sont autant de formes de réitération. A cela, il faudrait ajouter d'autres phénomènes comme les procédés définitoires (ou de reformulation) telles que la description spécifique¹³, l'explicitation etc.

Illustrons quelques-uns de ces différents cas:

1- *C'est l'histoire d'Alice. Sa mère vit au Brésil, son père à Singapour. Alice a un frère adoptif, un garçon handicapé mental qui s'appelle José.* (Répétition pure et simple).

2- *Désormais, on ne parlera plus de l'automobile comme avant. Fruit de cinq ans de recherches et de 7,5 milliards d'investissements, la nouvelle Citroën XM inaugure une nouvelle conduite¹⁴.* (Hyponymie ou hyponymie selon le terme envisagé en premier lieu d'analyse).

3- *Humble servante de son récit, Florence Seyvos passe le chiffon sur des personnages avec dévotion et cruauté.* (Description spécifique).

4- *Il aime se souvenir, ou plutôt ne pas oublier(..). Avec Chahine l'Egyptien, l'identité est un tempérament.* (Explicitation).

Cette réitération permet d'établir une co-référence au même objet désigné par la première unité. Il faut cependant ajouter que cette co-référence n'est pas indispensable. Ainsi, dans l'exemple suivant:

Why does this boy wriggle all the time? Other boys don't wriggle.

Pourquoi ce garçon gigote tout le temps? Les autres garçons ne gigotent pas.

¹³ Les sous-catégories (e) et (f) n'ont pas été signalées par M. A. K. Halliday et R. Hasan. Nous considérons d'ailleurs que la liste n'est pas exhaustive.

¹⁴ Page publicitaire parue dans le Nouvel Observateur du 16 -22 Novembre 1989, p. 140.

On peut remplacer *boys* par *children*. Il n'y aura par conséquent plus de coréférence; en revanche, l'effet cohésif reste le même. En partant d'une telle constatation, les auteurs sont amenés à élargir le concept de réitération.

(b) La collocation

Dans le modèle de la cohésion prôné par M. A. K. Halliday et R. Hasan (1976), la collocation se présente comme une extension du concept de réitération¹⁵. En effet, on peut constater que dans l'exemple précédent, en remplaçant le mot *boys* par un mot comme *girls*, la relation de cohésion est maintenue même si les mots reliés ne sont pas synonymes, et ne partagent pas non plus le même référent.

En observant ce phénomène, dans lequel la relation est de type *complémentarité*, les auteurs considèrent que la relation de cohésion est établie dès lors que deux items entretiennent entre eux une relation lexico-sémantique reconnaissable. Ce postulat va être de nouveau élargi pour admettre qu'il y a collocation dès lors que deux mots (ou plusieurs) ont tendance à apparaître dans le même environnement. Cette définition reprend celle qu'on trouve chez d'autres chercheurs comme par exemple Firth¹⁶ (1933) mais dans le présent cadre théorique, et pour la distinguer du concept de réitération, les auteurs l'assortissent de deux conditions:

1. les deux unités reliées ne sont pas co-référentes.
2. les unités en question ne constituent pas une réitération (ne sont ni une répétition, ni un synonyme etc..)

¹⁵ R. Hasan a révisé en 1984 le modèle de 1976 et a rejeté le concept de collocation. Elle a fondu le phénomène de collocation dans celui de la réitération.

¹⁶ Cf. Kirsten Malmkjær (éd.) (1991), *The Linguistics Encyclopedia*, London et New York, Routledge, pp. 301-304.. Par ailleurs, S. McRoy (1992) décrit ce phénomène comme une association de mots (*word association*) et le distingue de la notion de *collocation* qui, selon elle, recouvre le phénomène de l'*expression*.

Ainsi définie, cette notion recouvre toute une gamme de possibilités d'association et de rapprochement d'unités linguistiques.

c) La collocation: un phénomène de cohésion lexicale

Après avoir exposé les différentes ressources de la cohésion, il nous appartient maintenant de mieux éclairer cette notion de collocation, en dégagant ses caractéristiques. En partant de problèmes théoriques soulevés d'une manière générale dans M. A. K. Halliday et R. Hasan, nous discuterons dans ce qui suit les points théoriques suivants:

1. Cohésion structurelle / cohésion non structurelle.
2. Cohésion: relation anaphorique / cataphorique.
3. Cohésion endophorique / cohésion exophorique.

Ces éléments théoriques nous permettent de mieux saisir l'intérêt que présente la notion de collocation pour l'automatisation. En fait, les questions qui seront discutées ici peuvent s'appliquer également aux autres ressources de la cohésion.

(1) Cohésion structurelle vs Cohésion non structurelle

Dans leur modèle, M. A. K. Halliday et R. Hasan considèrent que la cohésion est une relation indépendante de la structure. Ils ne rejettent pas l'idée que la structure de la phrase apporte une part importante à la cohésion d'un énoncé (et par là au texte), mais ils considèrent que la notion de cohésion est plus générale que celle de la structure de la phrase. Cette conception de la cohésion a pour conséquence que la collocation semble ne pas inclure le phénomène connu sous le nom d'*expression*¹⁷. M. A. K. Halliday et R. Hasan laissent ce phénomène dans l'ombre, sans doute parce qu'il est trop structuré à leurs yeux.

¹⁷ Comme on le verra plus loin, une expression peut être figée comme *perdre le nord* ou plus ou moins figée comme *exercer un droit*.

Le phénomène de la collocation du type *expression* apporte sa part dans l'unité du texte. Les mots qui composent l'expression sont intrinsèquement cohésifs, mais la cohésion ainsi créée se manifeste uniquement dans les limites de la phrase. Ces mots n'ont donc pas cette vertu de souder entre eux les maillons du texte que sont les phrases. Leur intervention est plus localisée, et forme à travers le texte des îlots sporadiques, interagissant avec les différents réseaux constitutifs du texte.

Nous considérons cependant que ce type de collocation, c'est-à-dire une collocation dont l'effet cohésif provient beaucoup plus de propriétés syntaxiques et d'une disposition structurée des unités linguistiques, mérite d'être étudié sous l'angle de la cohésion. D'autant plus que c'est un phénomène qui doit être distingué de la structure de la phrase.

Nous montrerons, en conséquence, que le phénomène de l'expression, qui est une des formes de la collocation, intervient dans la formation de la cohésion. Son intervention se situe à un niveau différent de celui des autres formes de la collocation, certes, mais son apport à la cohésion du texte mérite d'être souligné. Il convient de distinguer entre deux types de collocation. C'est ce que nous verrons dans les deux points suivants.

(a) Collocation non structurée

C'est une association de mots qui ont tendance à apparaître dans le même environnement grâce à leurs relations sémantiques. Ce lien sémantique est présent dans le système linguistique et se manifeste de différentes manières dans le discours. La cohésion est vue ici comme une relation indépendante de la structure; elle peut se surajouter à elle mais ne se confond pas avec elle.

Comme nous l'avons mentionné précédemment, M. A. K. Halliday et R. Hasan soutiennent que la cohésion est un phénomène indépendant de la structure. Leur raisonnement est le suivant:

D'une part, si l'on considère que le texte est une unité composée de phrases, il serait difficile d'affirmer que cette unité constitue une structure globale dont les composantes seraient les phrases. En effet, aucun texte ne possède une structure telle qu'on l'entend quand

on parle de structure dans une phrase. Certes, on peut parler de structure narrative d'un texte littéraire, mais celle-ci n'est pas suffisamment rigoureuse. Elle reste, en revanche, complémentaire à la constitution de la textualité.

D'autre part, au niveau de la phrase elle-même, on peut constater qu'elle est régie par une structure que la syntaxe modélise et décrit. L'utilisation de certaines catégories est prévisible dans la mesure où, à chaque endroit de la phrase, on peut calculer les possibilités de combinaison sur le plan syntagmatique et les possibilités de commutation sur le plan paradigmatique. Ce comportement des unités qui composent la phrase fait que ces unités sont solidement ancrées dans une structure qui constitue en fait le garant de leur cohésion. Par ailleurs, les règles grammaticales interviennent pour réguler l'utilisation de termes cohésifs. Quand on veut, par exemple, référer une seconde fois à un nom déjà cité dans la phrase, on doit utiliser un pronom.

La structure de la phrase semble donc assurer la cohésion, mais selon M. A. K. Halliday et R. Hasan, cette exigence est trompeuse puisque seules quelques instances cohésives peuvent être traitées structurellement, et seulement quand les deux termes (celui qui réfère et celui auquel il est référé) sont présents dans la même phrase, il s'agit en particulier des éléments cohésifs catégorisés sous l'étiquette de la référence. On peut d'ailleurs montrer que les relations cohésives sont les mêmes quand les deux termes sont présents dans la même phrase ou non. Dans l'exemple suivant, les éléments *him* et *his* renvoient à *amiral*, même si on séparerait les deux tranches de la phrase par une ponctuation forte.

*If you happen to meet the admiral, don't tell him his ship's gone
down.¹⁸*

(b) Collocation structurée

C'est ce qui équivaut à une expression, ou selon d'autres terminologies, à une *lexie* ou à un *idiome*. Sa contribution à la cohésion du texte est différente de celle de la collocation au sens précédent. Il s'agit ici d'un autre niveau de la cohésion. Les mots qui forment une

¹⁸ Si vous rencontrez l'amiral, ne lui dites pas que son bateau a coulé.

expression entretiennent une relation de cohésion entre eux, et cette relation est d'ordre structurel. La cohésion, dans ce cas, se situe au niveau lexical¹⁹. Étant structurelle, elle se manifeste par conséquent dans les limites de la phrase. Elle permet de souder non pas les phrases entre elles, mais les composants mêmes de la phrase. Son intervention est locale. Grâce à elle, la structure de la phrase est consolidée.

Si l'on convient de ce qui précède, on devra admettre que si la cohésion agit généralement d'une manière indépendante de la structure, elle n'intervient pas moins au niveau structurel dans le cas des collocations au sens d'expressions.

Nous pensons aux collocations au sens d'expressions figées, mais également à des expressions dont on maîtrise mal le degré de figement. En effet, les expressions figées sont une sorte d'associations intégrées et adoptées par le système linguistique. À l'instar des synonymes, des antonymes et autres phénomènes du même genre, les unités qui forment une expression figée *colloquent* à un niveau qui se situe en deçà de la réalisation dans le discours, i.e. le niveau du système linguistique. C'est pourquoi ces unités sont fortement prévisibles (la présence de l'une implique la nécessité de l'autre), et c'est pourquoi nous les considérons à la fois comme relevant de la structure et comme susceptibles d'assurer la cohésion. Vues comme des blocs pré-construits, ces expressions sont d'emblée cohésives, en vertu même du figement de leur structure.

L'effet cohésif vient du fait que les mots interagissent (entretiennent un réseau de relations qui jouent au niveau de l'interprétation de ces mots) parce que déjà reliés dans le système linguistique, et donc reconnus d'emblée en tant que syntagmes. Nombreuses sont les expressions idiomatiques qui exercent un effet de cohésion en vertu de la soudure même des mots dont elles sont composées.

Il peut arriver que cet effet soit doublé d'un autre, si l'on tient compte du fait que les mots qui forment l'expression peuvent être en même temps reliés sémantiquement, en étant par exemple membres d'un même champ notionnel. Dans de tels cas, la cohésion est assurée à

¹⁹ F. Smadja (1993)

deux niveaux: un niveau structurel (rapport syntagmatique) et un niveau non structurel (rapport sémantique ou notionnel). Les exemples suivants illustrent ce phénomène:

- *Il voit rouge.*
- *Il faut appeler un chat un chat*
- *Prendre le taureau par les cornes*

Le tableau suivant montre les différentes réalisations de la collocation.

COLLOCATION			
Avec rapport notionnel		Sans rapport notionnel	
Intraphrastique		Transphrastique	Intraphrastique
<i>Structurée</i> (2)	<i>Non structurée</i> (1), (2)	<i>Non structurée</i> (3)	<i>Structurée</i> (4)

Tableau 1 Différentes réalisations de la collocation.

(1) *Il aimait jouer et c'est pourquoi il était content de voir son frère sortir le ballon.*

(2) *Il a vu rouge.*

(3) *Il aimait jouer, cela est sûr; il était d'ailleurs content de voir son frère sortir le ballon.*

(4) *Il faut battre le fer tant qu'il est chaud.*

En considérant l'exemple (1), on constate que les mots *jouer* et *ballon* ont une relation sémantique qui peut être perçue indépendamment du contexte. Elle est absolue. A ce titre, ces mots sont en collocation au sens non structurel du terme. On pourrait répliquer que ces deux mots appartiennent à la même phrase, et qu'ils sont par conséquent intégrés à une structure. Notre réponse est que l'appartenance à la structure de la phrase ne suffit pas. Car il faut encore que cette relation structurelle soit lexicalisée. *Jouer* et *ballon* ne forment pas de

collocation du point structurel parce que rien n'indique qu'ils sont reconnus comme faisant partie d'un seul bloc.

Dans (2), il y a collocation à deux titres. Le premier parce que les mots *vu* et *rouge* sont liés sémantiquement. Le second, parce qu'ils font partie d'un syntagme lexicalisé.

L'exemple (3) illustre un autre cas de collocation non structurée. En effet, les deux mots soulignés sont liés sémantiquement. D'un autre côté, ils relèvent de phrases différentes (à supposer que le point-virgule entame une nouvelle phrase), et la question de leur lexicalisation ne se pose pas.

Enfin, l'exemple (4) illustre le cas d'une collocation structurée (il y a lexicalisation), sans rapport notionnel.

Les collocations structurées peuvent elles-mêmes prendre plusieurs formes. On peut penser à différentes manières de les classer. Nous avons retenu, en la simplifiant, celle donnée par F. Smadja (1993)²⁰ qui a distingué trois types de collocations:

1. Les collocations prédicatives: C'est le cas de deux mots utilisés ensemble d'une manière répétitive et partageant la même relation syntaxique. Par exemple, un nom est fréquemment utilisé comme l'objet d'un verbe. Prendre une décision, exercer un droit, réussir un examen.
2. Les expressions figées. Par exemple: machine à vapeur. Le moins qu'on puisse dire, c'est que le sens de ces expressions n'est pas toujours décomposable.

Pour anticiper sur notre stratégie de désambiguïsation lexicale et sur l'exploitation que nous comptons faire de ce type de collocations, nous précisons que pour des raisons d'efficacité, nous avons privilégié le recours aux collocations structurées. Quant aux

²⁰ Smadja (1993) a admis un troisième type de collocation. Il s'agit de séquences qui obéissent à des schèmes particuliers (phrasal templates).

collocations non structurées, nous les avons considérées comme des sources d'informations d'appoint. D'un autre côté, dans les collocations structurées, celles qui nous ont servi le plus sont les collocations prédicatives.

(2) Cohésion: relation anaphorique vs relation cataphorique

La cohésion est fondée sur le lien établi entre une unité linguistique et une autre. Ce lien peut se faire dans deux directions. Une unité peut référer à une autre unité qui se trouve à gauche dans le texte; c'est le cas de l'anaphore. Dans la direction opposée, on parlera de cataphore.

Bien que la relation cataphorique soit techniquement possible, dans la pratique, elle est rarement utilisée. En effet, la cohésion semble être foncièrement anaphorique. Ceci est évident pour ce qui concerne la catégorie de la référence, mais reste vrai pour les autres catégories.

La collocation aurait de ce point de vue un statut particulier. Prise au sens d'association libre de mots du même champ notionnel, une unité en collocation ne porte pas en soi une indication de la nécessité de la rattacher à une autre pour être en mesure de l'interpréter. Dans l'absolu, la relation ne peut être vue en termes de direction. Mais sa présence à proximité d'autres unités du même champ est une indication que le lien est à établir dans la direction imposée par la lecture. S'il s'agit d'une deuxième lecture, l'interprétant peut percevoir les relations établies dans une double direction.

Prise au sens de lexie ou d'expression figée, la collocation est difficilement explicable en termes de relations phoriques. Les mots ici sont reliés dans le système linguistique (en langue); cette relation structurelle est scellée dans l'absolu. Le processus de lecture peut permettre de constater l'articulation d'une unité avec une autre, plutôt dans une double direction (le processus de lecture admet un va-et-vient).

L'intérêt de savoir la direction d'une relation cohésive se justifie quand on compte utiliser ces ressources au profit de la désambiguïsation lexicale automatique. En effet, il serait pertinent, une fois le mot ambigu repéré, de savoir s'il faut traquer les mots en collocation à gauche ou à droite. En fait, pour les collocations structurées, cela dépendra de l'expression envisagée, alors que pour les collocations non structurées, on doit investir les deux directions.

(3) Cohésion endophorique vs cohésion exophorique

Une référence est dite endophorique si le terme référé se trouve dans le texte. Elle est dite exophorique si le terme référé n'est pas présent dans le texte, et qu'il s'agit par conséquent d'un élément de la situation.

La relation endophorique relie les éléments du texte entre eux, et en raison de cela, elle est la seule à considérer comme assurant directement la cohésion du texte. En reliant les éléments du texte au contexte extratextuel, la relation exophorique semble, en revanche, ne pas agir directement au niveau de la cohésion du texte. Elle contribue toutefois à la création du texte. On est tenté d'ajouter que la référence exophorique joue un rôle dans le maintien de la cohérence du texte.

Comme notre intérêt va vers les collocations, nous retiendrons uniquement celles qui sont susceptibles d'intervenir d'une manière endophorique. En fait, le problème ne se pose que pour les collocations non structurées.

(4) Cohésion: une question de degrés

La cohésion se manifeste dans les textes d'une manière plus ou moins forte. Il est rare qu'un texte soit complètement dépourvu de cohésion. L'effort interprétatif est tel que le lecteur trouve souvent une cohésion dans ce qui semble être une suite décousue de mots. Le degré zéro de la cohésion est repoussé par la hantise du lecteur de voir des liens partout. D'un autre côté, il est difficile d'arrêter un niveau de cohésion comme étant l'idéal. La cohésion est une question de degrés. Cela est encore plus vrai quand on traite de la collocation. En effet, les relations sémantiques entretenues entre les mots sont, par excellence, variables et subjectives.

D'une manière générale, le fait de pouvoir multiplier les ressources est significatif de l'absence d'un niveau stable de cohésion. Dans les textes juridiques par exemple, on s'ingénie à atteindre le plus haut degré de cohésion²¹.

Dans le cas de la collocation non structurée, le caractère graduel de la cohésion dépend de plusieurs facteurs dont la solidité des liens sémantiques entre des unités appartenant au même champ notionnel ou encore la tendance de ces unités à apparaître dans le même environnement textuel. Mais force est de constater que ces liens sont d'une grande complexité et qu'on ne peut les envisager autrement qu'à travers une gradation.

Dans le cas de la collocation structurée, la soudure des mots entre eux est proportionnelle au degré de figement de la collocation envisagée. S'il est vrai que certaines expressions manifestent un large degré de figement, beaucoup d'entre elles se présentent comme des syntagmes plus ou moins figés. Les critères qui permettent de trancher sur certains cas douteux ne sont pas faciles à établir.

Pour qui veut tirer profit de la cohésion lexicale dans un processus de désambiguïsation automatique, il devra tenir compte de ces distinctions.

C. L'AMBIGUÏTÉ LEXICALE

La notion de cohésion que nous venons de développer se construit en partie grâce au réseau des relations qui existent dans le système linguistique ou qui se créent dans le discours entre les unités lexicales de la langue. Mais en situation d'ambiguïté, il y a comme un effet boomerang qui fait que la cohésion lexicale participe à la désambiguïsation de l'unité ambiguë²². Nous montrerons dans le chapitre suivant comment de proche en proche la cohésion lexicale intervient dans la détermination du sens et en particulier le rôle qu'elle peut

²¹ Il est intéressant de ce point de vue de voir qu'au Canada anglophone, par exemple, on s'appuie sur la jurisprudence plutôt que sur une législation, comme si cette dernière ne pouvait fournir suffisamment de cohésion et de cohérence pour régler les litiges. Voir l'article "Propos sur la traduction juridique" de Pierre LABRANCHE (1993) où l'auteur apporte des réflexions intéressantes sur le discours juridique.

²² De ce point de vue, la cohésion est un support incontournable de la cohérence.

théoriques susceptibles de nous éclairer sur les choix à prendre ultérieurement dans la stratégie de désambiguïstation lexicale automatique.

1. L'ambiguïté, une réalité linguistique

Si chaque unité linguistique devait correspondre à une seule signification, il n'y aurait pas lieu de parler d'ambiguïté lexicale; mais le nombre de mots augmenterait en conséquence. La langue procédant de l'économie, il est primordial que le signe linguistique ne soit pas le lieu de l'univocité systématique (à chaque forme, une signification et à chaque signification une forme). L'univocité existe cependant pour bon nombre de mots, particulièrement dans la terminologie spécialisée.

Cette réalité linguistique engendre forcément des situations de confusion. Pour n'évoquer que le niveau lexical, une même unité peut être le lieu d'une collision de plusieurs significations. Vues dans l'absolu, celles-ci sont d'une part distinctes et d'autre part mutuellement exclusives (C. Fuchs); placées dans le contexte, elles sont toutes écartées sauf une (nous faisons abstraction du cas où il y a ambiguïté effective). La question est alors de savoir à quel niveau se situe l'ambiguïté.

2. Les différentes réalisations de l'ambiguïté

Nous admettons à la suite de C. Fuchs, à qui nous empruntons la distinction et la terminologie, trois modes de réalisation de l'ambiguïté: *l'ambiguïté virtuelle*, *l'ambiguïté effective* et *l'équivoque réelle*.

a) *L'ambiguïté "virtuelle"*

Quand une unité linguistique est envisagée dans la langue (d'une manière absolue), elle peut avoir plusieurs interprétations. On peut dire que cette unité est *potentiellement* ou *virtuellement* ambiguë. Nombreuses sont les unités qui en dehors du contexte linguistique et du contexte situationnel, enclenchent plus d'une signification. Le mot *bière* en français fait penser à une certaine boisson alcoolisée mais également à un cercueil. Le mot *grève* renvoie à "un terrain plat (formé de sable, gravier) situé au bord de la mer ou d'un cours d'eau" mais

"un terrain plat (formé de sable, gravier) situé au bord de la mer ou d'un cours d'eau" mais également à "une cessation volontaire et collective du travail décidée par des salariés pour obtenir des avantages matériels" (Le Petit Robert, éd. 1992). Cette pluralité de sens, partant d'une même unité linguistique, n'est que virtuelle. On devrait dire *provisoire*, en tout cas *suspendue* dans le temps. Car viendra le moment où cette pluralité cède le pas à l'unicité; ce moment, c'est la mise en contexte, le *discours*. On passe d'une compétence éclatée à une performance unitaire. C'est ainsi que le mot *grève*, initialement doté d'une double signification, se voit aiguillé vers une seule:

Les salariés ont débattu de l'arrêt de la grève.

b) L'ambiguïté "effective"

Comme on vient de le signaler, il faut généralement attendre que l'unité en question soit insérée dans un environnement donné pour que le récepteur soit conduit vers une signification ou une autre. Mais il arrive que l'environnement textuel ne suffise pas. Quand c'est le cas et que la seule analyse linguistique (ne faisant intervenir que les contraintes de langue) ne réussit pas à lever le voile, on est alors en présence d'une ambiguïté *effective*. Ainsi, la phrase suivante laisse-t-elle le récepteur dans l'embarras, puisque l'une et l'autre des deux acceptions potentielles du mot *grève* sont plausibles.

*Les salariés, qui parlaient beaucoup, s'étendaient depuis une heure sur la grève.*²³

c) "L'équivoque réelle"

On cherche généralement à lever une ambiguïté effective en s'appuyant sur un contexte plus large qui est celui de la situation d'énonciation. Si malgré tout, ce dernier ne nous aide guère à trancher, il ne reste plus qu'à reconnaître que nous sommes face à une ambiguïté insoluble. C. Fuchs appelle ce type d'ambiguïté une *équivoque réelle*. Dit autrement, toute ambiguïté effective devient une équivoque réelle pour peu que le contexte extra-

²³ Naturellement, nous supposons ici que le récepteur saisit les nuances du français et qu'en l'occurrence, il a conscience que le mot *s'étendre* est lui-même à double sens.

linguistique ne vienne pas au secours du récepteur pour lever le voile. Le plus souvent, le contexte vient relayer nos connaissances linguistiques et de ce fait, les ambiguïtés effectives que nous rencontrons cessent de nous préoccuper. Il reste que beaucoup de textes tombent sous le coup de l'équivoque: nombreuses sont les conversations dont nous avons manqué quelques éléments, ou des textes dont il nous manque quelque information soit sur les conditions de production soit sur les acteurs mis en jeu et qui par conséquent nous laissent perplexes quant à la bonne interprétation à leur donner. En plus d'une analyse linguistique susceptible de repérer les ambiguïtés effectives, nous éprouvons souvent le besoin de faire appel à une connaissance du monde.

Ceci étant dit, l'équivoque ne s'appuie pas toujours, remarque C. Fuchs, sur une ambiguïté effective. En effet, il peut y avoir équivoque quand le récepteur n'arrive pas à associer le bon référent à un signe linguistique tel que le nom propre. Ainsi que le montre l'exemple suivant, pris chez C. Fuchs, l'équivoque relative à l'identité de *Marie* ne s'explique pas par une ambiguïté d'ordre linguistique.

- *Je suis l'époux de Marie.*

Pour anticiper sur la démarche que nous allons suivre dans la désambiguïsation automatique, remarquons que si cette forme d'extrême ambiguïté pose problème au récepteur humain, les choses sont *a fortiori* plus complexes pour le traitement automatique.

Loin de nous, par conséquent, l'idée de proposer une quelconque méthode pour résoudre les cas d'équivoque réelle, puisqu'on ne peut pas demander à la machine de dénouer des situations que l'homme lui-même n'arrive pas à clarifier.

En ce qui concerne l'ambiguïté effective, notre objectif serait, non pas d'en faire une analyse linguistique complète, mais d'esquisser un niveau d'analyse qui serait à même d'empêcher les ambiguïtés virtuelles de se transformer en ambiguïtés effectives. Il faut préciser que, quand elle est dépourvue de la moindre connaissance sur le contexte et qu'elle ne compte que sur un dictionnaire de lemmes, la machine aurait tendance à considérer toute unité ambiguë dans le dictionnaire comme étant également ambiguë à l'endroit où elle la rencontre.

Sans le contexte, le langage est, aux yeux de la machine, beaucoup plus que pour l'homme, source de confusion et d'indécision. Chaque unité est plurielle. Là où un lecteur humain, même quand il ne compte pas sur le contexte proche (microstructure), voit une forme usuelle, triviale et univoque (sans doute parce qu'il fait intervenir des paramètres de macrostructure), la machine s'applique à examiner toutes les significations. Des plus rationnelles à nos yeux à la plus invraisemblable. Pourvue de règles sur le contexte, la machine pourrait réduire les ambiguïtés virtuelles. Mais même dans ce cas, il y a toujours le risque que de nouvelles ambiguïtés s'ajoutent ou se créent pour peu que la règle soit trop forte. Il s'agira d'ambiguïtés dites *artefact*. L'affinement des règles, souligne C. Fuchs (1995: 17), force paradoxalement la machine à examiner les cas les moins attendus²⁴. Un phénomène de surgénération se produit et a comme conséquence immédiate de faire exploser les combinaisons.

L'ambiguïté est effective seulement par rapport à un contexte donné. Si on élargit ce contexte, elle peut être levée. On pourrait schématiser ce postulat par une courbe qui représente un degré d'ambiguïté proportionnel à l'étendue du contexte (voir Figure 1). Réduit à zéro, le contexte n'apporte aucune contribution et laisse le mot aussi ambigu qu'il puisse être étant placé dans un dictionnaire. Formé d'une combinaison dite collocation, le contexte devient source d'information et apporte son lot au processus de désambiguïstation. La phrase, ensuite le texte, et enfin la situation de production sont autant de recours possibles, chacun plus riche que l'autre.

²⁴ Voir l'exemple du groupe nominal dans Fuchs (1995).

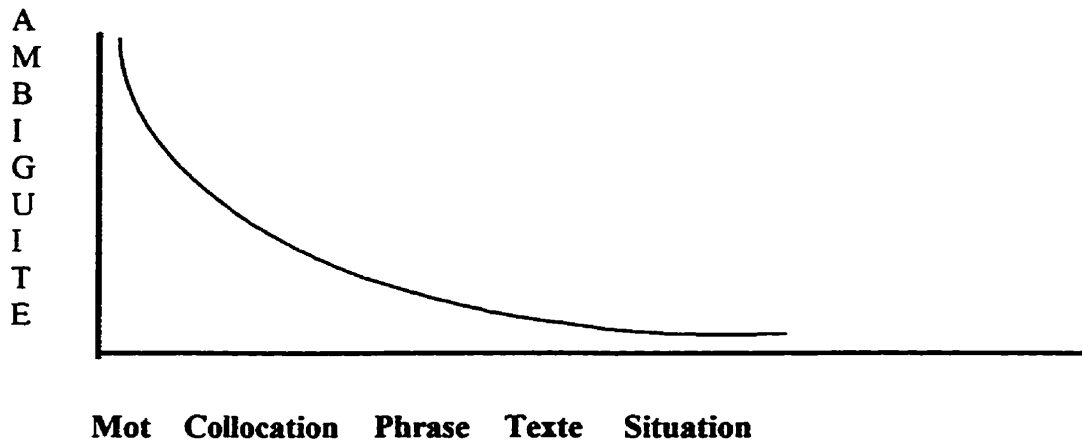


Figure 1 L'ambiguïté en contexte

Mais force est de constater que le schéma est trop simpliste. Car, comme le remarque C. Fuchs, il peut arriver que l'élargissement peut amener d'autres possibilités d'ambiguïté. Inversement, et c'est là la force de la stratégie que nous adoptons, une simple collocation peut suffire à déterminer la signification voulue. Sans parler des mots composés dans lesquels le sens n'est pas déduit d'une manière *componentielle* (exemple: *Hôtel de ville*), il y a bien des expressions dont le degré de figement est très variable, et qui constituent en soi un indice de signification des plus sûrs (exemple: *exercer un droit*).

Généralement, c'est la phrase qui constitue un contexte approprié pour lever une ambiguïté. Dans notre stratégie, nous utiliserons un contexte moins élaboré et moins étendu: Comme déjà mentionné, nous investirons le concept des collocations, fondamentalement dans les limites de la phrase. Le système développé se prête, cependant, à une exploration qui va au-delà de ces limites.

3. Homonymie vs polysémie

Un autre aspect de l'ambiguïté est relatif à la conception que l'on se fait de l'identité du mot (sa caractérisation pour parler comme C. Fuchs). En effet, face à une forme graphique donnée, le récepteur, du moins le linguiste, peut s'interroger sur l'existence dans la

langue en question d'une autre forme identique mais dont la signification, l'histoire et d'autres propriétés encore sont différentes. Si cette forme existe, le récepteur sera confronté à une ambiguïté qui se situe à un niveau pour ainsi dire *métalinguistique*. S'agit-il du même mot ou bien de deux mots différents? Ce type d'ambiguïté, qui interfère avec les distinctions déjà faites, ne fait que compliquer le tableau. Pour un traitement automatique, nous proposerons de nous passer de cette distinction. Mais auparavant, voyons de plus près de quoi il est question.

Dans une homonymie, les significations sont distinctes et exclusives les unes des autres. Par contre, dans une polysémie, les significations sont, dans certains cas, tellement apparentées qu'elles peuvent se confondre. Il y a une continuité du sens. Un certain flou. D'ailleurs certains textes, certains énoncés exploitent cette caractéristique du langage (textes littéraires impressionnistes).

Dans le système de désambiguïstation que nous avons développé, nous avons inclus le phénomène de l'homonymie, en le fondant dans la polysémie.

4. Perception humaine de l'ambiguïté

Dans un processus de lecture, les sujets humains doivent souvent lever différents cas d'ambiguïté. De quelle manière le font-ils? Doivent-ils envisager au niveau de chaque mot ambigu toutes les possibilités que le contexte permet (ambiguïtés effectives), ou au contraire n'envisager qu'une seule possibilité à chaque mot et continuer jusqu'à la fin de l'énoncé? De telles questions sont en fait l'objet de multiples expériences psycholinguistiques. Plusieurs chercheurs ont tenté d'apporter une réponse à ce problème délicat de l'accès lexical (voir G. Hirst, 1987: 84; M. K. Tanenhaus et Leiman, 1979; D. Swinney et D. T. Hakes, 1976).

On s'accorde généralement à dire (voir G. Hirst, 1987: 84; M. K. Tanenhaus et Leiman, 1979) que les sujets humains trouvent la bonne interprétation d'un lexème ambigu

sans effort apparent et sans même prendre conscience²⁵ de l'existence d'une ambiguïté lexicale.

Le cas du discours comique (humour, histoires drôles etc.) n'obéit cependant pas, dans toutes ses manifestations, à cette règle. En effet, comme le mentionne G. Hirst (1987: 84), l'effet comique peut résider précisément dans la perception de l'ambiguïté. Différents cas se présentent: Dans certains, le comique est provoqué quand l'interprétant remarque d'abord le premier sens, ensuite le deuxième; dans d'autres cas, le comique se déclenche si l'interprétant remarque les deux sens en même temps. En dehors de cette exception, trois hypothèses différentes semblent s'affronter.

La première hypothèse consiste, pour reprendre les termes de C. Fuchs, en une "activation unique d'une seule signification". Elle est connue sous l'étiquette de "THE PRIOR CHOICE hypothesis". Citons, à la suite de G. Hirst (1987), D. Swinney et D. T. Hakes (1976: 683):

"Prior context biases the interpretive process before the ambiguity itself is encountered... perhaps allowing only a single reading ever to be accessed"

Dans une activation simple, une seule signification est activée en vertu du critère de la fréquence ou de la plausibilité. Les autres significations étant écartées, le traitement est rendu plus rapide. Néanmoins, il y a un retour en arrière si l'interprétation retenue s'avère être, en cours de chemin, une erreur. Cela arrive dans le cas où la séquence est une phrase labyrinthe²⁶. Comme les phrases labyrinthes sont relativement peu nombreuses, on peut estimer qu'en fin de compte, ce traitement global de l'ambiguïté par le sujet humain est plus économique.

²⁵ G. Hirst (1987: 84, note) rapporte l'expérience de Mohanty selon laquelle la présentation d'une phrase lexicalement ou structurellement ambiguë à un sujet provoque chez ce dernier une augmentation légère mais significative du battement de coeur.

²⁶ En anglais: garden path.

Pour l'homme, le contexte, l'**influence sémantique** (semantic priming) et l'**activation propagée** (spreading activation) sont autant de ressources qui contribuent à activer une seule signification.

Dans la deuxième hypothèse, on considère que le sujet humain opère "une **activation multiple** de toutes les significations" (voir C. Fuchs, 1996: 69). Cette hypothèse est connue comme étant *the ALL-READINGS hypothesis*. Reprenons la citation donnée par G. Hirst (1987) des mêmes auteurs (D. Swinney et D. T. Hakes (1976: 683)):

"Prior context influences the interpretation of an ambiguity only after all readings are accessed; it aids in the selection of an appropriate reading to retain from those accessed, but does not affect the access process itself"

L'activation multiple implique un traitement en deux phases de l'ambiguïté. Dans une première phase, le récepteur humain prend en considération (*active*) toutes les interprétations possibles. Cette opération est estimée à moins de 200 millisecondes (voir C. Fuchs 1996; G. Hirst 1987); cette extrême brièveté explique la difficulté de toute introspection (voir M. K. Tanenhaus et Leiman, 1979: 427). Dans une seconde phase, une seule signification est retenue aux dépens des autres, jugées inadéquates.

Une troisième hypothèse est dite *The ORDERED SEARCH hypothesis*: celle-ci est expliquée par Hogboam et Perfetti (1975: 66) comme suit:

"The order of search [of senses of an ambiguous word] is determined by frequency of usage, the most frequent being first. The search is self-terminating, so as soon as an acceptable match occurs, no [other] entries will be checked"

On devra faire le rapprochement, mentionné également par C. Fuchs, entre le traitement de l'ambiguïté par le sujet humain et celui opéré par la machine. Pour anticiper, disons qu'une implémentation informatique se préoccupant de l'économie de l'effort trouverait son compte plutôt dans une imitation de l'activation simple. On peut simuler la première hypothèse. Le traitement se fera dans ce cas mot par mot, **séquentiellement**. Pour les besoins d'une désambiguïsation automatique, on peut tirer profit des mêmes concepts: l'activation

propagée (le processus automatique devra reprendre le schéma d'interprétation observé chez le sujet humain).

5. Perception automatique de l'ambiguïté

Un ordinateur est une machine sophistiquée qui a la capacité de manipuler des symboles et de faire des calculs. Quand il est encore dépourvu de programmes adéquats, il est capable de faire très peu de choses. Son système d'exploitation est prévu pour gérer différentes activités et non pour analyser des textes et encore moins les comprendre.

Un programme de lecture linéaire est relativement facile à concevoir, mais à ce stade, il est encore question d'une segmentation simplifiée de la chaîne linéaire qu'est le texte. En faisant abstraction des embûches ou des incohérences rencontrées à ce stade, on peut postuler que l'ordinateur, exécutant le programme élémentaire qu'on vient de lui incorporer, va reconnaître des formes distinctes graphiquement. En supposant qu'on l'enrichisse d'un dictionnaire répertoriant toutes les formes de la langue française et respectivement toutes les acceptions de ces unités, notre ordinateur pourra maintenant, grâce à un module *ad hoc* de comparaison, reconnaître à chaque fois que le cas se présente une forme ambiguë, c'est-à-dire une forme pourvue de plus d'une signification. Nous venons de schématiser à gros traits les étapes de la perception que l'ordinateur peut se faire d'une forme ambiguë. C'est évidemment d'ambiguïté virtuelle qu'il s'agit.

On répliquera qu'une suite possible dans le programme peut l'amener à lever cette ambiguïté en vérifiant si dans le contexte il existe quelque indice susceptible de l'aiguiller vers une seule signification, ou si au contraire, rien ne peut aider à la désambiguïsation et que, par conséquent, il doit conclure à l'existence d'une ambiguïté effective. Nous trouvons ce raisonnement juste et c'est pourquoi nous essaierons de le développer.

On notera, cependant, que l'ordinateur est désavantagé par rapport au sujet humain, dans la mesure où une analyse linguistique de la phrase reste une tâche difficile à implanter. Le sujet humain, pourvu d'une faculté phénoménale d'analyser et de faire des

supputations très complexes, a, en plus, la latitude de recourir, dans un mouvement tantôt progressif tantôt régressif, à un contexte plus large que la phrase.

Le sujet humain élimine les ambiguïtés virtuelles très rapidement (il est même rare qu'on assiste au contraire) et ne conclut à l'ambiguïté effective que quand c'est vraiment le cas, ou quand il n'arrive pas à effectuer l'analyse linguistique appropriée. Cette opération, même dans sa version éprouvante, reste encore plus efficace et plus rapide que tout programme artificiel. Il existe plusieurs raisons à cela: l'ordinateur pèche autant par sa rigidité que par sa puissance (Cf. C. Fuchs, 1996). Ceci dit, nous ne devrions pas occulter le fait que le langage, autant par sa complexité que par sa mouvance, échappe encore à une formalisation à la fois globale et pointue.

Dans la section qui suit, nous reviendrons sur différentes stratégies de désambiguïstation automatique adoptées par les chercheurs, et nous verrons que les enjeux à la fois conceptuels²⁷ et matériels²⁸ hypothèquent largement la perception qu'a l'ordinateur de l'ambiguïté.

Mais avant d'aller plus loin, considérons une dernière remarque sur un autre aspect de l'ambiguïté: l'ordinateur qui interprète les formes ne sait pas de prime abord s'il a affaire à un mot homonyme d'un autre, ou plutôt à un mot polysémique. A la charge donc du concepteur soit de bâtir une stratégie pour y remédier, soit de considérer que ce serait une tâche trop lourde et peu productive, et qu'elle mérite plutôt qu'on l'évite. Ce sera précisément notre façon de faire.

²⁷ La langue étant loin d'être un objet univoque, les chercheurs ont imaginé, pour en avoir le contrôle, des solutions parfois complexes (par exemple le recours aux graphes conceptuels). Mais l'implémentation de ces solutions n'allait pas de soi. Quand on parle de perception de l'ambiguïté par l'ordinateur, il faudrait par conséquent avoir à l'esprit à la fois la conception linguistique, la stratégie informatique (l'algorithme), la compatibilité de l'une avec l'autre, et enfin les possibilités techniques de l'ordinateur.

²⁸ La technologie des mono-processeurs qui s'oppose à la technologie des processeurs parallèles, permet un traitement en série.

D. STRATÉGIES DE DÉSAMBIGUÏSATION AUTOMATIQUE

Pour bien situer notre propre stratégie de désambiguïsation, il nous faudrait rendre compte d'autres travaux sur la question. En fait, chacune des stratégies concrétise à sa manière le dialogue entamé ces dernières années entre deux domaines apparemment distincts (l'un tourné vers la théorie et l'autre préoccupé par la technologie²⁹): la linguistique et l'informatique. Il serait en tout cas intéressant de faire le lien entre la nature des rapports établis entre les deux disciplines et le type de stratégie adopté pour résoudre le problème particulier qu'est celui de la levée automatique de l'ambiguïté lexicale.

Dans ce qui suit, nous tâcherons de donner, après une brève réflexion terminologique, un aperçu de quelques travaux relatifs à la question de la désambiguïsation lexicale automatique, en particulier ceux qui intègrent la notion de cohésion lexicale.

1. Linguistique informatique

Parce que le rapprochement de la linguistique et de l'informatique s'est opéré très récemment, et parce que ce rapprochement a pris de multiples formes, la discipline qui en est le résultat a reçu plusieurs étiquettes, chacune renvoyant à une pratique différente, du moins, à des postulats théoriques différents. C'est ainsi qu'on parle de "linguistique automatique", de "linguistique computationnelle", de "traitement automatique des langues", de "linguistique informatique" et d'"informatique linguistique". Dans ce qui suit, nous revenons sur cette question de terminologie, et nous aborderons par la suite les méthodes utilisées dans ce domaine.

La dénomination *informatique linguistique* renvoie, selon l'analyse de F. Rastier (1994) à une utilisation d'une "linguistique restreinte", essentiellement la morphosyntaxe. Par contre, le terme *linguistique informatique* renvoie à une pratique dans laquelle l'informatique n'est qu'un outil mis au service de la linguistique.

²⁹ Cf Rastier (1994)

Les méthodes de travail varient en fonction des applications visées et des positions théoriques adoptées. C'est ainsi que certaines recherches s'inscrivent dans ce qui est appelé *les industries de la langue* ("génie linguistique") (en anglais: language industries), alors que d'autres, tournées vers une vision plus radicale, se réservent l'étiquette de *recherche fondamentale*.

Dans le premier cas, l'objectif est souvent d'obtenir une application qui fonctionne et accomplit la tâche visée sans prétendre à une cohérence théorique; tout procédé permettant de contourner les problèmes devient légitime.

Dans le second cas, l'objectif est d'arriver à des applications qui seraient le fruit d'une démarche reposant elle-même sur une théorie linguistique homogène. Dans ce cadre, les démarches informatiques elles-mêmes sont choisies en fonction de cette ambition. Comme, à l'heure actuelle, il n'y a pas de modèle général de la langue qui soit à toute épreuve, et, comme, en informatique, la plus sophistiquée parmi les techniques que représente l'intelligence artificielle achoppe sur des problèmes de décidabilité³⁰, ce type de recherche ne répond pas encore aux attentes des chercheurs.

2. Exploitation de la cohésion lexicale

a) De l'insuffisance de la syntaxe et de la sémantique

Toute analyse dans le domaine du traitement automatique se justifie par les objectifs mêmes de ce traitement. En effet, certaines applications, telles que la concordance ou l'exploration du lexique (du moins pour les mots simples), ne font pas appel à une analyse syntaxique. En revanche, dès qu'on projette de tirer parti des relations des unités entre elles, on est tenté de faire une analyse syntaxique.

³⁰ Nous faisons allusion à la difficulté qu'ont par exemple les systèmes experts de rendre une décision fiable quand ils font face à une explosion combinatoire de nœuds à visiter. La complexité des heuristiques utilisées ne fait que souligner le problème de la décidabilité.

La syntaxe est une affaire de structure. *Structure* est entendu ici au sens le plus rigoureux. Les mots entretiennent entre eux des rapports prévisibles et calculables; il y a différents niveaux de relations; la position de chaque mot compte; les mots jouent un rôle dans l'ensemble où ils se tiennent; il y a enfin une hiérarchie décelable.

Ce dernier point mérite d'être souligné. En effet, même si la phrase se présente d'une manière linéaire, l'interprétation et l'évaluation de chaque mot nécessitent que soit pris en compte tout le réseau de relations qui sous-tendent l'unité de la phrase.

Comme le mentionne C. Fuchs, la syntaxe a ses propres problèmes, mais pour le traitement automatique, d'autres problèmes spécifiques s'ajoutent.

b) L'Intelligence Artificielle: "promesses et réalités"

Le problème de l'ambiguïté lexicale s'inscrit dans l'ensemble des problèmes liés à la compréhension du langage naturel par la machine.

On le retrouve au centre des interrogations que pose la linguistique informatique, en particulier ce que Y. A. Wilks (1988) appelle la sémantique computationnelle. Selon cet auteur, cette tendance se distingue de trois autres domaines qui se préoccupent de la compréhension automatique du langage, à savoir, l'analyse syntaxique, la sémantique logique, et les systèmes experts.

La sémantique computationnelle s'oppose à l'idée que la sémantique logique et l'analyse syntaxique soient nécessaires ni même suffisantes pour la compréhension automatique du langage. L'auteur rejette l'idée qui prévaudrait chez les défenseurs des systèmes experts selon laquelle une base de connaissances adéquate suffirait pour résoudre le problème du sens, car selon cette position, la base ne contiendrait que les significations appropriées. L'auteur rejette l'idée et conclut que le *savoir linguistique* et le *monde* sont inséparables.

Historiquement, il faut remonter jusqu'aux travaux de R. C. Schank (1975) et de Y. A. Wilks (1975). Plus récemment, la sémantique computationnelle s'est occupée du problème du sens. Dans ce qui suit, nous rendons compte de quelques travaux allant dans cette direction.

(1) Les mots experts

Le système développé par G. Adriaens et S. L. Small (1980) s'intitule *Word Expert Parser* (WEP). Il peut être considéré comme un des premiers systèmes d'intelligence artificielle qui a tenté de résoudre le problème de l'ambiguïté lexicale. C'est un programme qui s'inscrit dans le cadre des travaux sur la compréhension automatique du langage. Il se veut un modèle du processus de compréhension, au delà de son intérêt comme système basé sur la démarche de l'intelligence artificielle.

Les auteurs du programme partent de l'observation que les mots se caractérisent par une richesse de contenu (idiosyncratique) et regrettent que les conceptions de systèmes de désambiguïsation n'exploitent pas cet aspect. Ils déplorent que ces conceptions privilégient des approches qui manipulent des règles générales sur le langage (syntaxiques et / ou sémantiques), les mots n'étant que des occurrences (tokens) qui participent à la compréhension en vertu de leur inclusion dans les règles.

Considérant que cette démarche est incompatible avec le problème de la sélection de sens, les auteurs sont conduits à la conception d'un modèle radicalement différent.

Les auteurs rejettent la stratégie, généralement répandue, qui consiste à prévoir un programme général qui contrôlerait des composants séparés (morphologique, syntaxique, sémantique) et qui se présentent sous forme de règles structurelles statiques. Au lieu de cela, les auteurs avancent l'idée que les mots eux-mêmes doivent être vus comme des agents actifs, ou comme des *mots experts*.

Chaque mot peut déclencher un processus qui lui permet d'accéder à un ensemble de connaissances mémorisées: Les mots du contexte immédiat, les concepts traités, des

connaissances sur le processus lui-même, sur le discours, et enfin sur le monde. Toutes ces connaissances sont susceptibles d'aider dans une opération de désambiguïsation.

Une autre façon de définir un *mot expert* est de le considérer comme un réseau sémantique dont les nœuds sont des **questions** sur le contexte et dont les arcs correspondent aux réponses possibles à ces questions. Chaque feuille de l'arbre représentant une signification particulière au contexte examiné, telle qu'elle apparaît une fois le réseau examiné.

La désambiguïsation lexicale se fait en mettant en œuvre l'ensemble de ces questions. Parmi celles-ci, la question dite *view question* permet de vérifier si le concept rattaché au mot voisin est proche de celui rattaché au mot traité. Deux autres questions nous intéressent davantage dans cette stratégie. La première, dite *Literal question*, a pour rôle de vérifier si dans le voisinage du mot à traiter il n'y a pas une particule qui va avec ce mot. Pour un mot comme *throw*, les particules possibles sont: *away, up, in, out*. La seconde vérifie si le contexte ne contient pas une expression idiomatique comme par exemple *throw in the towel*, laquelle conduirait à une signification particulière du mot *throw* (give up).

Nous avons présenté ces deux concepts parce que dans notre propre stratégie, nous recourons à des concepts très proches. Néanmoins, dans notre vision, les deux concepts sont fondus dans une seule notion: nous ne faisons pas de distinction entre le littéral et l'idiome. Selon nous, toute expression ou collocation (y compris les éléments grammaticaux qui accompagnent le verbe) sera prise comme un littéral. Son apparition en tant que telle dans le contexte aura un rôle informatif.

Notre utilisation des expressions idiomatiques (qui, dans notre conception, font partie d'un concept plus général: la collocation) différera du modèle de G. Adriaens et S. L. Small. Nous avons regretté que dans le document consulté, nous n'ayons pas trouvé de plus amples informations sur la façon dont ces idiomes sont utilisés (selon quel principe ils sont retenus? quels éléments sont retenus? et quelle quantité?). Néanmoins, nous avons trouvé très intéressant le principe général qui sous-tend cette proposition. En effet, partant de leur réflexion sur les expressions idiomatiques, les auteurs en arrivent à considérer que tout fragment du discours est plus ou moins idiomatique. Cette idée serait à rapprocher selon les

auteurs de leur vision générale du langage selon laquelle les mots reflètent une richesse et un haut niveau d'idiosyncratie³¹. En reprenant cette pensée à notre compte, nous dirons que les collocations peuvent être suffisamment singulières pour qu'elles puissent être inductrices de significations particulières. Nous ne suivons pas en revanche les auteurs dans les conséquences qu'ils tirent de leur réflexion: selon G. Hirst (1987: 83, voir en particulier la note 10), Small considère que le langage ne se laisse appréhender qu'à travers des opérations locales sur les mots et non à travers des généralisations et des règles globales.

G. Hirst (1987) reproche à cette méthode le fait que les *mots experts* sont trop compliqués. Un mot comme *throw* aurait une description de six pages et pourrait en avoir dix fois plus.

(2) Les mots Polaroid

G. Hirst (1988: 73) a développé un système de désambiguïsation lexicale qui tient compte de travaux de psycholinguistique, notamment ceux de D. Swinney (1976), de Seidenberg, Tonenhaus, Leiman et Bienkowski (1982).

Le système est formé de plusieurs composantes: un analyseur syntaxique: *Paragram*, un interpréteur sémantique appelé *Absity*, un désambiguïsateur appelé *Semantic Enquiry Desk* et un système de représentation des connaissances et d'inférence connu sous le nom de *Frail*.

Ces différentes composantes fonctionnent en coopération pour le même objectif: la désambiguïsation lexicale. L'analyseur syntaxique construit une structure de la phrase. L'interpréteur sémantique en fait un "objet sémantique". Comme la stratégie générale prévoit

³¹ "When we look at the words of a natural language (by skimming through a dictionary, or by introspection), we find that they are objects with a very rich and highly idiosyncratic content." Page 14. "One of the linguistic ideas behind WEP was that all fragments of language are more or less idiomatic [Bolinger, 1979]. The view is closely related to the stress on idiosyncrasy in language, disputing the sharpness of the idiom notion." Page 30.

de traiter le mot ambigu aussitôt qu'il est rencontré³², le système commence par une représentation provisoire du mot. Celle-ci va céder le pas à l'attribution de traits réels au fur et à mesure que l'analyse avance. G. Hirst utilise la métaphore de la photo Polaroid pour expliquer le processus. Ce qui est pertinent dans cette image, c'est que le mot acquiert les propriétés qui le définissent de manière progressive, à la manière d'une photo Polaroid. L'analyse est considérée comme complète à la fin de la phrase. Le système peut manipuler le mot (objet sémantique) à différentes étapes intermédiaires.

Absity est un marqueur sémantique qui exploite le principe de la *prégnance sémantique* (semantic priming). En cas d'ambiguïté lexicale, *Absity* parcourt un réseau sémantique (composé de nœuds correspondant chacun à un concept) s'il rencontre un concept proche du mot traité, il active cette connexion.

Comme le remarque G. Hirst, la navigation dans le réseau sémantique est très coûteuse à cause de la trop grande quantité de nœuds. Le recours à des processeurs parallèles est envisagé. Les *mots Polaroid* se présentent comme des procédures séparées. Cela n'est pas sans rappeler les *mots experts* de Small.

Chaque mot expert contient un paquet d'informations lexicales susceptibles de permettre de choisir une signification. Les *mots Polaroid* communiquent entre eux selon des règles prédéterminées.

La désambiguïsation lexicale se fait sur la base des connexions établies entre les nœuds correspondant aux mots voisins. La connexion se fait (un chemin est créé) si la description du mot-source (frame) croise la description du mot voisin. Il y a croisement si ces descriptions respectives des mots voisins ont un composant en commun.

Des règles sont définies pour empêcher la prolifération des croisements malheureux. Le mécanisme du *mot Polaroid* intervient à la suite de l'intervention du marqueur

³² D'autres démarches proposent de surseoir à la désambiguïsation et de s'en occuper à la fin de l'analyse de la phrase. Cf. Dahlgren, McDowell, et Stabler (1989), cités par Susan W. McRoy (1992).

sémantique. Si ce dernier n'a pas rencontré de difficulté particulière, le mécanisme *mot expert* n'a rien à faire d'autre.

(3) Les préférences sémantiques

Le système développé par Y. A. Wilks (voir G. Hirst 1987) est fondé sur le concept des restrictions de sélection. Il ne s'agit cependant pas de restrictions absolues, mais plutôt de préférences. Ainsi, un verbe comme *drink* a comme préférence un sujet animé mais un sujet inanimé peut être accepté. Ce qui permet la prise en charge d'une phrase comme:

- *My car drinks gasoline.*

Les restrictions de sélection concernent, en fait, toutes les catégories. Elles sont d'ordre syntaxique, mais également sémantique. Ainsi, un mot comme *big* est étiqueté comme préférant un objet physique.

Cette stratégie semble avoir des limites. Selon G. Hirst (1987) qui cite B. Boguraev (1979), la méthode de Y. A. Wilks est inadéquate pour traiter les cas d'ambiguïté lexicale qui nécessitent un contexte global, une association avec les concepts véhiculés dans le contexte immédiat, ceux qui nécessitent un calcul inférentiel et enfin les cas de polysémie.

Un système qui utilise une démarche similaire est celui de B. Boguraev. Le système de désambiguïsation de B. Boguraev est basé sur l'approche de Y. A. Wilks, mais en utilisant des jugements sémantiques, au lieu des préférences sémantiques. Son objectif était la désambiguïsation lexicale et structurelle. Son système est basé sur un ATN (Augmented Transition Network).

(4) Des sources d'information multiples

La démarche proposée par Susan W. McRoy se présente comme une position intermédiaire entre deux démarches opposées: la première est fondée sur le lexique (*word-based*) et traite les mots à l'entrée en les considérant comme des chaînes de caractères et en observant leurs relations avec d'autres. La seconde approche est fondée sur la représentation

des connaissances et essaie d'établir une sorte d'*interlingua* pour représenter le contenu du texte. Alors que la première souffre du double problème de la multiplicité des significations pour un mot donné et l'unicité de signification pour des mots différents, la seconde pêche par son incapacité d'être opérationnelle pour le traitement concret de textes; la difficulté étant de construire une base de connaissances suffisante pour un traitement de grande envergure.

L'auteur préconise de travailler sur la base des significations au lieu des chaînes de caractères. Cela suppose naturellement que les mots soient annotés de leurs significations respectives. Tout en reconnaissant que le contexte sémantique et les restrictions de sélection sont un bon moyen pour lever l'ambiguïté, l'auteur considère que ces moyens ne sont ni suffisamment fiables, ni assez rapidement accessibles, pour être utilisés à eux seuls. Il faudrait par conséquent multiplier les ressources. En conséquence de quoi, l'auteur prévoit d'en utiliser toute une panoplie, dont les marques syntaxiques (*syntactic tags*), la fréquence des mots, les collocations, le contexte sémantique, les restrictions de sélection et les outils syntaxiques (*syntactic cues*).

Mais parmi ces moyens, la collocation (au sens d'expression ou d'association de mots) est considérée, au côté des marques syntaxiques et des propriétés morphologiques, comme la plus importante source d'information qu'on puisse mettre au service de la désambiguïsation lexicale. Cependant, cette source d'information n'est pas à prendre d'une manière absolue. Ainsi, si la collocation *Wait on* induit le sens *servir* comme dans *Mary waited on John*, il reste qu'un autre critère entre en jeu d'une manière plus forte. Il s'agit des préférences lexicales et des supputations de l'interprétant (*role-expected expectation*). Ici, on accorde une préférence au fait que le complément est un inanimé comme dans *Mary waited on the steps*. L'auteur en conclut que l'interprétant devrait opérer une évaluation de l'information émanant d'une collocation en la comparant à d'autres sources, plutôt que de l'utiliser comme filtre sémantique indépendant du reste, comme le suggère par exemple G. Hirst (1987). Sur cette question, notre opinion est qu'il est possible de retenir dans la collocation autant d'éléments nécessaires à la désambiguïsation, de sorte que la règle des préférences lexicales n'est plus nécessaire à évaluer. En l'occurrence, la collocation retenue devrait prendre la forme: *wait on steps*.

Cette démarche, qui combine différentes sources d'informations, a donné de très bons résultats, selon les estimations de l'auteur.

c) Pour un contexte littéral

Nous comptons, pour notre part, explorer la notion de collocation. En nous attachant fondamentalement à la deuxième acception du terme collocation, nous avons projeté d'exploiter la propriété d'induction de sens reconnaissable dans les collocations.

La collocation sera conçue comme un phénomène qui s'intègre au texte, et qui, ce faisant, contribue à sa cohésion. Imprégné de ce dernier constat, nous avons accordé de l'importance aux différentes combinaisons qu'une collocation peut subir dans le discours. Y a-t-il interruption? Ou au contraire une contiguïté? De telles questions se posent et nous rappellent que si la collocation au sens de l'expression est plus une question de cohésion de lexique, une sorte de sédimentation lexicale, elle ne se manifeste pas moins sous différentes couleurs dans le discours. Ce sont ces couleurs qui font l'objet de notre observation, et ces nuances seront déposées dans une base de données appropriées. Nous projetons une alimentation de la base par l'expert humain, qui saura quoi retenir dans une collocation et quel type d'agencement des mots est plus productif. L'expert tiendra compte des règles établies dans le système (des règles simples, il faut le dire) et transmettra son savoir dans la base. Pas sous forme de règles formelles, mais comme des tranches de discours, relevées dans leur littéralité.

Le propre de notre démarche c'est aussi notre intention de travailler avec des collocations attestées. Nous planifions d'extraire les collocations de manière quasi automatique à partir de textes socialement reconnus.

La base sera composée d'une part des mots formes, d'autre part des lemmes et des catégories. A cela, nous associerons des significations et des collocations. D'une certaine manière, cette façon de faire rappellera les mots experts vus plus haut. Mais la comparaison ne va pas loin, car dans notre cas, l'expertise n'est pas traduite en règles. Elle est transférée à

travers le littéral. Et du coup, elle est sûrement appauvrie. Ce qui en reste est cependant d'une grande force, sans oublier que la démarche procède d'une manière quasi naturelle.

E. CONCLUSION

Nous retenons de ce qui précède les idées suivantes:

D'abord, que la cohésion lexicale se manifeste comme une répétition ou comme une collocation. Ce dernier concept renvoie soit à une association de mots reliés sémantiquement et apparaissant dans le même environnement (c'est le phénomène de l'appartenance à un même champ notionnel), soit à une rencontre marquée dans le système linguistique et se manifestant de façon variée dans le discours (C'est le phénomène des tournures lexicalisées). Nous avons soutenu que l'un et l'autre apportent à son niveau une part de cohésion lexicale.

Après avoir exposé les différentes réalisations de l'ambiguïté, nous avons aussi montré quelques différences entre le sujet humain et la machine face à un énoncé ambigu. Nous avons également exposé quelques-unes des démarches suivies par d'autres chercheurs. Enfin, nous avons évoqué en termes encore abstraits, notre propre stratégie. Le chapitre suivant nous permettra de présenter une application qui tient compte de ces considérations théoriques.

CHAPITRE DEUXIÈME
COHÉSION LEXICALE ET
DÉSAMBIGUÏSATION AUTOMATIQUE:
APPLICATION

2. COHÉSION LEXICALE ET DÉSAMBIGUÏSTATION AUTOMATIQUE: APPLICATION

A. INTRODUCTION

La démarche de désambiguïstation lexicale automatique que nous proposons ici est somme toute assez simple et facile à comprendre. Considérant qu'un mot peut avoir plusieurs significations, l'idée est d'associer à chacune des significations un ensemble de collocations qui sont susceptibles de l'induire. Le programme de désambiguïstation aura pour tâche de vérifier si l'une ou l'autre de ces collocations est attestée dans le contexte du mot ambigu. Si c'est le cas, il conclura que la signification du mot en question sera celle à laquelle est associée la collocation examinée. Dans ce qui suit, nous expliciterons en détail ce propos laconique.

B. STRATÉGIE GÉNÉRALE DE DÉSAMBIGUÏSTATION

Considérons l'extrait suivant:

Le premier titre traite de la jouissance et de l'exercice des droits civils et il énonce les principes généraux en la matière.³³

Dans ce court extrait, constitué d'une seule phrase, le mot *droits* ne semble poser aucun problème d'interprétation (au sens dénotatif) à un lecteur pourvu d'une connaissance moyenne de la langue française. On pourrait épiloguer longuement sur la démarche cognitive à la base de cette facilité d'interprétation. Le lecteur envisage-t-il, par exemple, toutes les significations potentielles avant de se fixer sur celle qui, tout compte fait, sera la bonne? Cette question est sûrement intéressante à explorer, mais aux yeux de l'ordinateur, à moins d'indications contraires, toutes les possibilités sont *a priori* à examiner.

Le mot *droits* qu'on peut présumer comme déjà répertorié dans un dictionnaire électronique sera considéré par la machine comme possédant plus d'une signification, en l'occurrence, selon un résumé libre de la définition donnée dans le Petit Robert:

³³ Code Civil du Québec p. 2. Le mot *droits* a été souligné par nous.

Signification 1: Ce qui est permis en conformité à une règle morale, sociale, précise. (Mots associés: possibilité, pouvoir, qualité, autorisation, permission, faculté, habileté, prérogative, privilège)

Signification 2: Somme d'argent, redevance qu'une personne, une collectivité est en mesure d'exiger de quelqu'un (Mots associés: accises, contribution, *vieux*: imposition, impôt, redevance, taxe...) Exemple: *Acquitter un droit*.

Signification 3: Ensemble de règles qui régissent les rapports des hommes entre eux. (Mots associés: légalité, légitimité, justice, morale, juridique).

Laquelle de ces trois significations est attestée dans notre extrait? La question est surtout de savoir sur quoi le programme devra fonder son *jugement* pour opter pour la bonne signification, i.e. la signification effective dans le corpus (nommément: la signification numéro 3).

Partant de l'idée qu'une collocation donnée est souvent inductrice d'une signification donnée à l'exclusion d'autres, le programme pourrait, si seulement il disposait, pour chacune des significations de chacun des mots de la langue, des collocations les plus pertinentes (c'est-à-dire les collocations dont l'apparition dans le contexte permet d'attester pour le mot en question une signification particulière), balayer tout l'environnement du mot ambigu dans le texte dans l'espoir d'en trouver une. Plus concrètement, si on a préalablement associé à la signification numéro 3 la collocation "exercice des", et qu'une exploration du contexte gauche du mot *droits* nous permet de constater que la collocation y est présente, le programme pourrait conclure que c'est la signification numéro 3 qui est attestée. Dans le cas présent, nous avons une seconde chance d'aboutir, le mot *droits* est accompagné à sa droite du mot *civils*, sans compter le mot *jouissance* qui se trouve à sa gauche.

Convaincu de l'efficacité d'un tel raisonnement, nous avons planifié l'utilisation des collocations selon les indications suivantes:

-
- Un dictionnaire contiendrait les mots de la langue, et pour chacun de ces mots, la catégorie grammaticale et le lemme.
 - Associer à chaque lemme toutes les significations possibles.
 - Associer à chaque signification un certain nombre de collocations.

Le Tableau 2 illustre ces distributions, à travers le cas du mot *droit*:

Mot-forme	Lemme-catégorie	Signification	Collocation
"droit"	DROIT-NOM	1- Ce qui est permis en conformité à une règle morale, sociale / précise. Possibilité, pouvoir, qualité, autorisation, permission. Faculté, habileté, prérogative, privilège.	D: de propriétaire D: de jour D: accordé D: mettre en circulation D: de circuler D: de propriété G: cession du G: exercice de D: de reprise D: de obtenir G: suspension du G: suspendre D: de personne D: et obligation G: accorder D: suspendu
		2- Somme d'argent, redevance qu'une personne, une collectivité est en mesure d'exiger de quelqu'un. → accise, contribution, vx: imposition, impôt, redevance, taxe. Exemple: acquitter un droit.	D: fixé G: payer G: acquitter G: paiement des
		3- Ensemble de règles qui régissent les rapports des hommes entre eux. → légalité, légitimité, justice, morale. → juridique.	D: commun D: romain D: écrit G: qui de
	DROIT-ADJECTIF	Qui est sans déviation, / Dont la direction est constante... Tige droite, avoir le corps droit → raide. Ligne, voie droite → direct, rectiligne	G: ligne G: voie
		Qui est du côté opposé à celui du cœur de l'observateur. - Opposé à gauche → dextre La main droite.	G: extérieur

Tableau 2 Structure globale de la base de données.

Légende du Tableau 2:

D : Droite
G : Gauche

Chacune des colonnes du tableau ci-dessus appelle un commentaire. Commençons par la première. Celle-ci représente dans notre stratégie un dictionnaire des mots de la langue, dictionnaire qui devrait tenir compte des règles suivantes:

- 1- les différentes formes de la langue constituent les entrées du dictionnaire.
- 2- Une forme ne constitue qu'une seule entrée. Par conséquent, face à deux mots homonymes, notre choix est de les représenter dans une seule entrée, et de représenter les distinctions de significations de la même manière que pour le cas de la polysémie.

La deuxième colonne contient le duo lemme-catégorie. En plaçant dans le dictionnaire la catégorie et le lemme, nous nous assurons une économie dans la gestion des données. Toutes les informations qui vont suivre vont être reliées au lemme et non à la forme. Nous nous donnons en même temps le moyen de savoir rapidement si la forme est ambiguë du point de vue aussi bien de la catégorie que du lemme.

Dans la troisième colonne, nous retrouvons la ou les différentes significations du mot-lemme. Nous savons que la détermination du réseau des rapports de ces significations peut être assez complexe. Néanmoins, parce que notre démarche est à un stade expérimental, nous avons opté provisoirement pour la représentation des distinctions les plus grossières.

La dernière colonne représente la liste des collocations relatives à une signification donnée. Le choix de ces collocations se fait en fonction de leur force de désambiguïsation. On doit observer des critères de sélection, mais également une certaine syntaxe pour représenter ces collocations dans la base.

Une fois ces dictionnaires constitués et les relations établies, l'algorithme de désambiguïsation consisterait à faire une exploration de l'environnement textuel du mot

ambigu dans l'espoir de trouver une des collocations associées à une signification donnée. Le programme devrait examiner —tant et aussi longtemps que la comparaison n'a pas abouti à une concordance des deux données comparées— les collocations l'une après l'autre et ce pour chacune des significations. Il serait d'ailleurs judicieux de placer aussi bien les significations que les collocations selon l'ordre le plus rentable. Un ordre possible est celui de la fréquence d'usage pour ce qui concerne les significations et celui de la force de désambiguïsation pour les collocations. Par ailleurs, il serait économique, dans la représentation des collocations, de substituer les lemmes aux formes, ce qui permettrait de comparer les lemmes de la collocation à ceux du contexte du mot ambigu. Naturellement, cela suppose que chacun des mots du texte soit annoté du lemme approprié. Par exemple, pour le contexte suivant:

" la demande de la Société ou d'un agent de la paix, la preuve de cette immatriculation; 4 le même droit est accordé aux étudiants, aux coopérants ou aux stagiaires du Québec qui séjournent au lieu du domicile de cet étudiant, coopérant ou stagiaire"

on peut considérer que la collocation "droit accordé" est inductrice de la signification "pouvoir, privilège" du mot *droit* (plutôt que "*taxes, impôts*) . Il est avantageux de retenir comme items de cette collocation le lemme accorder, la comparaison devant se faire entre ce lemme et ceux associés respectivement aux mots du contexte. Ainsi, si l'on rencontre différentes formes de l'item *accorder* comme dans les contextes suivants, le système en tirera bénéfice d'une manière égale.

- le même droit est accordé aux étudiants

- les mêmes droits sont accordés aux étudiants

- le droit qu'on accordera aux étudiants

Le processus global tel qu'il a été décrit pourrait être vérifié manuellement, d'autant plus que l'hypothèse est théoriquement séduisante et défendable. Nous avons jugé, cependant, que la réalisation du système planifié aurait au moins l'avantage de mettre à l'épreuve nos hypothèses.

Nous avons pris soin de réaliser³⁴ successivement trois programmes qui, pris ensemble, constituent le système de désambiguïsation lexicale automatique.

Le premier, appelé MARQUEUR, est un outil de marquage des mots. Fondamentalement, il a été prévu pour assigner aux mots d'un texte donné les lemmes et catégories grammaticales respectifs de ceux-ci. Accessoirement, il répond à d'autres besoins. Ce logiciel fonctionne d'une manière semi-automatique.

Le deuxième, du nom de COLLOCATEUR, sert à extraire³⁵ d'un texte donné toutes les collocations répondant aux critères de l'utilisateur. Si le texte est déjà marqué, COLLOCATEUR fonctionne d'une manière complètement automatique.

Le dernier programme, appelé LOCALEX (il localise des lexèmes), est le programme-test par excellence de notre hypothèse. Il localise les mots sélectionnés par l'utilisateur, s'astreignant, chemin faisant, à montrer si l'occurrence trouvée répond ou non aux traits grammaticaux et sémantiques désignés par l'utilisateur.

Dans ce qui suit, nous reviendrons sur chacun des programmes pour en montrer les principes sous-jacents.

1. Le programme MARQUEUR

Notre besoin primordial pour réaliser un système de désambiguïsation était d'une part de développer un outil qui permettrait une lemmatisation de tous les mots du texte, et d'autre part la confection d'un dictionnaire des formes et d'un dictionnaire des lemmes. Nous

³⁴ Les programmes MARQUEUR et LOCALEX ont été conçus et réalisés en collaboration avec Michael Mepham. Le programme COLLOCATEUR a été initié par Michael Mepham et a été réalisé en collaboration avec Michael Mepham, Mario Toupin et Conrad Bourdon du C.T.I. (actuellement S.I.T.).

³⁵ Cf. F. Smadja (1993) qui présente XTRACT, un logiciel d'extraction automatique de collocations.

avons besoin également que tous les fichiers en question obéissent à une structure souple et maniable³⁶.

La sous-section suivante nous permettra d'expliciter les principes sous-jacents au logiciel MARQUEUR. Pour une explication plus pointue de son fonctionnement et une présentation de son interface, on peut se reporter au guide placé en annexe. Nous allons, en fait, examiner deux niveaux d'utilisation de MARQUEUR: le marquage au bénéfice de la désambiguïisation, et le marquage pour d'autres besoins³⁷.

a) Marqueur au service de la désambiguïisation

Nous avons mentionné précédemment que le système de désambiguïisation que nous proposons fonctionne sur des textes préalablement traités. Ce traitement, qui se fait sous MARQUEUR, consiste à associer à chaque unité du texte deux propriétés essentielles, chacune d'elle devant apporter une première couche d'informations sur le mot pris en tant que forme.

Il s'agit en premier lieu d'une sorte d'archi-lexème, un mot générique si l'on veut, appelé communément par les linguistes *lemme*. En second lieu, il s'agit de la catégorie grammaticale.

Le *lemme* est généralement défini comme étant la forme qui sert d'entrée dans un dictionnaire. Pour former un lemme, on peut utiliser les règles suivantes:

Un nom au pluriel se voit perdre la marque du nombre pour former un lemme.

Une forme conjuguée est ramenée à l'infinitif.

³⁶ Ce dernier point est une autre raison qui nous a encouragé à réaliser notre propre logiciel de lemmatisation, la première étant la non-disponibilité d'un tel programme dans le milieu universitaire où nous faisons notre recherche.

³⁷ Bien que cette extension du programme ne fasse pas partie des objectifs fondamentaux de notre thèse, nous avons cru utile d'apporter cette ouverture, gardant à l'esprit que cette latitude d'enrichir la base est un potentiel important, y compris pour une amélioration de la stratégie de désambiguïisation.

Un adjectif est ramené à la même forme au masculin singulier.

Pour le reste, on conserve la même forme pour constituer le lemme.

Le Tableau 3 illustre ces règles de conduite:

Mot-forme	Lemme + catégorie
bureau	BUREAU-NOM
bureaux	BUREAU-NOM
actuellement	ACTUELLEMENT-ADVERBE
grands	GRAND-ADJECTIF
partira	PARTIR-VERBE
par	PAR-PRÉPOSITION

Tableau 3 Règles à suivre pour former un lemme.

Au delà de ces consignes qui agissent sur l'aspect formel du lemme, se posent de vraies questions à portée théorique. (1) Une divergence de catégories grammaticales entraîne-t-elle la formation de deux lemmes différents? (2) Deux homonymes doivent-ils être représentés comme deux lemmes différents?

Commençons par la première question. En observant un dictionnaire comme le Petit Robert, on s'aperçoit qu'il utilise la catégorie grammaticale comme critère de démarcation de prime abord. Un mot comme *ferme* a une entrée en tant que nom et une autre en tant qu'adjectif ou adverbe. La catégorie grammaticale est donc, dans ce cas, un critère de premier plan pour déterminer les entrées.

Par contre, un dictionnaire de l'anglais comme le Cobuild English Dictionary adopte une attitude originale. Pour chaque forme, il n'y a qu'une seule entrée. Les subdivisions se font fondamentalement en fonction du sens, la catégorie n'étant utilisée comme ligne de démarcation que quand elle apporte une nuance de sens.

Il ressort de ces deux exemples que les dictionnaires n'adoptent pas tous la même attitude dans l'organisation des entrées³⁸. Dans notre cas, nous avons considéré que chaque

³⁸ Cf. J. DUBOIS et F. DUBOIS-CHARLIER (1990).

forme constitue une seule entrée. A chaque entrée correspondra un ou plusieurs lemmes. Le lemme est intimement lié à une catégorie grammaticale.

Quant à la catégorie, nous l'avons utilisée de manière à distinguer deux lemmes semblables en tous points sauf pour la catégorie. Quelques raisons nous ont conduit à ce choix: la première, c'est que MARQUEUR sert à opérer un premier niveau de désambiguïsation, le niveau morpho-syntaxique. Il était donc essentiel de distinguer deux lemmes s'ils ont deux catégories différentes. Une autre raison est d'ordre théorique.

Nous considérons en effet que les nuances de sens dépendantes de la catégorie sont suffisamment importantes pour mériter qu'on les retienne comme des propriétés discriminatoires. Non seulement la catégorie apporte une nuance de sens qui lui est inhérente (le verbe et le nom impliquent une action, l'adjectif implique un qualificatif etc..) mais le passage d'une catégorie à une autre peut être source d'un glissement de sens assez important, ainsi que le montrent les exemples mentionnés dans le Tableau 4 (les définitions ont été prises dans le Petit Robert).

Mot	Nom (masculin)	Adjectif	Verbe
quotidien	journal paraissant tous les jours	qui se fait, revient tous les jours	
marteau	outil de percussion, composé d'une masse métallique percée d'un trou dans lequel est fixé un manche	être marteau, fou, cinglé.	
livre	assemblage d'un assez grand nombre de feuilles portant des signes destinés à être lus.		remettre à l'acheteur ce qui a été commandé

Tableau 4 Différences de signification liées à la catégorie grammaticale.

La catégorie est donc intimement liée à la formation même du lemme. En effet, pour former un lemme, du moins dans notre conception, on utilise en plus d'une forme générique du mot, sa catégorie grammaticale. Ce raisonnement devrait être, en principe, maintenu pour subdiviser en plusieurs lemmes une forme ayant plusieurs sens. Nous en

convenons, mais pour des raisons pratiques³⁹, nous n'avons pas poussé la subdivision jusqu'aux significations du mot. A la différence de Y. Choueka et S. Lusignan (1985), nous avons exclu le recours au sens pour faire une distinction lemmatique". Le mot *son*, par exemple, est analysé comme *son-nom*, *son-possessif*. Les distinctions de sens que porte *son-nom* sont analysées à un autre niveau et ne conduisent nullement à parler de deux lemmes différents (*son* au sens 1 et *son* au sens 2).

Cette attitude nous satisfait d'autant plus que nous ne voulions pas déléguer à MARQUEUR une lemmatisation trop pointue. Il suffisait pour notre tâche de disposer des formes et de leurs lemmes respectifs. Le reste, c'est-à-dire la définition sémantique, devra faire l'objet d'une autre application, dédiée à manipuler les significations qu'on rattacherait aux lemmes. Le Tableau 5 constitue un autre exemple de lemmatisation. On peut observer en particulier le mot *droits*.

Mot-forme	Lemme + catégorie
bureau	BUREAU-NOM
bureaux	BUREAU-NOM
droit	DROIT-NOM
	DROIT-ADJECTIF
	DROIT-ADVERBE
fermé	FERMER-VERBE
	FERMÉ-ADJECTIF
actuellement	ACTUELLEMENT-ADVERBE

Tableau 5 Exemple de mots lemmatisés.

La seconde question posée plus haut consiste à décider si deux homonymes doivent être représentés comme deux lemmes différents ou comme un seul. Notre choix a été, ainsi qu'il a été annoncé au chapitre I, de représenter une seule forme. Si celle-ci est polycatégorielle, il y aura autant de lemmes que de catégories; sinon, on ne rattachera à la forme en question qu'un seul lemme. Le Tableau 6 illustre notre propos.

³⁹ Pour simplifier les relations des fichiers, entre autres. Il faut noter aussi que les différences des traits morpho-syntaxiques tels que le nombre, le genre et la structure argumentale (transitif, intransitif, etc...) ne provoquent pas dans notre conception de subdivision lemmatique.

Mot-forme	Lemme + catégorie
livre	LIVRE-NOM
	LIVRER-VERBE
grève	GRÈVE-NOM

Tableau 6 Exemple de lemmatisation de mots homonymes.

Quoiqu'il en soit, au delà de ces considérations théoriques, le recours au concept de lemme est très avantageux pour la structure des données compte tenu de l'économie qui en découle. Il est en effet plus pratique de travailler sur une seule entité générique que sur une quarantaine de formes variées. Cette préoccupation de l'optimisation pourrait nous conduire à une lemmatisation plus poussée, en l'occurrence en regroupant le nom et le verbe dans un hyperlemme, mais cela serait trop complexe à gérer. L'argument serait alors que certains noms ont le même noyau sémantique que les verbes, mais précisément, le principe n'est pas général, puisque, comme nous venons de le voir, dans certains cas, le passage à une autre catégorie amène une autre différence sémantique qui peut être importante.

Le procédé de marquage se fait d'une manière semi-automatique. Dans une première étape, on commence par segmenter le texte. Cette opération, à la base de la plupart des traitements informatiques des textes, se fait en fonction de la présence d'un blanc dans la chaîne de caractères qu'est le texte. Cette démarche n'est pas parfaite car deux problèmes au moins se manifestent au vu des résultats. Le premier est directement lié à la définition d'un mot. Ainsi, une entité comme *pomme de terre* va-t-elle être découpée en trois mots, engendrant de cette façon un problème difficile à résoudre. Comme notre objectif n'est pas de résoudre tous les problèmes rencontrés, nous n'avons apporté qu'une solution provisoire, mais qui est très efficace dans le contexte qui est le nôtre. Il s'agit de conserver ce découpage et de traiter toutes les composantes de l'entité en question comme ayant la même catégorie. De cette manière, les programmes qui vont suivre vont avoir affaire à la même catégorie, comme s'il s'était agi d'un mot unique. Ainsi, l'entité *pomme de terre* sera représentée comme suit:

Segment	Catégorie
<i>pomme</i>	Nom
<i>de</i>	Nom
<i>terre</i>	Nom

Tableau 7 Découpage d'un mot composé.

Cette solution peut sembler gênante. Nous en convenons, mais elle nous permet de contourner un obstacle qui autrement nous aurait détourné de notre problématique.

Le second problème soulevé par la méthode de découpage est mineur. Ce sont autant de phénomènes dus à l'imperfection ou à la diversité des conventions de typographie. Tel est le cas quand une ponctuation non rattachée au mot précédent est générée par le programme comme formant un mot. La solution à cette bétise doit être envisagée au niveau d'un traitement de texte (au sens trivial du terme) qui pourra à travers un balayage du texte remplacer les "blanc+ ponctuation" par une ponctuation sans le blanc.

La deuxième étape consiste à effectuer un marquage des mots du texte. Ce traitement se fait en deux phases. La première est automatique et ne demande aucun effort à l'utilisateur. Le principe sous-jacent à ce traitement est de comparer le mot pointé dans le texte à son corollaire dans un dictionnaire. Le dictionnaire en question recense les mots de la langue et pour chacun d'entre eux sa catégorie grammaticale et son lemme. Si le mot *bureau* figure dans les entrées de ce dictionnaire, le programme pourrait, à la rencontre de chaque occurrence de ce mot dans un texte soumis au marquage automatique, opérer un marquage en fonction de ce qui est inscrit dans le dictionnaire. Naturellement, dans ce traitement automatique, non seulement le mot doit se trouver dans le dictionnaire, mais en plus, il ne doit pas être ambigu quant à sa catégorie ou à son lemme.

Le dictionnaire est en fait une structure vide qui se remplit au fur et à mesure qu'on traite des textes. Chaque texte traité apporte son lot de mots nouveaux.

La deuxième phase est interactive. C'est à ce niveau que l'utilisateur apporte au programme sa connaissance de la grammaire. En effet, l'utilisateur est sollicité dans deux situations différentes. Voyant défiler sur l'écran le texte ligne après ligne, il aura à s'interroger sur le mot mis en relief. Ceux des mots qui ne sont pas mis en relief ont été traités automatiquement, il s'agit de mots univoques qui n'ont pas besoin d'attirer l'attention de l'utilisateur. Ce dernier, converti pour l'occasion en lexicologue, sera confronté alors aux cas suivants:

- 1- Le mot du texte est absent du dictionnaire (comme le dictionnaire est vide au départ, il est normal que les mots non encore rencontrés ne répondent pas à l'appel). Dans ce cas, deux cas se présentent:
 - a- Le mot est univoque: l'utilisateur vérifie le mot dans son contexte et lui attribue la catégorie et le lemme appropriés (ces informations vont être associées au mot du texte, mais enrichiront également le dictionnaire).
 - b- Le mot est plurivoque en langue (dans l'absolu, beaucoup de mots ont plus d'un lemme et plus d'une catégorie): l'utilisateur, en plus de décider de ce qui est attesté (i. e. désigner parmi les choix offerts le lemme-catégorie qui convient au mot mis en relief dans le texte), doit envisager de compléter le dictionnaire en ajoutant les propriétés potentielles du mot, à savoir les autres lemmes-catégories que le mot peut revêtir dans l'absolu.
- 2- Le mot est présent dans le dictionnaire, mais il est plurivoque: il lui a été précédemment associé deux ou plusieurs lemmes-catégories. Dans ce cas, le programme affiche à l'écran les différentes possibilités épinglées dans le dictionnaire. L'utilisateur, qui reste maître dans l'appréciation de ce qui lui est proposé (au besoin, il peut modifier les données du dictionnaire), doit sélectionner le lemme-catégorie approprié au mot en contexte.

Le Tableau 8 présente une table de décision qui classifie ces différentes situations:

Mot du texte			
Absent du dictionnaire		Présent dans le dictionnaire	
Univoque	plurivoque	Univoque	plurivoque
On le marque dans le texte. Cette marque est enregistrée également dans le dictionnaire	On le marque dans le texte. Cette marque est enregistrée également dans le dictionnaire. On ajoute les marques potentielles qui seront enregistrées dans le dictionnaire	Le mot n'est pas proposé au marquage interactif puisqu'il est censé avoir été traité par le marquage automatique.	On sélectionne les marques qui vont être associées au mot dans le texte. Les marques potentielles sont déjà enregistrées dans le dictionnaire, mais au besoin, on peut apporter une correction par des ajouts et/ou des suppressions.

Tableau 8 Les différents cas de figure dans le marquage.

En fait, nous utiliserons ces deux propriétés (lemme et catégorie) d'une manière inséparable. Voir Tableau 9. On doit noter d'autre part que plusieurs formes peuvent partager un seul lemme. De même, une seule forme peut être liée à plusieurs lemmes.

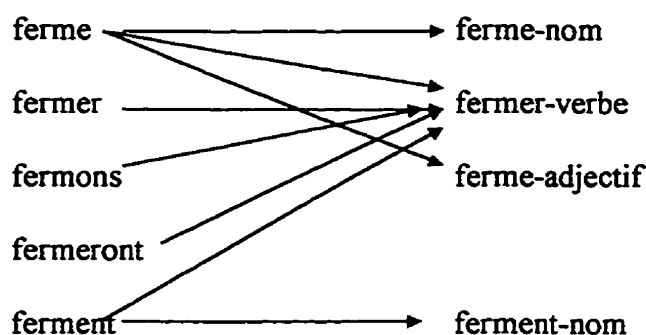


Tableau 9 Relation des formes et des lemmes.

Cet état de fait entraîne une complexité au niveau des relations à établir entre les formes et leurs lemmes respectifs. Une approche possible de ce problème est de structurer les données de la façon suivante: voir le Tableau 10.

MARQUEUR		
FORMES.DIC	LIENS.LIE	LEMES_DC2
F-1 Ferme	F-1 L-1	L-1 Ferme-NOM
	F-1 L-2	L-2 Fermer-VERBE
	F-1 L-3	L-3 Ferme-ADJECTIF
F-2 Fermer	F-2 L-2	
F-3 Ferment	F-3 L-2	
	F-3 L-4	L-4 Ferment-NOM

Tableau 10 Utilisation du fichier des liens.

Le fichier des liens LIENS.LIE (deuxième colonne) a pour fonction de lier le fichier des formes au fichier des lemmes. Grâce à cette structure de fichiers, il est possible de retrouver tous les lemmes associés à une forme donnée, et toutes les formes associées à un lemme donné. Bien entendu, en suivant cette méthode, nous évitons la redondance dans l'enregistrement de l'information, car aussi bien la forme que le lemme ne sont représentés qu'une et une seule fois.

b) Marqueur au service d'autres applications

MARQUEUR est un outil qui s'intègre parfaitement à notre système de désambiguïstation lexicale. C'est grâce à lui que les mots du texte traité peuvent être pourvus des deux propriétés susceptibles de lever une part de l'ambiguïté.

Au cours de la réalisation, il nous a paru intéressant d'ouvrir le logiciel à des besoins plus larges que le nôtre. Ainsi, le champ destiné à recevoir le lemme a été appelé *descripteur*; ce concept nous paraît plus approprié dans la mesure où l'utilisateur peut y inscrire la description qu'il juge pertinente, la définition qui répond à ses préoccupations. De son côté, la table des catégories grammaticales peut être partiellement ou complètement modifiée; l'utilisateur peut avoir besoin d'autre chose que de la catégorie grammaticale. Mais l'apport le plus saillant, à notre sens, est d'avoir permis à l'utilisateur d'associer aux mots une douzaine de propriétés supplémentaires. Des champs additionnels peuvent ainsi être créés et utilisés.

L'assignation aux mots de ces autres propriétés se fait interactivement dans une interface qui présente des champs configurés préalablement par l'utilisateur. Les informations ainsi saisies iront enrichir le dictionnaire associé au texte traité.

De cette manière, MARQUEUR se présente comme un outil ouvert à différents types de marquage. Sa flexibilité et les possibilités de manipuler des fichiers de type base de données font de lui un SGBD textuel. Appelons-le SGBDT (Système de Gestion de Base de Données Textuelles). Pour une illustration de ces aspects, nous renvoyons au guide de MARQUEUR placé en annexe.

Pour toutes ces raisons, il nous a paru plus exact d'appeler le logiciel en question MARQUEUR. Ainsi, on peut marquer des mots⁴⁰, dans le sens où on leur associe soit des informations disponibles dans une table⁴¹ prédéfinie, soit des informations à saisir dans des zones (champs de différents types⁴²) appropriées.

2. Le programme COLLOCATEUR

COLLOCATEUR est un programme qui a pour fonction essentielle d'extraire des collocations dans un texte donné. Une collocation est entendue ici dans un sens assez large. Il s'agit de séquence de mots liés syntagmatiquement. Le résultat de l'extraction se présente sous forme de liste de collocations. C'est dans cette liste que nous sélectionnerons les collocations qui serviront dans le processus de désambiguïisation. Cette fois-ci, les collocations sélectionnées sont moins arbitraires: il s'agit précisément de collocations de type structuré, tel que défini au chapitre I.

Le principe de base implanté dans le programme consiste à parcourir un texte d'une manière linéaire, en tâchant d'extraire les suites de mots qui répondent aux critères suivants:

⁴⁰ Le concept de tagging, en anglais.

⁴¹ Table au sens de fichier ou de liste.

- 1- Un mot est considéré comme candidat à la formation d'une collocation si sa catégorie grammaticale a été désignée par l'utilisateur comme membre potentiel d'une collocation.
- 2- On tient compte de la position du mot examiné dans la chaîne des mots. Un mot peut se trouver dans l'une ou l'autre des trois positions: initiale, médiane et finale. Une ponctuation forte ou un mot non candidat marque la fin d'une collocation.
- 3- Une collocation doit comprendre un minimum de deux mots.

Sachant que tout mot du texte est marqué de sa catégorie grammaticale (le texte est supposé déjà marqué) et que, d'autre part, il doit occuper dans une suite d'unités une des trois positions différentes (initiale, médiane, finale), il s'agira pour l'utilisateur d'évaluer les chances pour chaque catégorie d'être membre de collocation dans chacune des positions possibles.

Le programme, fort de ces informations assignées par l'utilisateur, procède à une exploration linéaire du texte. Admettons qu'il tombe sur la première suite:

*Le premier titre traite de la jouissance et de l'exercice des droits civils et il énonce les principes généraux en la matière.*⁴³

Faisons aussi l'hypothèse que l'utilisateur a retenu les paramètres synthésés dans le Tableau 11. Les trois colonnes de droite de ce tableau se lisent comme une réponse à la question: le mot pourvu de cette catégorie est-il candidat ou non à la formation d'une collocation, à la position mentionnée?

⁴² Caractère, numérique, logique.

⁴³ Code Civil du Québec p. 2.

Catégorie	Initiale	Médiane	Finale
Nom	Oui	Oui	Oui
Verbe	Non	Non	Non
Adjectif	Oui	Oui	Oui
Déterminant	Oui	Oui	Non
Pronom	Oui	Oui	Non
Adverbe	Non	Oui	Oui
Autre	Non	Non	Non

Tableau 11 Règles d'admissibilité dans Collocateur.

En considérant que le texte est préalablement marqué, le programme devrait agir de la façon suivante:

Le premier mot rencontré étant *le*, le programme s'enquiert de sa catégorie, il saura que c'est un déterminant. Comme ce mot est placé à l'initiale, et que dans cette position, il est candidat à la formation d'une collocation, ce mot va être retenu comme premier item d'une éventuelle collocation. Nous disons *éventuelle* car le succès d'une collocation est tributaire de la rencontre d'au moins un deuxième candidat.

La suite contient le mot *premier*. C'est un adjectif et est candidat dans la position où il se trouve, il sera donc retenu.

Le prochain mot est un nom. Il sera retenu. À ce stade, on aura retenu : *Le premier titre*. En continuant, on tombe sur le mot *traite* qui est un verbe et qui n'est pas candidat. On clôt la première collocation et on explore la suite du texte pour extraire d'autres collocations en suivant la même démarche.

Comme on le constate, le principe de fonctionnement de COLLOCATEUR ne repose sur aucun modèle syntaxique. On mise ici sur la nature grammaticale des mots (certaines catégories seraient autorisées, d'autres interdites); c'est le plan paradigmatique. On mise également sur la position des mots (pour une catégorie donnée, certains endroits de la chaîne sont plus productifs que d'autres); c'est le plan syntagmatique. On pourrait presque parler d'une syntaxe, si dans le processus il ne manquait pas le paramètre des relations entre les mots.

Néanmoins, cet algorithme combinatoire, sans être syntaxique, est très productif.

a) Collocateur au service de LOCALEX

Le programme COLLOCATEUR joue un rôle important dans le système de désambiguïisation. Son intervention est d'autant plus appréciable que la quantité des données à traiter peut être dissuasive pour tout effort manuel. COLLOCATEUR traite les données d'une manière complètement automatique.

Nous avons vu l'intérêt de disposer d'une liste de collocations et l'usage qui peut en être fait dans le système de désambiguïisation. Bien qu'il soit toujours possible d'aller cueillir des collocations d'une façon manuelle, soit dans des textes soit dans des dictionnaires, il est clair qu'un outil dispensant le chercheur de cette tâche fastidieuse ne peut être que le bienvenu. C'est la raison d'être de COLLOCATEUR. L'exécution de la tâche est automatique, mais en plus, le chercheur a le loisir de prendre la mesure des paramètres à retenir et des conséquences que ceux-ci impliquent.

Les collocations peuvent être extraites dans le respect de schèmes (patrons) décidés par l'utilisateur. On peut vouloir rechercher Nom + Verbe ou Verbe + Déterminant + Nom. Bien cibler ces combinaisons favorise avantageusement l'extraction des collocations structurées.

Un tri ultérieur se fera interactivement et permettra de nettoyer les "fausses trouvailles". Mais la récolte automatique est déjà intéressante. On peut voir un exemple dans le Tableau 12. (Remarquez que les collocations 7, 8 et 9 peuvent être considérées comme un bruit à écarter).

Référence	Collocation
1	Société religieuse
2	Caractère permanent
3	Mariages pour un temps
4	Ministre du culte
5	Célébration d'un mariage
6	Opportunité d'un examen médical prénuptial
7	Homme et un femme
8	Québec ministres du culte
9	Célébrant compétent et en présence de deux témoins

Tableau 12 Exemple de collocations extraites à l'aide de COLLOCATEUR.

En partant d'une conception très large de la notion de collocation, à savoir une suite de mots ayant des rapports syntagmatiques, on pourrait en déduire que n'importe quelle catégorie grammaticale pourrait participer à la formation d'une collocation. En paramétrant COLLOCATEUR en conséquence de cette vision (ce qui revient à répondre OUI à toutes les catégories quelle que soit la position), le programme retiendra systématiquement toutes les phrases. COLLOCATEUR n'aurait d'intérêt, on l'aura compris, que si l'on cherchait certains types de collocations, c'est-à-dire que si l'on privilégiait certains syntagmes au détriment d'autres. Ainsi, on peut vouloir des collocations verbales, ou nominales.

Il incombe malgré tout à l'utilisateur la tâche de sélectionner parmi les collocations obtenues celles qui conviennent le mieux à ses objectifs. Dans notre cas, il fallait retenir dans l'extrait préparé par COLLOCATEUR les collocations qui, en plus d'avoir une unité au plan syntagmatique, présentent un certain degré de lexicalisation, et, par ailleurs, avoir la propriété d'induire une signification particulière.

b) Collocateur au service d'autres applications

On peut penser à plusieurs applications qui pourraient bénéficier des services de COLLOCATEUR. Les collocations extraites à l'aide de ce programme peuvent servir de matière première pour des applications œuvrant dans des domaines comme l'indexation, la recherche documentaire, la terminologie, l'Analyse du Discours et bien d'autres. Le format du fichier adopté pour emmagasiner les collocations est une commodité supplémentaire.

Avant d'aborder le troisième programme, signalons qu'à ce stade et compte tenu de ce qui précède, les relations entre les différentes informations impliquées dans le processus de désambiguïation se présentent de la façon suivante (voir Tableau 13).

MARQUEUR			INTERACTIF	COLLOCATEUR
FORMES.DIC	LIENS.LIE	LEMME.DC2	SENS.SNS	EXPRES.XPR
F-1 Ferme	F-1 L-1	L-1 Ferme-NOM	L-1 sens 1 S-1	S-1 Expr 1
				S-1 Expr 2
				S-1 Expr 3
			L-1 sens 2 S-1	S-2 Expr 1
				S-2 Expr 2
	F-1 L-2	L-2 Fermer-VERBE		
	F-1 L-3	L-3 Ferme-ADJECTIF		
F-2 Fermer	F-2 L-2			
F-3 Ferment	F-3 L-2			
	F-3 L-4	L-4 Ferment-NOM		

Tableau 13 Relations entre les différentes informations.

3. Le programme LOCALEX

Comme nous l'avons spécifié précédemment, LOCALEX est le programme qui nous permettra de mettre en œuvre notre hypothèse sur la désambiguïation lexicale automatique. Sa tâche concrète est de localiser dans un texte donné toutes les occurrences d'un mot-clé entré par l'utilisateur, en montrant à chaque fois si le lemme, la catégorie grammaticale et la signification de l'occurrence examinée correspondent ou non à la requête.

Bien entendu, l'interface est conçue de manière à ce que l'utilisateur puisse sélectionner les critères de sa requête sans avoir à se préoccuper de la complexité du processus qui se déroule en tâche de fond. L'utilisateur doit sélectionner un mot parmi ceux fournis dans un index reflétant la réalité du texte. Le parcours de l'index est agrémenté de l'affichage dynamique des différents lemmes et catégories grammaticales du mot que l'utilisateur pointe (avant même de le sélectionner). Ces informations sont puisées dans le texte, objet de la recherche. Une fois le mot sélectionné, ce sont les différentes significations qui sont soumises au choix de l'utilisateur. Les différentes significations proposées proviennent, quant à elles, du dictionnaire implanté dans le logiciel. La recherche des occurrences présentes dans le texte se

fera ensuite dans le respect de ces trois contraintes: la catégorie grammaticale, le lemme et la signification.

Donc, pour chacune des occurrences repérées, le programme spécifiera si le lemme, la catégorie grammaticale et la signification mis en jeu dans le contexte où l'on se trouve correspondent bien à ce que l'utilisateur désire. Pour faciliter la lecture des réponses possibles de LOCALEX, et en même temps éviter de charger l'affichage du texte de trop de codes, nous avons opté pour l'utilisation d'un jeu de couleurs auxquelles nous avons attribué respectivement une signification particulière. Le Tableau 14 énumère ces couleurs et leurs significations respectives. Nous reprendrons ce même tableau, mais en l'enrichissant, au chapitre III.

Couleur	Interprétation
Rouge / Blanc	1- C'est le bon duo lemme-catégorie, et la bonne signification.
Noir / Blanc	2- C'est le bon duo lemme-catégorie, mais ce n'est pas la bonne signification.
Bleu / Blanc	3- Ce n'est ni le bon duo lemme-catégorie, ni la bonne signification.

Tableau 14 Codes adoptés par Localex.

Pour être en mesure de faire ces distinctions (i. e. pour procéder à la désambiguïisation), un certain nombre d'investigations s'imposent. La première tâche est relativement aisée. Elle consiste à déterminer si l'occurrence rencontrée est pourvue du même lemme-catégorie que celui désiré par l'utilisateur. Sachant que ces deux informations ont été enregistrées lors du marquage, cette vérification est triviale (une simple consultation des propriétés des mots du texte!).

La deuxième tâche est de déterminer la signification effective dans le contexte. Pour ce faire, le programme consulte le dictionnaire, localise dans ce dernier la forme objet de la recherche, et prend note du lemme et de la catégorie associés à cette forme (en cas de plurivocité, le programme connaît déjà les choix de l'utilisateur); à partir de là, le programme peut s'enquérir des significations rattachées au lemme en question. Il s'arrête sur celle

sélectionnée par l'utilisateur et se positionne au niveau des collocations qui lui sont rattachées. Et à ce stade commence la confrontation des collocations de la base au contexte de l'occurrence repérée; le contexte peut être à gauche ou à droite de l'occurrence traitée. Chaque collocation est porteuse du code qui impose la direction à explorer: G: pour contexte gauche, D: pour contexte droit.

Cette confrontation constitue l'opération la plus délicate de tout le parcours. Elle est aussi le pivot central du processus de désambiguïisation lexicale automatique. Connaissant la rigueur légendaire de l'ordinateur, et reconnaissant d'autre part que dans la confection des textes, le scripteur dispose après tout d'une grande marge de manœuvre, on ne peut s'empêcher de craindre que la collocation emmagasinée dans le dictionnaire et qui est associée à la signification désirée ne soit pas représentée telle quelle dans le contexte; quand bien même elle le serait, ne pourrait-on pas craindre une quelconque divergence de forme⁴⁴ qui ferait échouer la concordance souhaitée? La crainte est justifiée et l'inquiétude a ses raisons d'être. Néanmoins, il y a lieu d'envisager les moyens de repousser ces appréhensions.

Le système se caractérise par quelques points forts qu'il convient de souligner. Comme nous l'avons signalé précédemment (voir chapitre I), dans notre stratégie, nous avons misé sur la probabilité de retrouver dans le contexte des collocations inductrices d'une signification donnée. Le risque que nous courions était que même dans le cas où la collocation prévue dans la base serait présente dans le contexte, il ne soit pas certain que cette dernière soit une image parfaite de la première. Car il faut que :

1. les mêmes items de la collocation de la base soient présents dans le contexte (du moins les lemmes qui les représentent);
2. ces items se retrouvent dans un espace limité (limites de la phrase), avec comme condition supplémentaire qu'une rupture de la structure ne soit pas fatale (un item trop éloigné risque de ne plus appartenir à la même

⁴⁴ Même dans les collocations figées, certaines parties peuvent être manipulées librement.

collocation);

3. ces items apparaissent dans le contexte de manière ordonnée.

En vérité, ces exigences ne sont pas toutes indispensables. Du moins, pas avec cette rigueur. Précisément, pour apporter une solution au premier problème (celui de la présence de chaque item dans le contexte), nous avons posé l'hypothèse qu'une sélection appropriée des items de la collocation permet d'augmenter les chances d'adéquation des items de la collocation emmagasinée à ceux rencontrés dans le contexte: si la recherche se fait uniquement sur la base des items pertinents, elle a plus de chance d'aboutir (elle est moins contraignante que si on exigeait de retrouver par exemple toutes sortes de mots de liaison).

Cette même sélection permettra de prévenir en partie le deuxième risque (concernant la rupture de l'unité de la collocation), dans la mesure où nous considérons qu'en dehors des mots essentiels de la collocation, un grand nombre de choses peuvent être ignorées par la recherche. Ainsi, pour une collocation comme *exercer droit*, les séquences suivantes sont toutes aussi valides.

- *Excercer un droit*

- *Exercer son droit*

- *Excercer tous les droits*

- *Etc.*

Nous avons dit *en partie* car, en effet, le risque d'une rupture fatale n'est pas exclu. Nous reviendrons sur cette question dans le chapitre III.

Pour prévenir le dernier problème (lié à l'ordre des items), il suffit de prévoir tous les ordres possibles que peuvent prendre les items de la collocation et représenter ces combinaisons dans la base. Ainsi, la collocation à *temps plein* sera enregistrée au même titre que à *plein temps*.

Lors de l'exploration du contexte, LOCALEX recherche les différents items de chaque collocation. En fait, chaque item de la collocation est comparé à la forme du mot du contexte et également au lemme correspondant (le texte étant marqué, la base est déjà alimentée du lemme de chaque mot). Cette stratégie présente, d'une part, une économie dans l'alimentation de la base (nous enregistrons un lemme au lieu d'une quarantaine de formes différentes) et, d'autre part, un gain de temps appréciable lors de la recherche (la comparaison ne se fait qu'une seule fois au niveau d'une unité au lieu d'une quarantaine de fois).

D'un autre côté, nous spécifions la direction de la recherche pour chaque collocation enregistrée dans la base. Ceci est indispensable car la position du contexte à explorer n'est pas indifférente à la discrimination sémantique. Il peut arriver cependant qu'une collocation soit favorable à une signification donnée indépendamment de sa position (à gauche ou à droite); par exemple: *suspendre un droit* est équivalente à *un droit suspendu*. Pour de tels cas, nous plaçons dans la base les deux collocations, comme le montre le Tableau 15 .

Collocation	Direction
G:suspendre	à gauche
D:suspendu	à droite
D :suspendre	à droite

Tableau 15 Représentation de collocation dans les contextes gauche et droit.

Ce que nous avons vu jusqu'à présent concerne l'utilisation des collocations structurées. Nous rappelons que dans notre projet, nous prévoyions également de recourir, mais seulement d'une manière limitée et donc complémentaire, aux collocations non structurées.

Le système mis en place accueille ces collocations au même titre que les premières. Dans le même champ, on peut placer aussi bien les unes que les autres. La procédure d'exploitation de ce type de collocations est de toute façon la même.

LOCALEX n'est cependant pas conscient qu'il recourt à une information d'un autre type. Si c'est le sens 1 du mot examen (épreuve contrôlée) qui l'intéresse, il utilisera

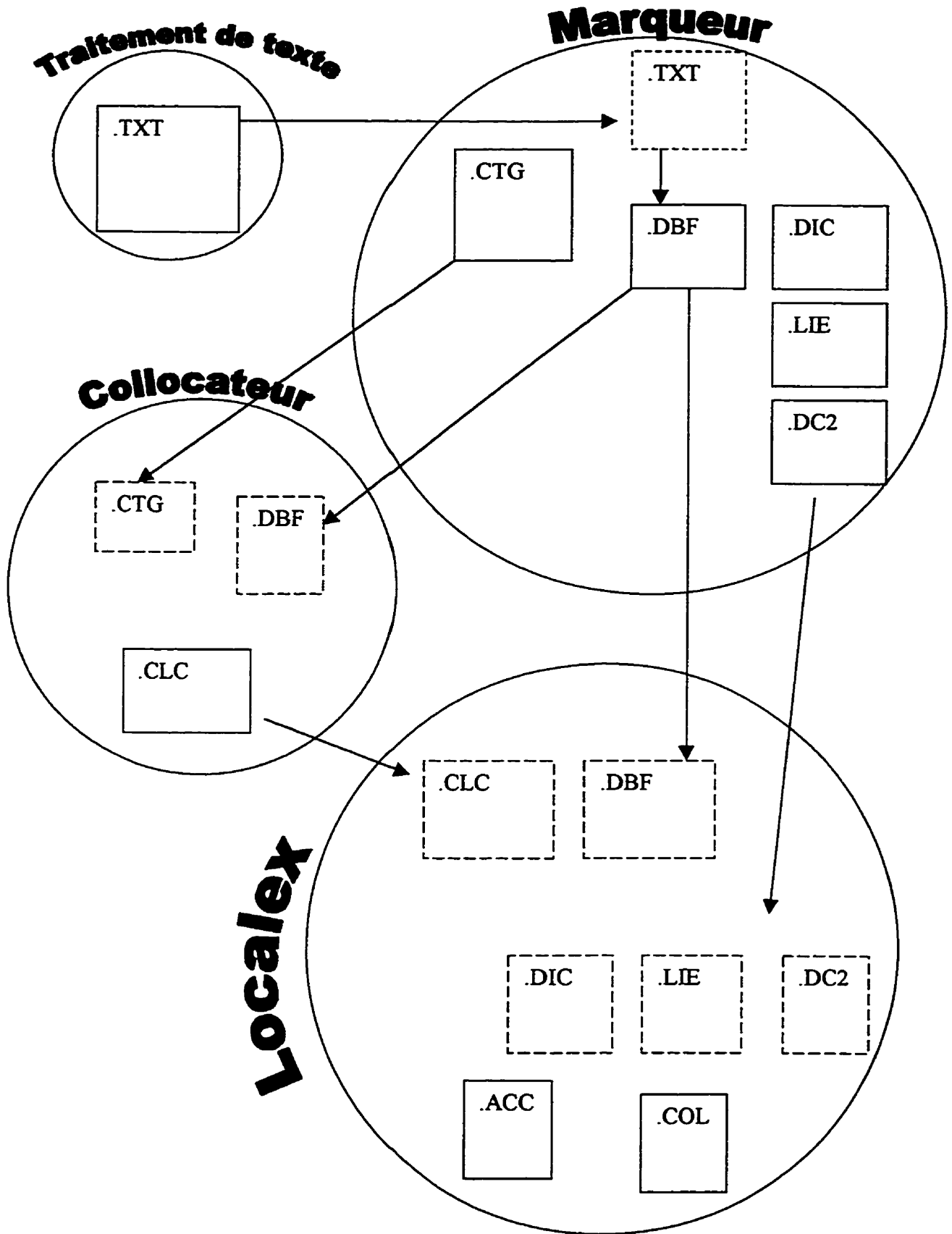
indifféremment les collocations du type: *réussir un examen* que le mot *concours*, la première étant de type structuré, la seconde étant de type non structuré.

4. Homogénéité et interdépendance des programmes

Le système de désambiguïisation est formé de l'ensemble des trois programmes. Mais c'est LOCALEX qui met en application notre hypothèse sur l'exploitation des collocations (conçues comme une des marques de la cohésion des textes). Dans notre stratégie, nous avons prévu un prétraitement du texte objet de la recherche et une préparation d'une base de connaissances comprenant un jeu de dictionnaires reliés (pouvant fonctionner comme un seul dictionnaire) et détenteurs des formes des mots de la langue, des lemmes et catégories respectifs de ces dernières, des significations rattachées à ces lemmes et enfin des collocations que l'expert a associées à ces significations. Toutes ces informations sont structurées avec la plus grande rigueur.

Il était par conséquent primordial que les trois programmes fonctionnent d'une manière homogène. Les programmes coopèrent entre eux et cela ne pouvait se faire sans prévoir des passerelles entre les fichiers. Pour ne donner qu'un exemple, chaque programme de niveau inférieur récupère les fichiers du programme précédent et les intègre dans un ensemble homogène que nous avons appelé *Projet* (voir les différents guides placés en annexe). On peut consulter la Figure 2 pour se faire une idée de l'architecture de cette coopération entre les programmes. On peut également se reporter à l'annexe (E.) pour prendre connaissance de la structure des fichiers utilisés dans le système.

Figure 2 Coopération entre les trois programmes et transfert des fichiers.



Chacune des trois applications peut utiliser deux types de fichiers: des fichiers importés (inclus dans un cadre au contour discontinu) et / ou des fichiers originaux (symbolisés par un trait plein). D'autre part, les fichiers se distinguent par la nature de l'information qu'ils contiennent. L'extension du fichier en est un témoin, décodable d'après le Tableau 16.

Extension du fichier	Signification
ACC	Fichier des acceptions ou significations
CLC	Fichier des collocations provenant de COLLOCATEUR
COL	Fichier des collocations retenues dans LOCALEX
CTG	Fichier des catégories grammaticales
DC2	Fichier du dictionnaire des lemmes
DIC	Fichier du dictionnaire des formes
LIE	Fichier des liens entres les formes et les lemmes
TXT	Fichier texte

Tableau 16 Extensions des fichiers et leurs significations

Les logiciels qui répondent à la tâche visée sont en même temps des outils de test. Un des bénéfices de la réalisation des logiciels est que ceux-ci permettent d'obtenir les conditions nécessaires à la mise à l'épreuve de la stratégie de désambiguïsation sans obstacle au niveau de la manipulation des données.

Il faudrait ajouter que les logiciels qui forment le système de désambiguïsation sont faciles à utiliser. Cette souplesse est le résultat de nombreux arrangements dont les principaux sont:

1. Intégrité et fluidité des données et compatibilité des fichiers.
2. Homogénéité des structures.
3. Modularité des programmes.
4. Menus intuitifs et ergonomie générale.

Ajoutons cette dernière remarque: l'utilisation de COLLOCATEUR est considérée comme une étape provisoire. Son rôle est d'aider à la préparation de la base. Il s'ensuit qu'on peut prévoir deux modes d'utilisation de notre système de désambiguïisation lexicale. Une utilisation maximale profite de l'apport de COLLOCATEUR qui relève les collocations du texte. Les résultats sont en principe très fructueux. Une utilisation minimale se passe des services de COLLOCATEUR. La recherche se fait sur la base des collocations déjà présentes. En principe, c'est ce mode qui correspond à une utilisation normale du système. Au préalable, naturellement, l'alimentation de la base reste à faire. Nous récapitulons ces deux modes d'utilisation dans le Tableau 17.

Type d'utilisation	Prétraitement	Résultat escompté
Le système sera appliqué à des textes dans un mode de traitement <u>minimal</u> .	Marquage du texte à l'aide de <i>MARQUEUR</i> . Le texte devient prêt pour une exploration par <i>LOCALEX</i> .	L'efficacité du processus de désambiguïisation sera fonction de la probabilité de rencontrer dans le texte les collocations relevées dans la base
Le système sera appliqué à des textes dans un mode de traitement <u>maximal</u> .	Marquage du texte à l'aide de <i>MARQUEUR</i> . Collecte des collocations à l'aide de <i>COLLOCATEUR</i> . Enrichissement du fichier des collocations (inclus dans <i>LOCALEX</i>) par les collocations relevées grâce à <i>COLLOCATEUR</i> . Le texte devient prêt pour une exploration par <i>LOCALEX</i> .	L'efficacité du processus de désambiguïisation sera largement accrue par le fait que les collocations relevées dans le texte sont représentées dans la base du système.

Tableau 17 Modes d'utilisation du système de désambiguïisation lexicale.

5. La base des connaissances lexicales

Dans notre stratégie, nous avons prévu de recourir à deux types d'informations. Les premières sont d'ordre morphosyntaxique et lexical (caractérisation du lemme et de la catégorie grammaticale), les secondes concernent ces tranches de discours très particulières

que sont les collocations (particulières en ce sens qu'à travers leur construction, leur récurrence dans le discours, elles finissent par devenir le corrélat de significations déterminées).

Nous avons développé les outils nécessaires pour chacune de ces bases de données. Comme nous venons de le dire dans la section précédente, ces programmes s'intègrent bien au système dans sa totalité. LOCALEX ne pouvait fonctionner sans le secours des deux autres programmes. Mais cette nécessité ne devrait pas cacher un fait essentiel à nos yeux, que ces deux premiers programmes sont (de par leur existence à l'état actuel et de par leur évolutivité) un argument en faveur de la philosophie que nous avons adoptée. Cette manière de faire consiste à utiliser les ressources langagières dans toute leur matérialité. Prendre le langage à bras-le-corps en quelque sorte. D'où le recours à des bases de connaissances dépourvues de formalisme. Nous avons en fait considéré que les données brutes sont riches d'informations, et qu'il était possible d'exploiter cette richesse sans avoir à l'envisager sous l'angle métalinguistique. C'est dans cet ordre d'idées que nous faisons les remarques suivantes:

- La préparation des données est en grande partie automatisable et le traitement est simple et facile à effectuer. Même MARQUEUR peut bénéficier d'une plus grande automatisation, avec la prise en compte de règles contextuelles. Il s'ensuit que le succès de l'opération de désambiguïisation lexicale (le succès de LOCALEX) n'est pas hypothéqué par un travail manuel démesuré.
- Le travail manuel correspond à des opérations lexicographiques de fond: inventorier les propriétés des mots, recueillir des collocations, comme pour les dictionnaires conventionnels. Les informations ainsi recueillies sont d'ordre très général et à l'exception de quelques transformations au niveau des collocations, elles constituent un fonds de données indépendant des spécificités du système de désambiguïisation.
- Les connaissances lexicographiques utilisées (pour l'ensemble ou un sous-ensemble de la langue) représentent un travail d'une étendue et d'une complexité telles que seule la puissance de traitement informatique conjuguée

à l'efficacité des supports numériques permettrait, en pratique, d'entreprendre. Il est par conséquent pertinent d'avoir développé (et intégré dans le système) des logiciels d'acquisition de ces connaissances.

C. CONCLUSION

Les trois programmes que nous venons de présenter constituent un système complet de désambiguïisation lexicale automatique. Chacun d'entre eux accomplit une part du travail. MARQUEUR attribue aux mots lemme et catégorie. COLLOCATEUR extrait des séquences suffisamment cohésives pour être envisagées comme des collocations. Un simple regard de l'expert humain (chargé de préparer la base lexicale) suffit pour rejeter les suites non désirées et conserver celles dites structurées. Notons au passage que nous n'avons pas prévu d'outils susceptibles de recueillir des collocations non structurées. Enfin, LOCALEX va tirer profit du marquage pour accomplir un premier niveau de désambiguïisation: le niveau morphosyntaxique. Il exploitera ensuite les collocations (au sens structurel du terme), préalablement rattachées à des significations particulières, pour opérer une discrimination sémantique.

Nous avons montré que le succès d'une telle démarche dépendait en partie de la cueillette des collocations. Une certaine expertise est supposée lors de l'alimentation de la base. Mais le succès vient aussi des règles établies pour représenter adéquatement les items de chaque collocation. Une grammaire contextuelle et computationnelle est à inventer.

CHAPITRE TROISIÈME
MISE A L'ÉPREUVE ET RÉSULTATS

3. MISE A L'ÉPREUVE ET RÉSULTATS

A. INTRODUCTION

Dans ce dernier chapitre, nous rendons compte de la mise à l'épreuve du système de désambiguïsation lexicale automatique proposé et nous procédons à l'examen des résultats.

Partant du fait qu'un certain nombre de conditions sont satisfaites (chapitre II), il s'agira de voir pour certains mots (voir infra les critères de sélection de ces mots) si le logiciel LOCALEX est en mesure de repérer toutes les occurrences de ces mots dans le contexte et d'identifier celles qui sont pourvues de la signification recherchée. Le logiciel adopte pour chaque situation une couleur d'affichage différente (voir le Tableau 22 où sont présentées ces conventions).

En procédant à cette mise à l'épreuve du système, nous essaierons de répondre aux objectifs suivants:

1. Montrer le taux de succès obtenu en mettant en application notre stratégie de désambiguïsation.
2. Identifier les limites de la démarche aux niveaux linguistique, conceptuel et technique.
3. Formuler des propositions pour dépasser les écueils rencontrés.

Dans ce qui suit, nous évoquerons des questions de méthode, en particulier la démarche que nous suivrons dans cette mise à l'épreuve, et nous présenterons plus loin les résultats des tests qui seront accomplis.

B. QUESTIONS DE MÉTHODE

Pour répondre à l'objectif fondamental du test, c'est-à-dire montrer la validité du système, il conviendrait de définir une démarche susceptible de nous permettre de répondre

aux objectifs fixés. Cette exigence est d'autant plus pertinente qu'il y a beaucoup de paramètres qui entrent en jeu et qu'il est nécessaire de pouvoir discerner l'intervention de chacun. Rappelons que le système procède à différents niveaux de repérage (forme, catégorie, lemme, signification) et qu'en plus de cela, il y a des cas de rejet, et des cas d'acceptation. Ajoutons à cela que les raisons du rejet comme celles de l'acceptation peuvent varier. Le point suivant esquisse la démarche générale. Nous verrons qu'il sera nécessaire de distinguer entre deux types de test. Enfin, dans le second point, il sera question des paramètres qu'il conviendrait d'observer.

1. Démarche générale

Nous tenons à souligner dès à présent que la vérification ne vise pas à confirmer le bon fonctionnement des différents programmes, au sens où l'on dit qu'un programme *tourne*. Notre objectif ici n'est pas non plus de relever les erreurs de programmation. Bien que ces aspects puissent avoir leur importance (confectionner des outils qui fonctionnent d'une manière satisfaisante est en soi une tâche intéressante et fructueuse⁴⁵ pour qui travaille dans le domaine de la linguistique informatique), il reste que notre problème initial est d'explorer une idée: exploiter les ressources de la cohésion dans un processus de désambiguïsation lexicale automatique. On sera donc plus attentif aux résultats que la démarche proposée aura permis d'obtenir, aux problèmes soulevés et aux voies qui auraient pu être suivies.

Nous partons de l'hypothèse que le corpus est soumis à un traitement préalable: le corpus doit être marqué. Cette tâche se fait à l'aide du logiciel MARQUEUR. Une autre phase du traitement préalable consiste en l'extraction des collocations en utilisant COLLOCATEUR. Une troisième condition consiste en l'alimentation des bases dans LOCALEX.

⁴⁵ Voir l'article de Nancy IDE (1987: 209-215) où l'auteur, rendant compte d'une conférence sur les bases philosophiques et les méthodes à adopter dans l'enseignement de l'informatique aux étudiants des sciences humaines, oppose un noyau dur favorable à l'apprentissage des techniques de la programmation à une attitude moins ambitieuse qui prône l'initiation à des outils informatiques appropriés au domaine.

La première opération du test consiste à faire une série de recherches lexicales (LOCALEX se prête à ce genre de recherches, il a été prévu pour cette tâche) et de vérifier le comportement du logiciel.

Il y a au moins deux dimensions à vérifier dans le système proposé, sachant que ce dernier recourt d'une part à des règles (sous forme d'algorithme) et d'autre part à des connaissances (sous formes de bases de données structurées): pour chacune de ces dimensions, nous concevons une démarche différente. La première, destinée à interroger la stratégie de désambiguïsation et la validité de l'algorithme, consiste à faire abstraction des problèmes relatifs à l'alimentation de la base et à se concentrer uniquement sur les problèmes dus à notre démarche. Dans ce cas, il faut travailler avec une base de collocations exhaustive pour le corpus choisi⁴⁶. Ce type d'épreuves nous paraît fondamental pour étayer la justesse de notre hypothèse de travail. C'est pourquoi, nous comptons ici lui consacrer une part importante. Nous appellerons ce test: le test de la validité de la démarche.

La seconde manière de procéder, destinée à évaluer la rentabilité de la base, consiste à vérifier si une base de collocations extraite d'un texte témoin est rentable pour d'autres textes dont on n'a pas relevé les collocations. Une des raisons qui nous ont conduit à recourir à l'exploitation des collocations dans un système de désambiguïsation automatique, c'est qu'intuitivement, nous croyions à l'hypothèse selon laquelle un type de discours donné est caractérisé par une distribution homogène du lexique au point qu'un corpus témoin⁴⁷ fournit une base de collocations suffisante pour traiter d'autres corpus du même type.

Si les résultats de cette seconde démarche sont probants, cela n'apportera sans doute pas la preuve de la validité de cette dernière hypothèse; il faudrait pour atteindre ce but recourir à une recherche de plus grande échelle et à un arsenal théorique plus approprié, mais cela donnera une indication en sa faveur et, du même coup, nous confortera dans la voie que nous avons choisie. D'autres recherches pourraient être menées pour répondre avec plus de

⁴⁶ En plus des collocations relevées à l'aide de COLLOCATEUR, nous avons répertorié dans la base des collocations repérées *manuellement*.

⁴⁷ Représentatif de par l'usage qui en est fait dans la société et de par l'autorité dont il jouit.

force à la question de la distribution du lexique dans un type de discours donné. Précisément, nous croyons que l'utilisation de la méthode suivie ici pour l'exploration de plus larges corpus pourrait être un outil heuristique susceptible d'apporter un éclairage sur cette question.

A défaut de ces conditions, et dans l'intention de montrer que la technique de désambiguïsation proposée est valable, particulièrement l'idée qu'un nombre de collocations réduit peut être étendu au delà du corpus de base, nous comptons appliquer ce second test d'une manière limitée. Nous appellerons ce test: le test de la représentativité de la base.

2. Paramètres à observer

Le système de désambiguïsation proposé procède à différents niveaux de repérage: la forme, la catégorie, le lemme, et la collocation. On devra par conséquent être attentif à chacun de ces niveaux. On devra également s'interroger pour confirmer si chacune des décisions de LOCALEX est valide aux yeux de l'expert humain. Nous expliciterons ces différents paramètres dans les points suivants.

a) L'adéquation de la forme

Il s'agit à ce premier stade de vérifier si LOCALEX a bien repéré la forme qui lui a été spécifiée dans la requête. L'opération consiste dans la comparaison de forme à forme et n'est censée présenter que de faibles difficultés conceptuelles. Si, toutefois, il y a échec de localisation, celui-ci peut se justifier par une mauvaise orthographe du mot du texte.

b) L'adéquation de la catégorie et du lemme

Ce repérage consiste à vérifier pour chaque occurrence rencontrée si le lemme et la catégorie inscrits lors du marquage correspondent respectivement au lemme et à la catégorie sélectionnés par l'évaluateur du programme.

c) L'adéquation graphique des collocations

Si les deux types de repérage précédents ne sont pas censés poser des problèmes d'adéquation, il en va tout autrement pour le repérage des collocations. Il y a à ce niveau un degré d'impondérable que notre stratégie s'est appliquée à réduire.

Au lieu de nous investir dans une démarche déclarative propre à l'Intelligence Artificielle, par exemple, en fondant nos règles et la base de connaissances dans le moule de la logique du premier ordre et en adoptant une programmation de type Prolog, nous avons opté pour une démarche procédurale. Nous avons cependant emprunté à l'Intelligence Artificielle l'idée de la séparation de la base de connaissances (les significations, les collocations) de l'algorithme qui les manipule.

Mais au-delà de ces arrangements, nous avons surtout compté sur la façon de représenter les collocations. Ce qui doit être souligné est que la collocation enregistrée est réduite à un substrat composé uniquement des items que l'expert humain juge comme étant les plus essentiels. Autrement dit, la collocation est nettoyée de tous les éléments trop généraux (les déterminants par exemple) ou dont la suppression n'enlève rien à la cohésion de la collocation.

(1) Les différents types d'adéquation**(a) Adéquation continue**

Dans LOCALEX, la comparaison se fait entre une collocation de la base et son corollaire dans le contexte. On parlera d'*adéquation continue* de ces deux séquences comparées quand les mots repérés dans le contexte (collocation contextuelle) sont contigus de la même manière que le sont les items de la collocation de la base. Comme exemple, voir le Tableau 18:

collocation de la base	contexte
droit de circuler	le droit de circuler
droit accordé	le droit accordé
droit de propriétaire	le droit de propriétaire

Tableau 18 Exemples d'adéquation continue.

(b) Adéquation discontinue

On dira que la collocation est discontinue quand d'une part les items rencontrés dans le contexte sont séparés par un ou plusieurs mots, et que d'autre part, cette séparation ne provoque pas une déstructuration de la collocation; comme exemple, voir le Tableau 19.

Rappelons que cette discontinuité n'est pas jugée négativement. Tout au contraire, nous la prévoyons dès l'alimentation de la base, et nous l'intégrons à notre stratégie d'optimisation des données.

collocation de la base	contexte
droit de obtenir	le droit d'en obtenir
droit accordé	le droit est accordé
droit de propriétaire	le droit de ces propriétaires

Tableau 19 Exemples d'adéquation discontinue.

(c) Adéquation disloquée ou pseudo-collocation

On dira que la collocation a une adéquation disloquée quand, d'une part, les items rencontrés dans le contexte sont séparés par un ou plusieurs mots et que d'autre part, cette séparation entraîne une déstructuration de la collocation; comme on le voit dans l'exemple du Tableau 20, le mot propriétaire n'a plus rien à voir avec la particule de localisée par LOCALEX.

collocation de la base	Contexte
droit de propriétaire	Pour obtenir l'immatriculation d'un véhicule routier et le droit (2358) <u>de</u> mettre ce véhicule en circulation, le <u>propriétaire</u> de ce dernier doit: 1 satisfaire aux conditions et aux formalités établies par règlement; 2 (paragraphe abrogé);

Tableau 20 Exemple de la pseudo-collocation.

Remarquons que le terme *dislocation* pourrait très bien être érigé comme l'antonyme du mot *collocation*. Mais il a le défaut de laisser entendre que la séquence en question avait le statut d'une collocation et qu'elle a perdu ce statut. C'est pourquoi nous allons lui préférer le terme *pseudo-collocation*. Le concept est vu ici comme une forme d'échec. Une autre remarque, et non des moindres, s'impose ici: les items dont on dit —en partant de la définition de la collocation structurée— qu'ils sont disloqués, peuvent en fait former, grâce à cette relation privilégiée qu'est tout simplement leur co-occurrence, une collocation non structurée. Si tel est le cas, l'échec de l'une est compensé par la réussite possible de l'autre. C'est pourquoi, malgré tout, la pseudo-collocation n'entraîne pas systématiquement un dysfonctionnement de la stratégie. Elle serait généralement préjudiciable quand les items dont la collocation est constituée n'auraient aucune raison d'être associés dans un contexte proche. Dans l'exemple du Tableau 20 Les items *droit* et *propriétaire*, quoique disloqués, conservent la relation de collocation au sens non structuré, ce qui permettra de préserver le succès de la bonne signification.

(2) L'inadéquation et ses causes

Si l'on exclut le cas où la base ne contiendrait aucune collocation relative à la signification recherchée, les raisons de l'inadéquation peuvent être imputées au contexte. Différents cas se présentent, entre autres:

1. Carence de la base.

2. Contexte absent ou pauvre.
3. Contexte ambigu.

Nous reviendrons plus loin sur chacun de ces cas.

d) Les verdicts de Localex

Pour toutes les formes d'adéquation, nous avons besoin de savoir si la décision de LOCALEX est justifiée. Quand c'est le cas, on dira que cette décision est *heureuse*. Dans le cas contraire, on parlera de décision *malheureuse*. Un troisième cas de figure est possible, il s'agira de la situation où même l'expert humain n'est pas en mesure de prendre une décision. C'est notamment le cas de l'ambiguïté effective. On parlera alors de cas *indécidable*. Illustrons ces trois cas:

(1) Heureuses

En admettant que la base contienne la collocation *droit accordé* comme un indice du sens (1) du mot *droit*, et que LOCALEX indique que dans l'exemple suivant, c'est bien ce sens qui est attesté, l'expert humain, qui jugera ensuite le comportement de LOCALEX, pourra dire que le logiciel est parvenu à une décision heureuse.

- le même droit est accordé aux étudiants, aux coopérants ou aux stagiaires du Québec qui séjournent au lieu du domicile de cet étudiant, coopérant ou stagiaire.

(2) Malheureuses

Supposons que l'on soit intéressé par la signification 1 du mot *condition* pour lequel le dictionnaire de la base dispose des deux significations suivantes (Tableau 21), et que par ailleurs on associe à la signification (1) la collocation *fixer condition* et à la signification (2) la collocation *conditions atmosphériques*, on peut constater que, dans l'exemple suivant, LOCALEX se laissera tromper par la présence de l'item *fixer*, si c'est la signification (1) qu'il cherche en premier. Il décrètera que c'est la signification (1) qui est

attestée parce qu'il n'aura pas vu que l'item *atmosphériques* est présent (il ne l'aura pas vu parce qu'il a été conçu pour arrêter l'investigation dès qu'il rencontre la première des collocations associées à la signification en question : l'item *atmosphériques* ne sera décelé que s'il y a échec de la première signification). L'expert humain dira que c'est une décision malheureuse.

Condition (nom)	
Signification 1	Signification 2
État dont l'existence est indispensable pour qu'un autre état existe. <u>Exemples :</u> Remplir les conditions exigées. C'est une condition nécessaire. Clause, stipulation.	Ensemble de faits dont dépend quelque chose.

Tableau 21 Significations du mot condition.

- Les phares blancs prescrits au présent chapitre doivent être solidement fixés au véhicule et ajustés de façon à donner, dans des conditions atmosphériques normales et sur une route horizontale, un éclairage permettant au conducteur du véhicule de distinguer une personne ou un objet à une distance de 150 mètres.

(3) Indécidables

Les cas indécidables correspondent à des contextes pauvres ou polyvalents. Dans l'exemple suivant, la collocation *se soumettre à un examen* n'est pas exploitable.

- . Toutefois, la Société peut exiger que ce titulaire se soumette à un *examen* lorsque le permis échangé est un permis autorisant la conduite d'un véhicule de commerce, d'un véhicule d'urgence, d'un taxi, d'un autobus ou d'un minibus.

Que LOCALEX aboutisse à une signification ou à une autre (sur la base d'une collocation donnée), il aura tort, car ici, l'expert humain n'a rien, du strict point de vue des ressources linguistiques (c'est-à-dire en ne faisant pas intervenir les connaissances du monde), qui puisse orienter sa décision vers une signification ou une autre.

Le Tableau 22 récapitule les différents paramètres à observer.

Verdicts de LOCALEX	Adéquation de la Forme	Adéquation de la catég. / lemme	Adéquation de la collocation	Évaluation de l'expert humain	
				Décision	Commentaire
Noir sur bleu	Non	—	—	Décision Heureuse	La forme rencontrée ne constitue pas une occurrence du mot recherché. Affichage texte normal
				Décision Malheureuse	La forme rencontrée constitue une occurrence du mot recherché, mais LOCALEX ne l'a pas relevée.
Bleu sur blanc	Oui	Non	—	Décision Heureuse	La forme rencontrée constitue une occurrence du mot recherché Et le marquage est correct
				Décision Malheureuse	La forme rencontrée constitue une occurrence du mot recherché Mais il y a erreur de marquage
Noir sur blanc	Oui	Oui	Non	Décision Heureuse	La forme rencontrée constitue une occurrence du mot recherché, et le marquage est correct. Le contexte ne contient aucune collocation favorable à l'acception recherchée. Par ailleurs, pas d'autres indices.
				Décision Malheureuse	La forme rencontrée constitue une occurrence du mot recherché et le marquage est correct. (a) carence de la base, (b) carence du contexte, (c) algorithme, (d) limites de la stratégie.
				Indécidable	La forme rencontrée constitue une occurrence du mot recherché et le marquage est correct; Ambiguïté effective
Rouge Sur blanc	Oui	Oui	Oui	Décision Heureuse	La forme rencontrée constitue une occurrence du mot recherché, et le marquage est correct - Adéquation continue - Adéquation discontinue - pseudo-collocation - La collocation identifiée est pertinente. L'analyse linguistique le confirme.
				Décision malheureuse	- Adéquation continue - Adéquation discontinue - pseudo-collocation - La colloc. identifiée est non pertinente. - conflit

Tableau 22 Les différents paramètres à observer

Légende du Tableau 22

"—" = non applicable.

Remarquons que LOCALEX n'est pas conscient de la continuité ou de la discontinuité des collocations repérées. Cela est cependant facile à implémenter. En revanche, seul l'expert humain est en mesure de constater qu'il y a pseudo-collocation.

C. TEST DE LA VALIDITÉ DE LA DÉMARCHE

Trois points seront discutés dans cette section: le corpus et la base des connaissances, les réponses de LOCALEX et enfin l'évaluation des réponses de LOCALEX.

1. Corpus et base des connaissances

a) Le corpus

Le texte I est un extrait du Code de la Sécurité Routière du Québec⁴⁸. Cet extrait est de 78 pages comprenant 7 titres chacun à son tour composé de plusieurs chapitres.

b) Choix des mots objet de la recherche

Nous avons parcouru le texte afin de repérer les mots susceptibles de répondre aux objectifs de notre test. Cette exploration a été effectuée à l'aide de MARQUEUR (fenêtre de révision) mais elle aurait pu être faite autrement.

Les mots candidats au test ont été retenus en fonction des critères suivants, appliqués de manière hiérarchique:

⁴⁸ Nous avons utilisé une version électronique du Code de la Sécurité Routière. On peut se procurer la disquette chez Les Publications du Québec. Celle-ci porte les références suivantes: Titre: C-24, Code de la Sécurité Routière Version: 8 mai 1996 Fichier: C24_2.wp5 Format: Wordperfect 5.1.

- 1- Le mot-forme doit être polysémique⁴⁹. Par exemple, les mots *droit* et *véhicule* ont chacun au moins deux significations différentes. A ce stade, ils sont tous les deux candidats au test.
- 2- La polysémie en question doit être associée au même lemme-catégorie. Pour rester avec le même exemple, le mot-forme *droit* en tant que *droit-nom* a au moins deux significations. Il reste candidat au test. De son côté, le mot-forme *véhicule* a en tant que *véhicule-nom* au moins deux significations. Il reste également candidat. Cette contrainte a pour objectif de permettre de tester le système sur des cas d'ambiguïté pour le traitement desquels on doit recourir aux collocations plutôt qu'aux marques du lemme et de la catégorie, qui sont de moindre intérêt puisque l'apport vient d'une intervention interactive.
- 3- Au moins deux acceptions doivent être attestées dans le texte. Pour le mot *droit*, la vérification a permis de confirmer qu'au moins deux acceptions de *droit-nom* sont attestées dans le texte. Le mot est retenu. Par contre, le mot *véhicule* n'a qu'une signification attestée. Il ne doit pas être retenu.

Remarquons qu'en observant ces critères (en particulier le deuxième) dans la sélection des mots à examiner, nous nous donnions une chance d'effectuer le test sur des mots dont la polysémie est indépendante de la catégorie. Mais nous risquions de ne pas avoir de mots dont la polysémie est tributaire de la catégorie. Fort heureusement, dans le lot des mots retenus, certains mots sont polycatégoriels (avec des significations spécifiques à la catégorie) et de plus, pour au moins une catégorie donnée, ils sont polysémiques. Ce qui nous permettra de vérifier le pouvoir de désambiguïsation du système aussi bien à un niveau où la catégorie intervient (apport de MARQUEUR) qu'à un niveau où elle n'intervient pas (apport de l'exploitation des collocations).

Le résultat de cette sélection est un ensemble de 11 mots (voir le Tableau 24)

⁴⁹ Nous rappelons que le phénomène de l'homonymie a été fondu dans la polysémie.

c) Les acceptions

Les différentes acceptions ont été prises dans le dictionnaire Le Petit Robert. Pour simplifier la démonstration, nous avons retenu les significations les plus pertinentes à notre propos.

d) Les collocations

Les collocations ont été extraites à l'aide du logiciel COLLOCATEUR; certaines collocations ont été repérées manuellement et ajoutées à la base. Le Tableau 23 montre les collocations retenues pour la signification 1 du mot *droit*. Celles ajoutées manuellement sont soulignées.

Collocation
<u>D: de propriétaire</u>
D: de jouir
D: accordé
D: mettre en circulation
D: de circuler
D: de propriété
G: cession du
<u>G: exercice de</u>
D: de reprise
D: de obtenir
G: suspension du
<u>G: suspendre</u>
D: de personne
D: et obligation
<u>G: accorder</u>
<u>D: suspendu</u>

Tableau 23 Les collocations extraites automatiquement ou manuellement.

Les résultats du test devant confirmer (ou infirmer) la validité de la démarche, nous avons écarté la question de l'exhaustivité de la base en retenant dans la base toutes les collocations présentes dans le texte.

2. Les réponses de Localex

Nous renvoyons le lecteur au relevé, placé en annexe, des différentes occurrences et de leurs contextes respectifs.

Les mots retenus pour le test sont de fréquence variable dans le texte. Le Tableau 24 présente les différents mots retenus et leur fréquence respective. Les données sont triées selon l'ordre décroissant de leur fréquence.

Mot	Nombre d'occurrences
Droit(s), (e)	74
Condition(s)	67
Examen(s)	39
Remettre	34
Objet	23
Application	19
Avis	18
Remise	15
Pièce	11
Charge	6
Porter	4

Tableau 24 les différents mots retenus et leur fréquence (Test 1)

Remarquons que les formes du pluriel ont été comptabilisées au même titre que les formes du singulier. Dans les exemples que nous avons relevés, la marque du pluriel n'apportait aucune différence sémantique. Les deux formes sont du reste reliées au même lemme.

Seuls les mots *droit* et *charge* sont attestés comme polycatégoriels (le mot *droit* se présente dans le corpus comme nom et comme adjectif; le mot *charge* se présente comme nom et comme membre d'une locution conjonctive : *à charge de*). Le mot *remise* est polycatégoriel dans l'absolu, mais il n'est attesté que sous la forme d'un nom. Dans notre test, nous n'avons

traité que les cas de *droit-nom* et de *charge-nom*⁵⁰. Notons par ailleurs que chaque mot-lemme traité a plusieurs significations. Nous n'avons travaillé qu'avec une seule signification pour chaque mot. Bien entendu, le test aurait pu être élargi en traitant toutes les catégories et toutes les significations, mais nous avons considéré que les résultats obtenus sont suffisamment significatifs. Le Tableau 25 relate les différentes significations de chaque mot, et sur une trame de gris, les significations qui ont fait l'objet du test.

⁵⁰ Si LOCALEX rejette les occurrences de ce mot dont la catégorie est adjectif, cela est suffisant pour conclure que le programme est en mesure d'assurer une levée d'ambiguïté fondée sur une divergence de catégorie grammaticale.

Mot	Catég.	Signification 1	Signification 2	Signification 3
Droit	Nom	Ce qui est permis en conformité à une règle morale, sociale / précise. possibilité, pouvoir, qualité, autorisation, permission. Faculté, habileté, prérogative, privilège.	Somme d'argent, redevance qu'une personne, une collectivité est en mesure d'exiger de quelqu'un. → accise, contribution, vx: imposition, impôt, redevance, taxe. Exemple: acquitter un droit.	Ensemble de règles qui régissent les rapports des hommes entre eux. → légalité, légitimité, justice, morale. → juridique.
Condition	Nom	État dont l'existence indispensable pour qu'un autre état existe. Exemples: remplir les conditions exigées. C'est une condition nécessaire. Clause, stipulation.	Ensemble de faits dont dépend qqch.	
Examen	Nom	Action de considérer, d'observer avec attention. Considération, étude, inspection, investigation, observation, recherche, analyse etc... Spécialement: examen médical.	Epreuve destinée à déterminer l'aptitude d'un candidat... Examen et concours. Examen écrit, oral. Examen partiel. Composition, interrogation.	
Remettre	Verbe	61 1- Mettre de nouveau	Mettre en la possession ou au pouvoir de qqn.	
Objet	Nom	CONCRET: Chose solide ayant unité et indépendance... Chose; Familier: bidule, machin, truc. Instrument, outil. Affaire, article. Bagatelle, brouille. Colifichet.	ABSTRAIT: Matière sur laquelle porte un droit, une obligation, un contrat. Exemple: l'objet d'un litige, d'un procès.	
Application	Nom	Action de mettre une chose sur une autre de manière qu'elle la recouvre... Placage: l'application d'un enduit sur un mur. Pose: application de ventouses.	Action de faire porter sur quelque chose. - Mise en pratique: mettre en application une loi.	

Mot	Catég.	Signification 1	Signification 2	Signification 3
Avis	Nom	Ce que l'on pense, ce que l'on exprime sur un sujet. Jugement, estimation, opinion, pensée, point de vue, sentiment.	Ce que l'on porte à la connaissance de qqn. Annonce, communication, information, message, note, notification, nouvelle, renseignement. Préavis. Indication, ordre.	
Remise	Nom	Mettre de nouveau Faire passer dans un autre état ou un état antérieur. Exemple: remettre en circulation.	Mettre en la possession ou au pouvoir de qqn. Exemple: remettre un papier au destinataire.	
Pièce	Nom	Tout écrit servant à établir un droit, à faire la preuve d'un fait. Exemples: pièces justificatives. Pièces d'identité.	Chacun des éléments dont l'assemblage forme un tout organisé. Exemples: pièces d'une machine.	
Charge	Nom	Ce que porte ou peut porter une personne, un animal, un véhicule.. Fardeau, poids	Fonction dont qqn a tout le soin; responsabilité publique. Dignité, emploi, fonction, ministère, office, place, poste.	Ce qui met dans la nécessité de faire des frais, des dépenses... A charge: être à la charge de qqn., vivre à ses dépens.
Porter	Nom	avoir sur soi Porter une bague, un insigne.	Se porter comme Se porter garant.	

Tableau 25 Mots recherchés dans le corpus et leurs significations respectives.

Le Tableau 26 présente les résultats du test.

Mot	-A- Occurrences écartées pour non-adéquation de catégorie- lemme (Bleu)	-B- Occurrences écartées pour non-adéquation de sens (Noir)	-C- Occurrences retenues pour adéquation de la catégorie-lemme et du sens (Rouge)	Total
Droit(s), (e)	6	9	59	74
Condition(s)	0	3	64	67
Examen(s)	0	12	27	39
Remettre	0	14	20	34
Objet	0	6	17	23
Application	0	5	14	19
Avis	0	3	15	18
Remise	0	5	10	15
Pièce	0	3	8	11
Charge	1	3	2	6
Porter	0	1	3	4
Total	7	64	239	310

Tableau 26 Les réponses de LOCALEX.

A ce stade, on peut déjà relever le fait suivant: la recherche a été lancée pour la signification la plus répandue dans le texte (ce dernier fait, nous l'avons vérifié au préalable, manuellement) et LOCALEX confirme cette dominance (voir colonne -C- du Tableau 26). Ce qui peut laisser croire que, jugé sommairement, LOCALEX est satisfaisant. Si l'on devait se contenter de la requête: "Quelle est la signification dominante dans tel texte pour tel mot?", LOCALEX semble s'y prêter. Mais ce raisonnement ne tient pas, car jusqu'à présent, nous n'avons pas encore évalué individuellement les réponses de LOCALEX. Nous ne savons pas encore si les occurrences repérées par le programme comme satisfaisantes le sont également à nos yeux.

Pour évaluer ces résultats, il faudrait par conséquent regarder les réponses de LOCALEX pour chaque occurrence de chaque unité examinée. Seul ce travail d'observation méticuleuse permettra de rendre compte du comportement du programme et d'apprécier sa capacité de désambiguïsation. Le jugement de l'expert humain est ici indispensable, d'autant plus qu'il y a des cas dont on dira avec raison qu'ils sont indécidables même pour ledit expert humain. Il y aura donc trois cas à distinguer:

- Décision heureuse
- Décision malheureuse
- Cas indécidable.

3. Évaluation des réponses de Localex

Nous en venons maintenant à la question fondamentale pour le test. Localex a-t-il agi convenablement? Nous examinerons les différentes réponses à partir du tableau suivant :

Mot	-A- Occurrences écartées pour non- adéquation de lemme-catégorie (Bleu)				-B- Occurrences écartées pour inadéquation sémantique (Noir)				-C- Occurrences retenues pour adéquation du lemme-catégorie et du sens (Rouge)				Tot
	H	M	I	S/T	H	M	I	S/T	H	M	I	S/T	
	Droit(s), (e)	6	0	0	6	9	0	0	9	59	0	0	
Condition(s)	0	0	0	0	0	0	3	3	64	0	0	64	67
Examen(s)	0	0	0	0	0	0	12	12	26	0	1	27	39
Remettre	0	0	0	0	14	0	0	14	18	2	0	20	34
Objet	0	0	0	0	4	0	2	6	17	0	0	17	23
Application	0	0	0	0	2	1	2	5	14	0	0	14	19
Avis	0	0	0	0	2	0	1	3	15	0	0	15	18
Remise	0	0	0	0	4	0	1	5	10	0	0	10	15
Pièce	0	0	0	0	3	0	0	3	8	0	0	8	11
Charge	1	0	0	1	3	0	0	3	2	0	0	2	6
Porter	0	0	0	0	1	0	0	1	3	0	0	3	4
Total	7	0	0	7	42	1	21	64	236	2	1	239	310

Tableau 27 Évaluation des réponses de Localex

La colonne -A- du Tableau 27 montre la capacité de LOCALEX à rejeter l'occurrence dont la catégorie ou le lemme ne correspondent pas à ce qui est demandé. Parmi les onze mots qui ont servi au test, seuls les deux mots *droit* et *charge* sont polycatégoriels. Le premier est attesté comme nom et comme adjectif. Comme la recherche a été lancée pour le nom, toutes les occurrences de *droit-adjectif* (6 cas) ont été, à juste titre, rejetées. Quant au

deuxième, il est attesté comme nom et comme membre d'une locution prépositionnelle: "à charge". Comme seule la catégorie du nom a fait l'objet de la recherche, le programme a écarté, avec raison, la seule occurrence pourvue d'une catégorie autre que le nom.

Ce premier niveau de désambiguïsation, dont la difficulté a été résolue au niveau du marquage, se présente au niveau de LOCALEX comme un fruit à cueillir. Le programme ne semble, en effet, buter sur aucun écueil relatif à ce type de repérage. L'identification de la catégorie et du lemme a été efficace à 100 %. Aucun cas malheureux, n'a été observé. L'opération étant d'un certain point de vue triviale, le succès devrait être interprété comme une confirmation que le logiciel fonctionne comme prévu. Nous ne pouvons néanmoins nous empêcher d'apprécier aussi bien l'apport de MARQUEUR que la capacité de LOCALEX de tirer parti de cet acquis.

La colonne -B- du même tableau (Tableau 27) montre la capacité de LOCALEX à rejeter les occurrences dont la signification ne correspond pas à ce qui est demandé.

LOCALEX semble commettre très peu d'erreurs de rejet malheureux. Il n'y a qu'une seule occurrence qui a été rejetée à tort. Notons toutefois que cette erreur aurait pu être rattrapée si on avait pris le soin de rajouter la collocation qui lui est reliée, mais nous nous sommes refusé de le faire, car en fait cette occurrence (*application de aa*⁵¹) fait partie d'un type d'énoncés qu'il serait difficile d'exploiter selon notre méthode. Il s'agit de collocations dont l'un des items qui les composent est interprétable en référence à d'autres mots dans le texte. Sont compris dans cette catégorie les déictiques (*ceci, cela...*), les mots assurant une substitution (*premier, second* etc.) et toutes les unités à valeur métalinguistique (*ce mot, ce terme, ce nom* etc.). Nous avons jugé qu'il ne serait pas pertinent de répertorier ce type de collocations parce qu'elles contiennent des éléments vides de sens en eux-mêmes et par conséquent peu inducteurs de sens. Dans notre stratégie, nous privilégions les collocations dont la valeur désambiguïsatrice est donnée d'emblée, sans passer par un calcul supplémentaire.

⁵¹ Il s'agit ici d'une énumération aa, ab,ac, etc.

D'un autre côté, 21 occurrences ont été rejetées par le programme alors que ce sont des cas indécidables. Le programme les a rejetées parce qu'il n'a rencontré dans le contexte aucune collocation connue dans la base. En bonne logique, on pourrait penser que la base aurait dû être enrichie, même *a posteriori*, mais, contrairement à la situation précédente, nous sommes ici en présence de cas où le contexte examiné est difficilement exploitable. Le contexte est soit nul, soit pauvre ou encore polyvalent:

- ◆ Le contexte peut être *nul*, et dans ce cas, le mot ambigu conserve son ambiguïté virtuelle. Le recours à des mots au delà des limites de la phrase pourrait être utile, mais dans la stratégie adoptée, nous avons travaillé avec des collocations syntagmatiques, limitées aux frontières de la phrase.
- ◆ Le contexte peut être *pauvre*. Expressions trop générales, mots vagues...
- ◆ Mais il peut aussi contenir un mot qui serait inducteur d'au moins deux significations conflictuelles. Nous avons appelé ce type d'environnement un *contexte polyvalent*. Le mot *examen* illustre ce phénomène une douzaine de fois.

Exemples: *sans examen, subir un examen, réussir l'examen.*

Cet échec ne doit pas être pris dans l'absolu. Les énoncés ambigus échappent aussi à l'expert humain, à moins que l'on ne fasse appel à un contexte plus large que la phrase, ou à des connaissances du niveau de la macrostructure.

La colonne -C- du même tableau (Tableau 27) montre la capacité de LOCALEX à retenir les occurrences dont la signification correspond à ce qui est demandé.

On voit que le taux de succès est élevé. La capacité de LOCALEX à retenir avec **bonheur** les occurrences répondant totalement aux critères de la requête (lemme, catégorie et signification) s'élève à 236 sur 239 cas, ce qui représente un pourcentage de 98,74%. On note d'autre part que deux occurrences ont été relevées à tort, et une autre retenue alors qu'en

principe, la décision à son égard est difficile à prendre. Les deux cas malheureux sont les suivants:

- II: **INTERDICTION DE REMETTRE** (occurrence 18282) **UN VÉHICULE ROUTIER EN CIRCULATION, SUSPENSION DES PERMIS ET DES LICENCES SECTION I:**

- **INTERDICTION DE REMETTRE** (occurrence 18298) **UN VÉHICULE ROUTIER EN CIRCULATION, SUSPENSION DES PERMIS D'APPRENTIS- CONDUCTEURS, DES PERMIS PROBATOIRES ET DES PERMIS DE CONDUIRE.**

LOCALEX s'est laissé leurrer par la collocation *remettre permis* (il n'a pas pu déceler la présence de la collocation *remettre en circulation* qui se trouve être la bonne parce qu'il a été conçu pour arrêter l'investigation dès qu'il rencontre la première des collocations associées à la signification en question). Or, dans ce contexte, il se trouve que c'est une pseudo-collocation (le mot *suspension* a déstructuré cette collocation).

Une autre stratégie aurait été de recourir à la recherche de collocations non structurées, mais cela n'aurait pas été pertinent pour ces deux cas. A moins qu'on ne retienne le mot *circulation* comme ayant une affinité sémantique avec le verbe *remettre*, ce qui ne semble pas évident (ce type d'affinité doit être perçu dans l'absolu). Il aurait été beaucoup plus efficace de permettre à LOCALEX de tenir compte de la collocation structurée *en circulation*, par exemple, en lui permettant de tenir compte de toutes les collocations en présence et de privilégier la plus proche d'entre elles. On voit, par conséquent, que d'une manière générale la collocation structurée est plus efficace que la collocation non structurée. Le problème qui se pose est que des items impliquant des significations divergentes peuvent se retrouver dans un espace très réduit. Or, l'utilisation des collocations non structurées doit affronter un double problème. Si l'on tient compte de l'ensemble du lexique (en travaillant sur la macrostructure du texte), la conclusion globale risque d'être démentie par un contexte local. Si au contraire, on s'appuie sur une observation du contexte local, la conclusion résultant de la présence d'un item donné risque d'être trop faible (pas assez significative) et par conséquent démentie par la présence d'un autre item tout aussi proche. Il s'ensuit que la collocation non structurée doit être utilisée plutôt dans des stratégies incluant une comparaison des contextes. S'il n'y a pas de

conflit de sens, alors le recours à cette source d'information peut être utile. S'il y a conflit, on doit trouver le moyen de trancher entre les collocations (ou ensembles de collocations) associées aux significations respectives.

Revenons maintenant au cas indécidable:

- 90. Le titulaire d'un permis de conduire valide délivré au Canada, par une autre autorité administrative peut, s'il s'établit au Québec, échanger sans examen ce permis contre un permis de conduire délivré par la Société, sur paiement des droits et des frais fixés par règlement et du montant fixé en vertu de l'article 151 de la Loi sur l'assurance automobile (chapitre A-25).

Ici, le contexte est jugé comme polyvalent, car la préposition *sans* qui accompagne le mot *examen* peut induire des significations conflictuelles; on ne pouvait donc la retenir comme membre d'une collocation. LOCALEX a dû se rabattre sur la collocation dont il dispose, à savoir: *de société*. Mais celle-ci s'est avérée être une pseudo-collocation, à cet endroit du contexte.

Le Tableau 28 présente les mêmes résultats mais en les distribuant autrement, de manière à mettre en évidence les succès et les échecs de LOCALEX, toutes tâches confondues.

Il faudrait tenir compte des collocations du contexte qui sont conflictuelles mais qui n'ont pas été vues par LOCALEX, parce qu'il s'arrête dès qu'il tombe sur la première collocation.

Mot	-A- Heureuses				-B- Malheureuses				-C- Indécidables				TOTAL
	Bleu	Noir	Rouge	S/T	Bleu	Noir	Rouge	S/T	Bleu	Noir	Rouge	S/T	
Droit	6	9	59	74	0	0	0	0	0	0	0	0	74
Condition	0	0	64	64	0	0	0	0	0	3	0	3	67
Examen	0	0	26	26	0	0	0	0	0	12	1	13	39
Remettre	0	14	18	32	0	0	2	2	0	0	0	0	34
Objet	0	4	17	21	0	0	0	0	0	2	0	2	23
Application	0	2	14	16	0	1	0	1	0	2	0	2	19
Avis	0	2	15	17	0	0	0	0	0	1	0	1	18
Remise	0	4	10	14	0	0	0	0	0	1	0	1	15
Pièce	0	3	8	11	0	0	0	0	0	0	0	0	11
Charge	1	3	2	6	0	0	0	0	0	0	0	0	6
Porter	0	1	3	4	0	0	0	0	0	0	0	0	4
Total	7	42	236	285	0	1	2	3	0	21	1	22	310

Tableau 28 Les succès et les échecs de LOCALEX.

La colonne -A- regroupe tous les cas de succès obtenus de LOCALEX dans les différentes tâches d'identification: acceptation ou rejet de la catégorie et du lemme, acceptation ou rejet de la signification rencontrée. Le résultat obtenu pour cette activité est 285 occurrences sur 310, soit un pourcentage de 91,93 %. (le bleu: 7 cas sur 285, soit 2,45 %, le noir: 42 cas sur 285, soit 14,73 %, le rouge: 236 cas sur 285, soit 82,80 %).

La colonne -B- permet de souligner les échecs de LOCALEX pour ces mêmes tâches: 3 sur 310, soit un pourcentage de près de 1 %.

Enfin, la colonne -C- fait état des cas indécidables (contexte nul ou pauvre). Nous en avons compté 22 sur 310, soit un pourcentage de 7.09 %.

Cela dit, il convient de noter que ce taux de succès ne peut être apprécié à sa juste mesure si l'on ne tient pas compte de la procédure qui a mené à ce succès. Est-elle tout simplement fondée? Nous répondrons à cette question dans ce qui suit.

4. Force et limites de la stratégie de désambiguïsation

La force du système réside sans doute dans les règles établies pour la sélection des collocations à emmagasiner. Ceci parce que l'expertise du linguiste qui opère ces sélections peut jouer à différents niveaux: le plus important ici est celui de la contiguïté des items. En effet, nous avons observé que quand les items les plus essentiels de la collocation sont représentés, la contrainte de la contiguïté n'a qu'une force très relative. Les résultats du Tableau 29 sont assez révélateurs à cet égard: un nombre important de cas de discontinuité constituent des cas de bonne décision de la part de LOCALEX. Même les cas de pseudo-collocation ne constituent pas un échec systématique. Bien au contraire. La démarche accuse néanmoins quelques points faibles.

a) *Adéquation continue / discontinue / disloquée*

Il est frappant de constater que dans le lot des occurrences repérées avec plein succès (couleur rouge) et jugées comme des repérages heureux, il y a un nombre important de cas de continuité (92). Pour apprécier ce résultat, il faut se souvenir que lors de l'alimentation de la base en collocations, seuls les items les plus essentiels ont été retenus. Quand on retrouve ces items dans le même ordre et avec la même contiguïté, cela signifie qu'aucun élément ne s'est intercalé entre les items prévus. La sélection opérée s'est révélée pertinente au plus haut point.

Mais il faut noter que même dans le cas où la continuité ne serait pas assurée, nous n'aurions rien à craindre de la discontinuité: elle génère aussi des cas de succès, confirmés par l'analyse linguistique de l'expert. Les résultats obtenus viennent corroborer cette affirmation: 131 cas de discontinuité sur un total de 236 cas de décision heureuse. La discontinuité signifie qu'un élément quelconque s'est glissé entre les items essentiels de la collocation, mais il s'agit d'une intrusion sans conséquence sur le pouvoir de désambiguïsation de la collocation. Ainsi, si la collocation prévue est *droit de propriété* et que le contexte s'avère être *droit de cette propriété*, l'élément *cette* ne nuit nullement à la signification initiale.

Enfin, fait plus intéressant encore, en cas de pseudo-collocation, les résultats sont favorables au succès du repérage: 9 cas de pseudo-collocation heureuse contre 2 cas de pseudo-collocation malheureuse et seulement un cas indécidable.

Mot	Rouge									Tot
	Heureuses			Malheureuses			Indécidables			
	cont	disc	Pseud	cont	disc	Pseud	cont	disc	Pseud	
Droit	27	26	5	0	0	0	0	0	0	58
Condition	31	31	0	0	0	0	0	0	0	62
Examen	18	8	0	0	0	0	0	0	1	27
Remettre	1	17	0	0	0	2	0	0	0	20
Objet	0	17	0	0	0	0	0	0	0	17
Application	4	10	1	0	0	0	0	0	0	15
Avis	3	10	2	0	0	0	0	0	0	15
Remise	4	6	0	0	0	0	0	0	0	10
Pièce	3	3	1	0	0	0	0	0	0	7
Charge	1	1	0	0	0	0	0	0	0	2
Porter	0	3	0	0	0	0	0	0	0	3
Total	92	132	9	0	0	2	0	0	1	236

Tableau 29 Collocations continues, discontinues, pseudo-collocations

Les cas de pseudo-collocation sont en majorité des cas de succès et cela coïncide avec l'observation suivante: quand nous décrivons une collocation comme étant une pseudo-collocation, nous faisons référence à la collocation structurée. Pseudo-collocation signifie ici que les mots de la collocation du contexte ne sont pas unis dans le même syntagme.

b) Carence et surenchère de collocations

Un des écueils que notre stratégie a à affronter est la carence du contexte en termes de collocations. C'est le cas de mots isolés dans le texte. Les titres⁵² par exemple peuvent se présenter comme des mots uniques. Nous avons vu plus haut d'autres formes de carence. Mais cet écueil n'est qu'une facette du problème, car LOCALEX bute tout aussi bien sur une situation où il y a plusieurs collocations en présence, pour peu que celles-ci soient en conflit. Comme notre programme a été conçu pour suspendre la recherche dès qu'une collocation est

repérée, et qu'aucune comparaison n'est faite, une deuxième ou troisième signification concurrente peut être *passée sous silence*. Une solution sera envisagée dans la section des solutions proposées à la page 116, mais dans l'état actuel de l'algorithme, il serait intéressant de remarquer que dans l'alimentation de la base, on pourrait placer les significations selon l'ordre de la fréquence d'usage, de manière à ce que LOCALEX examine toujours en premier les collocations rattachées à ces significations favorites.

Autant le problème se pose avec un texte pauvre, autant l'équilibre se fait avec un contexte riche en collocations. Cette situation n'est pas rare. LOCALEX, qui n'examine qu'une collocation à la fois, et qui arrête ses investigations une fois qu'une concordance est atteinte, n'est pas conscient de cet embarras du choix. Dans une perspective d'exploitation des collocations structurées, en dehors du cas de la pseudo-collocation, il n'est nul besoin d'enfoncer d'autres portes ouvertes. Une seule collocation peut suffire. Mais dans une perspective d'utilisation de collocations non structurées, il serait important de pouvoir multiplier les preuves. Ainsi, le processus de désambiguïsation a plus de chance d'aboutir s'il est montré que le mot *droit*, par exemple, coexiste avec le mot *jurisprudence*, en plus du mot *obligations*.

D. TEST DE LA REPRÉSENTATIVITÉ DE LA BASE

Comme nous l'avons vu au chapitre précédent, les trois logiciels fonctionnent comme un ensemble. D'autre part, les programmes ont été conçus de manière à permettre la validation des données qu'ils manipulent. Ce sont autant de caractéristiques qui facilitent la mise à l'épreuve.

1. Présentation du test

Ce second test a, rappelons-le, comme objectif principal de montrer si les collocations d'un texte sont rentables dans la désambiguïsation lexicale appliquée à un autre

⁵² Quand un fichier se présente en format texte, le titre se confond avec les paragraphes voisins. Quand il est entouré de ponctuation forte, il se retrouve isolé. Il serait intéressant d'étudier les moyens de reconnaître automatiquement un titre dans différents formats de texte.

texte dont on n'a pas extrait les collocations. Accessoirement, nous aimerions évaluer l'importance de cette rentabilité, les ajustements qu'il faudrait apporter, et l'intérêt qu'on aurait à alimenter davantage la base des collocations.

Nous avons retenu un texte du même domaine que le premier. Il s'agit d'un texte de droit, en l'occurrence du Code Civil du Québec⁵³ (chapitre de la famille). Ce texte comprend 60 pages⁵⁴. Il a été marqué⁵⁵ à l'aide du logiciel MARQUEUR, et comme nous l'avons déjà prévu, nous n'en avons extrait aucune collocation. Il fera l'objet de recherches lexicales mais en tenant compte uniquement des collocations provenant du premier corpus (le Code de la Sécurité Routière).

Pour ce second test, nous n'avons retenu que deux mots. La raison de cela est que ce sont les seuls mots, parmi les 11 traités dans le premier test, qui disposent d'un nombre significatif de collocations. Nous présentons dans le Tableau 30 les différentes significations du mot *droit* (celle utilisée pour la recherche lexicale est imprimée avec une trame en gris). Le Tableau 31 montre les collocations extraites du premier corpus et utilisés telles quelles dans le présent corpus. Le même type d'informations est donné pour le mot *condition*, respectivement, dans le Tableau 32 et dans le Tableau 33.

⁵³ Nous avons utilisé une version électronique du Code Civil du Québec. On peut se procurer la disquette chez Les Publications du Québec. Cette dernière porte les références suivantes: Titre: C-64, Code Civil du Québec, Version: 22 juin 1995, Fichier: 91- F064.WP, FORMAT: Wordperfect 5.1.

⁵⁴ Mise en page au format de traitement de texte commun, simple interligne.

⁵⁵ Nous avons dû recourir à quelques arrangements pour accomplir ce marquage. En particulier, nous avons considéré certains mots fréquents comme univoques. Exemple: les formes *le* et *les* ont été marquées uniquement comme des déterminants. Cela a permis d'alléger la tâche du marquage.

Droit (nom)		
Signification 1	Signification 2	Signification 3
<p>Ce qui est permis en conformité à une règle morale, sociale / précise.</p> <p>possibilité, pouvoir, qualité, autorisation</p> <p>permission.</p> <p>Faculté, habileté, prérogative, privilège.</p>	<p>Somme d'argent, redevance qu'une personne, une collectivité est en mesure d'exiger de quelqu'un.</p> <p>→ accise, contribution, vx: imposition, impôt, redevance, taxe. Exemple: acquitter un droit.</p>	<p>Ensemble de règles qui régissent les rapports des hommes entre eux.</p> <p>→ légalité, légitimité, justice, morale.</p> <p>→ juridique.</p>

Tableau 30 Significations du mot "droit".

Collocations emmagasinées

Signification 1	Signification 2	Signification 3
<p>D: de propriétaire</p> <p>D: de jouir</p> <p>D: accordé</p> <p>D: mettre en circulation</p> <p>D: de circuler</p> <p>D: de propriété</p> <p>G: cession du</p> <p>G: exercice de</p> <p>D: de reprise</p> <p>D: de obtenir</p> <p>G: suspension du</p> <p>G: suspendre</p> <p>D: de personne</p> <p>D: et obligation</p> <p>G: accorder</p> <p>D: suspendu</p>	<p>D: fixé</p> <p>G: payer</p> <p>G: acquitter</p> <p>G: paiement des</p>	<p>D: commun</p> <p>D: romain</p> <p>D: écrit</p> <p>D: qui de</p>

Tableau 31 Collocations emmagasinées pour la signification 1 du mot "droit".

Condition (nom)	
Signification 1	Signification 2
État dont l'existence indispensable pour qu'un autre état existe. <u>Exemples :</u> Remplir les conditions exigées. C'est une condition nécessaire. Clause, stipulation.	Ensemble de faits dont dépend quelque chose

Tableau 32 Significations du mot "condition".

Collocations emmagasinées

Signification 1	Signification 2
G: assortir de	D: atmosphérique
G: satisfaire à	
G: satisfaire aux	
G: assorti de	
D: établir	
D: et formalités	
G: et modalités	
D: attacher	
D: de obtention	
D: conformer aux	
D: au permis	
G: supprimer	
D: de délivrance	
G: respecter	
D: prévu par règlement	
D: attaché	
D: établi	
G: selon	
G: à la	
G: fixer	
D: fixé	
D: fixer	
G: remplir	
G: conformément aux	

Tableau 33 Collocations relatives à la signification 1 du mot "condition".

2. Résultats et commentaires

LOCALEX a repéré un nombre important d'occurrences pour chacun des deux mots (voir Tableau 34) . Le test peut donc être accompli.

Mot	Nombre d'occurrences
Droit(s), (e)	85
Condition(s)	13

Tableau 34 les différents mots retenus et leur fréquence (Test 2)

Une première constatation: il y a 25 occurrences affichées en rouge par LOCALEX et 60 occurrences en noir. En attendant de vérifier si le système a bien agi, accusons ce premier constat: LOCALEX rencontre dans un texte *inédit* des collocations *venant d'ailleurs*, et s'en sert.

Mot	-A- Occurrences écartées pour non-adéquation de catégorie- lemme (Bleu)	-B- Occurrences écartées pour non- adéquation de sens (Noir)	-C- Occurrences retenues pour adéquation de la catégorie-lemme et du sens (Rouge)	Total
Droit(s), (e)	0	60	25	85
Condition(s)	0	11	2	13
Total	0	71	27	98

Tableau 35 Les réponses de LOCALEX (test 2)

Maintenant, si l'on regarde de près le comportement de LOCALEX (voir Tableau 36), on s'aperçoit que dans les occurrences signalées en noir (occurrences écartées pour non-adéquation de sens), LOCALEX a eu tort de décider un rejet dans la quasi-totalité des cas. (59 sur 60 sont des cas malheureux). La raison qui explique ce résultat est simple: c'est la carence de la base des collocations. LOCALEX tombe sur autre chose que ce qu'il connaît déjà.

Mais ce résultat n'est pas décourageant, car dans les cas signalés en rouge, tous sont des cas de succès. Leur pourcentage est important (27 sur 85), compte tenu du nombre modeste des collocations enregistrées dans la base (voir Tableau 36).

Mot	-A- Occurrences écartées pour non- adéquation de lemme-catégorie (Bleu)				-B- Occurrences écartées pour inadéquation sémantique (Noir)				-C- Occurrences retenues pour adéquation du lemme-catégorie et du sens (Rouge)				Tot
	H	M	I	S/T	H	M	I	S/T	H	M	I	S/T	
	Droit(s), (e)	0	0	0	0	1	59	0	60	25	0	0	
Condition(s)	0	0	0	0	0	11	0	11	2	0	0	2	13
Total	0	0	0	0	1	70	0	71	27	0	0	27	98

Tableau 36 Évaluation des réponses de LOCALEX (Test 2)

Cependant, même si l'expert humain confirme pour tous ces cas que LOCALEX ne s'est pas trompé dans sa décision, nous ne pouvons prendre pour acquis que ce succès est totalement mérité. Ainsi que nous l'avons fait dans le premier test, il faudrait regarder si les collocations repérées dans le contexte (et donc prises pour des preuves déterminantes) sont bien légitimement rentables (elles ne le seraient pas si les différents items recherchés ne forment plus une unité). Autrement dit, il faudrait s'assurer que nous n'avons pas compté sur des pseudo-collocations⁵⁶ pour prendre une décision sur la signification attestée en contexte.

Les résultats (voir Tableau 37) montrent que 7 occurrences ont été résolues grâce à des collocations continues. 10 occurrences ont été résolues sur la base de collocations discontinues. Cependant que 10 autres ont bénéficié —abusivement— de la présence de pseudo-collocations.

Ainsi que nous l'avons fait remarquer précédemment, le phénomène de la pseudo-collocation est un phénomène malheureux pour notre stratégie, un talon d'Achille dans notre arsenal, mais fort heureusement, le danger est atténué par un autre phénomène à l'œuvre dans le texte. Il se trouve que les mots repérés en contexte, s'ils ne sont qu'une fausse réplique de la collocation enregistrée dans la base, peuvent avoir cependant des affinités sémantiques. L'on se retrouve alors face à une collocation de type non structuré. Dans l'exemple suivant, la collocation emmagasinée est *droits et obligations*. Dans le contexte, les items *droits* et

⁵⁶ Auquel cas, ce serait une sorte de collocation labyrinthe.

obligations apparaissent comme désunis; nous ne pouvons compter sur le caractère structurel de la séquence, mais nous pouvons par contre, et c'est une chance assez fréquente, tenir compte de leur cohésion lexicale (relation sémantique) pour lever l'ambiguïté de l'un par l'effet produit par l'autre.

- le tuteur, s'il en existe, perd ses droits et est libéré de ses devoirs à l'endroit de l'adopté, sauf l'obligation de rendre compte.

Ce phénomène ne s'est pas présenté dans notre corpus d'une manière aussi tranchée que dans l'exemple ci-dessus, mais force est de constater que ce phénomène relève de relations sémantiques peu systématisables. A défaut d'explicitier ces relations complexes, nous profitons du fait que nombre de collocations structurées (vues du point de vue syntaxique et lexical) constituent en même temps des collocations non structurées (vues du point de vue sémantique et cognitif). La conséquence de cela est qu'il faut privilégier, dans l'alimentation de la base, les collocations qui ont ce double statut.

Mot	Rouge									
	Heureuses			Malheureuses			Indécidables			Tot
	cont	Disc	Pseud	cont	disc	Pseud	Cont	disc	Pseud	
Droit	7	9	9	0	0	0	0	0	0	25
Condition	0	1	1	0	0	0	0	0	0	2
Total	7	10	10	0	0	0	0	0	0	27

Tableau 37 Les succès et les échecs de LOCALEX (Test 2)

Ajoutons cette dernière remarque. Quand nous avons fait l'extraction des collocations du premier texte, nous nous sommes contenté de retenir uniquement les collocations qui sont attestées dans le corpus. Nous avons pourtant été tenté d'enrichir la base par des collocations qu'il est facile de dériver sur la base de celles déjà acquises. Par exemple, en utilisant la collocation verbale *exercer un droit*, on pourrait obtenir *exercice du droit*. Si nous avons appliqué ce principe, qui ne nuit pas à la démarche générale, le taux de succès de LOCALEX aurait été plus élevé, car effectivement, nous avons remarqué qu'un nombre important de cas pouvaient être résolus par le recours à une collocation dérivée. L'idée du recours à un hyperlemme (englobant nom et verbe dans une même entité) aurait été d'un

secours certain, mais nous l'avions écartée pour des raisons de cohérence du système (voir chapitre I).

E. SOLUTIONS PROPOSÉES

Parmi les problèmes rencontrés, certains sont difficiles à résoudre. Nous pensons en particulier à celui de la pseudo-collocation. Au stade actuel, LOCALEX n'est pas conscient des différentes réalisations dans le contexte des collocations (continue, discontinue, pseudo); en lui faisant reconnaître la première catégorie, il serait ensuite envisageable de recourir à d'autres sources d'informations, exclusivement dans les cas où il n'y a pas de continuité.

Précisément, pour aller dans cette voie, nous pensons qu'il reste possible de tenir compte du paramètre de la distance. En vérité, on ne sait pas trop à quelle distance il faut limiter la longueur d'une collocation. Dans le domaine littéraire, les phrases fleuves comme celles de Proust sont légion. Inversement, une phrase, comme une collocation, peut *tourner court* à tout moment. Cette voie semble fermée, mais ce ne sera pas le cas si on recourt à ce paramètre uniquement dans des situations de conflit. La comparaison pourrait être décisive.

Un autre niveau d'intervention serait d'enrichir les collocations en tirant parti d'une information grammaticale déjà disponible. Il s'agit de la catégorie grammaticale que MARQUEUR a rendue disponible. Ce genre de solutions va à l'encontre de la philosophie que nous avons défendue: travailler à partir du littéral. Si malgré tout, nous proposons cette solution, c'est parce que d'une part, c'est une information déjà enregistrée, et que d'autre part, il est possible de suivre une notation aussi simple que ce qui a été fait jusqu'à présent. Voici pour le mot *droit*, un exemple de collocation enrichie:

D:préposition propriété

De cette manière, les contextes suivants peuvent être considérés comme des corrélats valides pour cette collocation hybride:

- *Droit à la propriété*

- *Droit de propriété*

- *Droit sur la propriété*

Naturellement, l'algorithme devrait être modifié en conséquence.

F. CONCLUSION

La stratégie de désambiguïsation, fondée sur l'utilisation des collocations (considérées elles-mêmes comme une marque de cohésion lexicale) a été mise à l'épreuve. Nous voulions savoir si elle était efficace et en même temps évaluer les raisons du succès comme celles de l'échec. Nous pouvons maintenant conclure que la démarche suivie est concluante à plus d'un titre. Si les résultats du premier test peuvent être nuancés, en particulier pour la raison que les collocations présentes dans le texte ont été toutes saisies préalablement, c'est en même temps un argument pour dire que plus la base est étoffée, plus la chance de réussir l'opération de la levée d'ambiguïté est grande. D'ailleurs, le second test a été concluant dans ce sens. Une quantité négligeable de collocations a été suffisante pour lever le voile sur un nombre important d'occurrences ambiguës.

CONCLUSION GÉNÉRALE

4. CONCLUSION GÉNÉRALE

Nous avons projeté dans notre recherche d'utiliser les ressources de la cohésion lexicale dans un processus de désambiguïsation automatique, et nous avons été amené à privilégier le concept de collocation. Un long chemin a donc été parcouru depuis notre projet initial.

Ayant considéré qu'une part de l'ambiguïté lexicale peut être résolue en déterminant la catégorie grammaticale et le lemme du mot à traiter, nous avons jugé important de développer un programme qui effectue cette tâche. Nous ne pouvions cependant le rendre complètement automatique. Un travail complémentaire est nécessaire si l'on veut augmenter considérablement la part d'automatisation du programme⁵⁷.

Il reste que MARQUEUR, qui est d'ailleurs ouvert à l'attribution de toutes sortes de propriétés, est un programme utile. Il permet tout simplement d'annoter les mots d'un texte donné. Pour les besoins de notre thèse, nous l'avons utilisé pour marquer deux textes qui ont fait l'objet d'expérimentation sur le système de désambiguïsation.

Dans notre stratégie, nous avons planifié d'emmagasiner un nombre important de collocations. Chaque ensemble de collocations devant être rattachées à une signification donnée. Dans notre conception, les collocations doivent être extraites de textes attestés, si possible de textes reconnus socialement. Et c'est pourquoi, nous avons trouvé que la réalisation d'un logiciel capable de recueillir ces collocations, sur la base de règles simples, était d'un grand intérêt. Le logiciel COLLOCATEUR a répondu à ce besoin.

Enfin, il fallait mettre en œuvre notre stratégie. Nous avons donc conçu et réalisé un programme approprié à cette tâche. Il s'agit de LOCALEX. Celui-ci avait pour mandat de permettre de localiser dans un texte donné les occurrences d'un mot donné, avec comme contraintes, que l'occurrence soit pourvue de la catégorie grammaticale, du lemme et de la

⁵⁷ Dans nos activités d'enseignant à l'Université Ibn Tofail, au Maroc, nous avons dirigé des travaux d'étudiants dans ce sens.

signification désignés par l'utilisateur. C'est là une façon parmi d'autres de procéder à la levée d'ambiguïté.

Comme il ne suffit pas de voir fonctionner un logiciel, nous avons défini deux types de test. Le premier devait permettre de voir si l'exploitation de la cohésion lexicale (de la manière dont nous l'avons conçue) était efficace. Il nous fallait à cette étape supposer que la question de l'alimentation de la base ne se posait pas. C'est pourquoi, nous avons dû, pour accomplir cette première mise à l'épreuve, prendre soin de fournir au préalable à LOCALEX, toutes les collocations qui se présentaient dans l'environnement des occurrences visées par la recherche. Le second type de test avait pour objectif de montrer, à supposer que le premier test ait été concluant, dans quelle mesure on pouvait compter sur des collocations présentes dans la base (et enregistrées selon le protocole que nous avons conçu) pour lever automatiquement l'ambiguïté de mots repérés dans des textes inédits (nous ignorons tout de leur contexte).

Nous avons vu dans les résultats du premier test que le taux de succès était très élevé. Il y a, bien entendu, plusieurs niveaux d'évaluation et nous avons eu l'occasion d'en apprécier les détails. Nous pouvons cependant rappeler que le taux de réussite de LOCALEX, si l'on confond tous les types de repérage (rejet et acceptation du lemme-catégorie, rejet et acceptation de la signification) s'est élevé à 91,93 %.

Les performances du second test sont moins glorieuses: 27,55 % si l'on confond toujours tous les types de repérage. Mais ces résultats sont cependant loin d'être décourageants. Tout au contraire, si l'on se rappelle que dans cette seconde épreuve, nous avons exploité pour chacun des deux mots soumis au test, les collocations extraites d'un corpus limité, que par ailleurs nous n'avons pas bénéficié des collocations dérivées, et qu'enfin, le résultat de cette évaluation ne fait pas entrer en compte l'apport des collocations non structurées, si l'on se rappelle tout cela, on conclura que notre hypothèse intuitive initiale se confirme. Selon cette hypothèse, nous avons considéré qu'un type de discours donné est caractérisé par une distribution homogène du lexique au point qu'un corpus témoin fournit une base de collocations suffisante pour traiter d'autres corpus du même type. L'hypothèse elle-

même croise une des questions qui reste à résoudre mais dont la littérature commence à faire état: il s'agit de déterminer les rapports entre la cohésion des textes et leur cohérence⁵⁸.

Il faut d'autre part observer que les résultats du second test ont révélé que LOCALEX a globalement bien agi dans ses décisions d'acceptation de la signification recherchée. Même si certaines de ces décisions sont fondées sur des pseudo-collocations, il reste que LOCALEX a pris les décisions qu'il faut.

Au terme de ce parcours, il nous reste à mentionner que cette étude, qui aura permis d'apporter un éclairage sur certains aspects de la cohésion lexicale, notamment dans ses rapports avec l'ambiguïté lexicale et les possibilités de l'automatisation, aura sans doute aussi ouvert la voie à l'étude **en contexte** des collocations (au sens structuré), et ce dans une perspective d'automatisation. Il sera ainsi important d'être éclairé, pour une collocation donnée, sur les possibilités de continuité ou de discontinuité et de figement total ou partiel, et sur l'impact de ces éléments sur une exploitation informatique. Nous croyons que des outils comme COLLOCATEUR peuvent aider à découvrir ce potentiel. Et nous croyons que LOCALEX donnera de meilleurs résultats s'il bénéficie de ces connaissances.

Si nous avons si peu exploité la collocation au sens non structuré, c'est parce que nous avons privilégié l'efficacité du système. Notre étude aura cependant montré les enjeux théoriques et pratiques de ces notions. Nous avons, par exemple, relevé que la pseudo-collocation (du point de vue des exigences structurelles) peut coïncider avec la collocation non structurelle.

Dans l'exploitation même des collocations structurées, nous n'avons pas tout exploré. Ainsi, aurions-nous pu, en nous inspirant de l'analyse de F. Rastier (1994), tenir compte, en plus du pouvoir inductif des collocations et de la propagation du sens, de l'effet d'inhibition sémantique de certaines collocations.

⁵⁸ R. Hasan (1984)

Ce travail, multidisciplinaire et multiforme, n'aura pas manqué de soulever des questions, dont beaucoup sont restées sans réponse. Nous avons opté pour une manière concise et tranchée de présenter les problèmes, mais notre propos était d'arriver à tenir un raisonnement général et à en montrer le degré de faisabilité. C'est dans les mêmes termes que nous évoquerons quelques-unes des perspectives de cette thèse.

Nous estimons qu'une amélioration de l'algorithme (en permettant par exemple la comparaison des collocations reconnues dans le contexte) conjuguée à une analyse plus complète du comportement des collocations dans le discours, permettrait d'atteindre des performances de plus grande envergure. Néanmoins, nous restons persuadé que le système que nous avons développé est en mesure, au prix d'adaptations ciblées, de rendre des services appréciables dans un domaine comme la recherche documentaire. Il n'est pas toujours nécessaire d'atteindre une efficacité sans faille pour profiter de l'apport de l'automatisation.

Ainsi, avec une base de données d'envergure raisonnable, il est tout à fait envisageable d'explorer des sites Internet à la recherche de passages de textes comprenant les mots qu'on veut avec la signification qu'on veut. Mais bien plus intéressante encore serait la possibilité d'opérer la sélection sur des textes entiers (pratiquement sur des fichiers). Pour repérer, par exemple, les textes qui parlent de *avocat* au sens de fruit, la démarche consisterait à comptabiliser pour chaque texte les occurrences satisfaisantes. Le fichier est retenu si le nombre d'occurrences atteint un seuil raisonnable (paramétrable par l'utilisateur). Pour une application de ce type, où le besoin n'est pas très pointu, il importe peu que le programme fasse quelques faux pas.

Dans la même veine, une application visant l'indexation ou la classification des documents pourrait s'appuyer sur la base des collocations. Admettons que celles-ci soient organisées en fonction d'une typologie des discours, alors il deviendrait possible de décider de la *nature* d'un texte dès lors qu'un seuil de collocations est atteint (seuil paramétré en fonction de l'expérience et de la connaissance des domaines analysés).

Parmi les ouvertures possibles, nous pensons aussi à la possibilité d'intégrer le moteur de COLLOCATEUR dans LOCALEX. L'intention est de faire profiter LOCALEX des

collocations non encore enregistrées dans la base. L'utilisateur qui voit les résultats de l'extraction des collocations (qui se fait juste avant que les différentes occurrences ne soient localisées) peut, s'il le désire, retenir quelques-unes d'entre elles et les associer à une signification ou à une autre. Cette tâche, dynamique et interactive, est devenue prévisible grâce à la façon dont nous avons conçu la représentation des données: littérales et quasi naturelles.

Par certains aspects, les différents programmes constituent des outils qui peuvent aider à des études de lexicographie, voire d'Analyse du Discours. Bien entendu, des adaptations bien ciblées pourraient mieux les préparer à de telles activités.

5. BIBLIOGRAPHIE

A. LA LINGUISTIQUE TEXTUELLE

- BELLERT, Irena (1970), « On a condition of the coherence of Texts » in *Semiotica*, Vol. II,4.
- BLUMENTHAL, Peter (1991), « La linguistique du texte dans quelques langues de spécialité » in *Hermes, Journal of Linguistics* 7, pp. 7-28.
- BROWN, Gillian, YULE, George (1983), *Discourse Analysis*, Cambridge University Press, 288 p.
- CHARAUDEAU, Patrick (1983), *Langage et Discours : Eléments de Sémiolinguistique*, Paris, Hachette Université. 175 p.
- CHAROLLES, Michel (1976), « Grammaire de texte -Théorie du Discours -Narrativité » dans *Pratiques*, Vol. 11/12.
- CHAROLLES, Michel, PEYTARD, J. (1978), « Enseignement du récit et cohérence du texte » dans *Langue Française*, Larousse, Vol. 38.
- CHEVALIER, Jean Claude (1977-), « Peut-on ne pas être éclectique ? » dans *Le Français dans le Monde*, Larousse, Vol. 129.
- CHEVALIER, Jean Claude (1978), « Analyse du Discours et sa Signification » dans *Littérature*, Larousse, Vol. 18.
- COMBETTES, Bernard (1975), *Pour une linguistique textuelle*, Nancy II : C.R.D.P., 139 p.
- DIJK, Teun Adrianus (1977), *Text and Context, Explorations in the Semantics and Pragmatics of Discourse*, London ; New York, Longman, 261 p.
- DIJK, Teun Adrianus Van (1972), *Some Aspects of Text Grammars. A Study in Theoretical Linguistics and Poetics*, Paris, Mouton, 377 p.
- DUBOIS, Jean et DUBOIS-CHARLIER, Françoise (1990) "Incomparabilité des dictionnaires" dans *Langue Française* 87.
- DUCROT, Oswald (1979), « Les Lois du Discours » dans *Langue Française*, Larousse, Vol. 42.
- DUCROT, Oswald (1980), *Les Mots du Discours*, Minuit, 241 p.
- FOUCAULT, Michel (1980), *L' Ordre du Discours*, Leçon Inaugurale au Collège de France prononcée le 2 Déc. 1970, Gallimard, 81 p.
- GARDIN, J. C. (1974), *Les Analyses de Discours*, Delachaux et Niestlé.

- GARNHAM, A. OAKHILL, J.V. & JOHNSON-LAIRD, P.N. (1982) « Referential continuity and the coherence of discourse. » dans *Cognition*, 11, 29-46.
- GIVON, T. (Ed.) (1983), *Topic continuity in Discourse : A quantitative cross-language study*. Amsterdam : Benjamins.
- GRICE, Paul (1979), « Logique et Conversation » dans *Communications*, Seuil, Vol. 30.
- GRIMES, J. (1975) *The thread of discourse*. The Hague. Mouton.
- GUESPIN, L. (1976), « Types de Discours ou Fonctionnements Discursifs » dans *Langages*, Larousse, p. Vol. 41.
- HALLIDAY, M.A.K. (1976), R. HASAN, *Cohesion in English*, Longman.
- HARRIS, Zellig (1969), « Analyse du Discours » in *Langages* (publié initialement dans *Language* Vol 28, 1952), Larousse, p. Vol. 13.
- HASAN, R. (1984), «Coherence and cohesive harmony» dans FLOOD, J. (éd.) *Understanding Reading Comprehension*, Newark, DE, International Reading Associations, pp. 181-219.
- HALLIDAY, M.A.K. (1977), « Language Structure and Language Function » in *New Horizons in Linguistics* ed. LYONS, Penguin Books.
- HOBBS, J. (1978) « Coherence and coreference » dans Technical note 168, SRI International.
- HOBBS, J.R. (1979), « Coherence and coreference ». *Cognitive science*, 3, pp. 67-90.
- HOBBS, J.R. (1983), « Why is discourse coherent ? » dans NEUBAUER, F.(Ed.) *Coherence in natural language texts* (pp. 29-70). Hambourg : Buske.
- HOBBS, J. R., MOORE, R. C. (1985), *Formal Theories of the Commonsense World*, Norwood, NJ : Ablex.
- KERBRAT-ORECCHIONI, Catherine (1980), *Enonciation de la Subjectivité dans le Langage*, Armand Colin, 290p.
- LABRANCHE, Pierre (1993), *Propos sur la traduction juridique* dans Patrick Drouin, Dir., *Actes des 7èmes journées de linguistique*, Université Laval.
- LANGLEBEN, M. (1981), « Latent Coherence, Contextual Meanings and the Interpretation of a Text » in *Text*, Mouton Publishers, p. Vol. 1,3.
- LOCKMAN, A. & KLAPPHOLZ, A. D. (1980), Toward a Procedural Model of Contextual Reference dans *Discourse Processes*, 25-71p. Vol. 3.
- LUNDQUIST, L. (1980), *La Cohérence Textuelle*, Nyt Nordisk Forlag, Arnold- Busck Kobenhaven.

- MAINGUENEAU, D. (1976), **Initiation aux Méthodes de l'Analyse du Discours : Problèmes et Perspectives**, Hachette, 192 p.
- MEYER, B.J.F., (1975) **The organisation of prose and its effects on memory**. Amsterdam : North-Holland.
- MOKHTARI, Abdelkrim (1985), **Cohérence des textes et argumentation**. Thèse de 3^{ème} Cycle. Université de Jussieu (Paris 7).
- MULLER, Charles (1973), **Initiation aux méthodes de la Statistique Linguistique**, Paris, Hachette Université, 187p.
- PARISI, D., CASTELFRANCI, C. (1976), « Discourse as a Hierarchy of Goals » in Working Papers, Università di Urbino, p. Vol. 54,55.
- POLANYI, L. (1986) **The linguistic discourse model : Towards a formal theory of discourse structure** (Tech. Rep.) No 6409. Cambridge, MA :BBN Laboratories.
- ROULET, E. (1980), « Stratégies d'Interaction, Modes d'Implication et Marqueurs Illocutoires » dans **Cahiers de Linguistique Française**, Université de Geneve.
- SANDERS, Ted J. M., SPOOREN, Wilbert P. M., NOORDMAN, Leo G. M. (1992), « Toward a Taxonomy of Coherence Relations » in **Discourse Processes -A Multidisciplinary Journal** (Ed. Roy O. Freedle), Norwood New Jersey, Ablex Publishing Corporation, 1-35p. Vol. 15,1.
- SLAKTA, D. (1975), « L' Ordre du Texte » dans **Etudes de Linguistique Appliquée**, Paris, Didier, p. Vol. 19.
- VIGNAUX, Georges (1973), « Le Discours argumentatif écrit » dans **Communications**, Vol. 20, pp. 101-159..
- WEINRICH, Harald (1973), **Le Temps : Le Récit et le Commentaire**, Trad. : Michèle Lacoste, Seuil, 333 p.

B. LA LINGUISTIQUE INFORMATIQUE

- ADRIANENS, Geert et SMALL, Steven L. (1988), *Word Expert Parsing revisited in a Cognitive Science Perspective.*, dans Steven I. Small, Garrison W. Cottrel, Michael K. Tanenhaus éd. **Lexical Ambiguity Resolution**, 1988. Morgan Kaufmann Publishers, Inc.
- BOGURAEV, Branimir Konstantinov (1979) **Automatic resolution of linguistic ambiguities**. Thèse de Doctorat. Rapport technique 11, Computer laboratory, University of Cambridge.

- COURTOIS Blandine (1990) « Un système de dictionnaires électroniques pour les mots simples du français » dans *LANGUE FRANCAISE*, Paris, Larousse, Vol. 87. pp. 11-22.
- CHOUKA Yaacov et LUSIGNAN Serge, (1985) « Disambiguation by Short Contexts » *Computers in the Humanities*, Vol 19, no 3 pp. 147-157.
- DAHLGREN, Kathleen (1976), *Referential Semantics*. Dissertation, Los Angeles, University of California. University Microfilms.
- DAHLGREN, Kathleen (1985), « Social Terms and Social Reality » dans *Folia Linguistica Historica*, 107-126p. Vol. 6.
- DAHLGREN, Kathleen (1985), « The Cognitive Structure of Social Categories » dans *Cognitive Science*, 379-398p. Vol. 9.
- DAHLGREN, Kathleen (1988), « Using Commonsense Knowledge to Disambiguate Word Senses » In F. St. Dizier & V. Dahl, eds., *Natural Language Understanding and Logic Programming 2*, Amsterdam, North Holland.
- DAHLGREN, Kathleen (1988), *Naive Semantics for Natural Language Understanding*, Kluwer Academic Publishers, 258 p.
- DAHLGREN, Kathleen (1989), *Discourse Coherence and Naive Semantics*, Universitat Duisburg Gesamthochschule. L.A.U.D.
- DAHLGREN, Kathleen, McDOWELL, J. (1986), « Kind Types in Knowledge Representation. » dans *Proceedings COLING 86*.
- DAHLGREN, Kathleen, McDOWELL, J. (1986), « Using Commonsense Knowledge to Disambiguate Prepositional Phrase Modifiers » dans *Proceedings AAAI 86*.
- FUCHS, Catherine, (1993), *Les traitements automatiques des langues*, Hachette, 304 p.
- FUCHS, Catherine (1996) *Les ambiguïtés du français*, Ophrys, 184 p.
- HIRST, Greame (1981), « Anaphora in natural language understanding: A survey » dans *Lecture notes in computer science* 119. New York: Springer-Verlag.
- HIRST, Greame (1987), *Semantic Interpretation And The Resolution Of Ambiguity*, Cambridge University Press, 263 p.
- IDE, Nancy, (1987) *Computers and the Humanities Courses: Philosophical Bases and Approach*, in *Computers and the Humanities*, 21, 4
- KINTSCH, W. & DIJK, T. A. Van (1978), « Toward a model of text comprehension and production » dans *Psychological Review*, 85, pp. 363-394.

- MANN, W., THOMPSON, S. (1987), « Rhetorical Structure Theory : Description and construction of Text Structures » dans G. KEMPEN (Ed.), **Natural language generation** (pp. 85-95). Boston, MA : Kluwer Academic Publishers.
- McROY, Susan W. (1992) "Using Multiple Knowledge Sources for Word Sense Discrimination" dans **Computational Linguistics**, The MIT Press, pp. 1-30, Vol. 18,1.
- MEPHAM, M. (1973), **Computation in Language Text Analysis**, Québec, CIRB.
- MEPHAM, Michael (1986), « Lemmatiseur : un outil interactif pour les données textuelles », dans **Langues et Linguistique**, no 12, Qubec, Universit Laval.
- MEPHAM, Michael (1989), « La Terminologie grammaticale dans l'analyse des textes. » dans **Les Actes du Colloque de l'Association Belge de Linguistique appliquée**, COLONSTER- BELGIQUE.
- MEPHAM, Michael (1989), « Le Traitement interactif des documents » dans C. OUELLON, **Actes du Coll. La description des langues naturelles en vue d'applications informatiques**, Québec, Université Laval.
- MEPHAM, Michael (1989), « LEXITEX : Textual Analysis Software » dans **Tools for Humanists**, Toronto, Center For Computing in the Humanities : U. Toronto, 49-51p.
- MEPHAM, Michael (1991), « Efficient Pre-processing for the Creation of Large-scale Full-text Data Bases » dans **Les Actes du Colloque RIAO 91**.
- MEPHAM, Michael, TOUPIN, D., LEPAGE, D. (1989), **Lémmatiseur/Lexitex**, QUEBEC, C.T.I. Univerité Laval, 357 p.
- MOKHTARI, Abdelkrim (1993) « Cohésion et désambiguïsation lexicale automatique : cas de l'anaphore » dans **Actes des 7^e Journées de Linguistique**, CIRAL, Université Laval, Québec.
- MOKHTARI, Abdelkrim (1994) « Collocateur. Première version d'un logiciel de reconnaissance semi-automatique d'expressions » dans **Actes des 8^e Journées de Linguistique**, CIRAL, Université Laval, Québec.
- MORRIS, Jane & HIRST, Graeme, « Lexical Cohesion Computed by Thesaural Relations as an Indicator of the Structure of the Text » dans **Computaional Linguistics** pp.21-44 Vol. 17,1.
- PITRAT, Jacques, (1985) **Textes, Ordinateurs et Compréhension**, Eyrolles, , 200 p.
- RASTIER, François (1989), **Modèles sémantiques pour les traitements automatiques des langues naturelles**, Nanterre, EC2.

- RASTIER, François (1991), **Sémantique et recherches cognitives**, PUF, Paris.
- RASTIER, François, CAVAZZA, Marc, ABEILLÉ, Anne, (1994) **Sémantique pour l'analyse, de la linguistique à l'informatique**, Masson. P. 2.
- REICHMAN, R. (1985), **Getting Computers to Talk Like You And Me**, Cambridge, MA : MIT Press, p. .
- RICH, Elaine (1991), **Intelligence Artificielle**, New York, Mc Graw Hill, 621 p.
- SCHANK, R. C., ABELSON, R. P. (1977), **Scripts, Plans, Goals and Understanding**, Hollsdale, NJ : Erlbaum.
- SMADJA, Frank (1993), "*Retrieving Collocations from Text: Xtract*", dans **Computational Linguistics**, Vol 19, no 1.
- SWINNEY, David et HAKES, David T. (1976), "*Effects of prior context upon lexical access during sentence comprehension*" **Journal of verbal learning and verbal behavior** 15 (6), 681-689.
- WILKS, Yorick Alexander (1975) "*A preferential, pattern-seeking semantics for natural language inference*" **Artificial Intelligence**, 6, 53-74.
- ZERNIK, Uri (1990), "*Tagging word senses in corpus: the needles in the haystack revisited.*" Technical Report 90CRD198, GE Research and Development Center, Schenectedy NY. Dans **Proceedings, AAAI Symposium on Text-Based Intelligent Systems**, Stanford, CA, 25-29.

C. L'INFORMATIQUE

- ANTONOVICH, Michael, (1991) **Foxpro 2 Programming Guide**, The lance A. Leventhal, microtrend series. 579 p.
- BARD, Dick, (1993) **Foxpro, Event Driven Programming – How To Build Multi-Window Applications**. Brady Publishing, 550 p.
- CAILLIBOT, Robert, SIMIAN, Gérard (1992), **Intelligence Artificielle I et II (IFT 17586)** Notes de Cours AUT 92, Département d'Informatique, Univ. Laval.
- DREYFUS, Hubert L. (1984), **Intelligence Artificielle, mythes et limites**, Paris, Flammarion, 443 p.
- GOLDSTEIN, Larry Joel (1987), **PASCAL An Introduction to Modern Programming**, Holt, Rinehart and Winston, Inc., 619 p.
- GOLEY IV, George F. (1993) **Creating Foxpro Applications – The Professional Programmer's Guide to Foxpro 2.5**, Que, 547 p.

- GROSZ, B.J. & SIDNER, C.L. (1986), « Attention, Intentions and the Structure of discourse » dans **Computational Linguistics**, 12, 175-204.
- IDE, Nancy M. (1987), **PASCAL for the Humanities**, University of Pennsylvania Press -Philadelphia, 375 p.
- KRUSE, Robert L. (1987), **Data Structures and Program Design**, Second Edition, New Jersey, Prentice-Hall, Inc., 586p.
- POWELL, James E. (1992) **Foxpro 2 Developer's library**, The lance A. Leventhal, microtrend series. 411 p.
- STARKEY, J. Denbigh, ROSS, Rockford J. (1984), **Fundamental Programming with PASCAL**, West Publishing Company.
- STONE, Harold S. (1972), **Introduction to Computer Organization and Data Structures**, McGraw-Hill Book Company, 320 p.
- STUBBS, Daniel (0), **Data Structures With Abstract Data Types and Pascal**.
- Turbo Prolog - BORLAND (1986-), **Turbo Prolog : The Natural Language of Artificial Intelligence**, Borland International, Inc., 221 p.

D. LA RECHERCHE DOCUMENTAIRE

- ALBERICO, Ralph & MICCO, Mary (1990), **Expert Systems For Reference and Information Retrieval**, MECKLER, 395 p.
- MAULDIN, Michael L., CARBONNEL, Jaime (Préfacier)(1991), **Conceptual Information Retrieval - A case Study in Adaptive Partial Parsing**, Norwell (?), Kluwer Academic Publishers, 215 p.

I. ANNEXES

A. GUIDE D'UTILISATION DE MARQUEUR

1. PRÉAMBULE

Depuis que l'informatique a étendu son domaine d'application à celui des textes après avoir conquis le champ du numérique, les besoins de traiter des textes, de les analyser ou tout simplement de les préparer à l'analyse se sont fait sentir. Les programmes qui aident le chercheur à analyser les textes sont maintenant les bienvenus, d'autant plus que l'avantage qu'on peut tirer d'un traitement automatique est devenu moins suspect aux yeux de ceux qui considèrent la machine comme étant incompatible avec tout ce qui relève du symbolique.

Le produit que nous présentons ici n'a pas la prétention de faire une analyse du texte, si le mot *analyse* est compris au sens fort. Néanmoins, la part d'automatisation proposée est loin d'être négligeable.

Il s'agit d'un programme de marquage de textes, ou plutôt de marquage des mots du texte. Chaque mot a des propriétés qu'on peut relever et noter. Il serait fastidieux de le faire à la main, surtout que le besoin de faire des traitements automatiques sur ces mêmes propriétés est de plus en plus grand.

MARQUEUR tente d'apporter une solution à ce genre de besoins.

MARQUEUR a hérité beaucoup d'un logiciel développé par le C.T.I.¹ sous l'instigation de Michael Mepham et en collaboration avec d'autres chercheurs. Ce logiciel portait le nom de *Lemmatiseur*. Mais **MARQUEUR** diffère sur beaucoup de points de *Lemmatiseur*, et d'autre part ses objectifs sont plus avancés: alors que le premier se limitait à l'attribution des lemmes, **MARQUEUR** offre le moyen d'associer aux mots du texte une

¹ Centre de Traitement de l'Information, actuellement S.I.T. (Service de l'Information et des Technologies)

dizaine de propriétés dont la plupart sont au libre choix de l'utilisateur. Le résultat du travail est le marquage du texte (le texte sera en quelque sorte annoté au niveau de chaque mot si on le souhaite), mais également un jeu de dictionnaires prêts à toutes sortes d'usage. **MARQUEUR** a aussi l'avantage d'offrir des fonctions intégrées de gestion de l'ensemble de la base (texte et dictionnaires).

Le logiciel a vu le jour sous une version initiale en septembre 1994. Nous l'avons présenté à un public de chercheurs intéressés, dans le cadre des activités du C.I.R.A.L. le 19 septembre 1994. Les encouragements ressentis nous incitent maintenant à envisager l'amélioration du produit initial au point de le rendre opérationnel, d'une manière autonome, pour un public de chercheurs plus large.

Le coauteur Abdelkrim Mokhtari a bénéficié d'une bourse d'échanges entre départements accordée par l'AUPELF-FICU pour apporter ces améliorations. Nous sommes très reconnaissant aux responsables de l'organisme qui ont placé leur confiance dans ce projet. Nous remercions également la Directrice du CIRAL, Mme Denise DESHAIES, qui a très aimablement accepté de nous accueillir durant la période d'été 1996.

2. FONCTIONS

a) Description

Le logiciel en question a été nommé **MARQUEUR** parce qu'essentiellement, il s'occupe de marquage de textes. Par marquage, nous entendons l'attribution de propriétés à chacun des mots du texte analysé. Les propriétés peuvent varier et le choix de leur nature comme de leur nombre est à la discrétion de l'utilisateur. Une grande partie du travail est faite automatiquement, le reste se fait d'une manière interactive.

Le processus de marquage consiste à attacher à différents mots du texte une sorte d'archi-lexème, comme l'infinitif pour une forme conjuguée d'un verbe, ou le mot au singulier pour un mot au pluriel. La démarche consiste à exploiter d'une part un dictionnaire recensant

différentes formes et leurs lemmes respectifs, et d'autre part à traiter individuellement, en interactif, les cas ambigus ou absents du dictionnaire.

Le logiciel **MARQUEUR** innove sur plusieurs aspects. Dans *Lemmatiseur*, le dictionnaire ne détenait les différents lemmes du mot que si le hasard du traitement permettait de les enregistrer. (Chaque nouveau lemme pour une même forme donnée étant enregistré au moment où on le rencontre dans un texte.) Ce qui devait générer des lacunes dans le traitement. **MARQUEUR** résout cette difficulté en permettant à l'utilisateur d'apporter à chaque forme tous les lemmes utiles reliés à cette forme. (Ex: la forme "juge" aura le lemme *juger_verbe* et un autre "juge_nom".)

Mais **MARQUEUR**, qui garde la même démarche que *Lemmatiseur*, remplace désormais le terme "Lemme" par celui de descripteur, ceci parce que l'objectif maintenant est plus large qu'une simple lemmatisation. Il s'agit désormais d'assaisonner les mots de toutes sortes d'ingrédients que le chercheur (linguiste, historien ou autre...) juge pertinents à ses objectifs. Et à commencer par un descripteur qui va par la suite permettre d'opérer une discrimination des mots. Ensuite, une ou plusieurs autres propriétés qu'on est libre d'imaginer. Les propriétés en question, une fois saisies, iront d'une part s'incruster sur les mots du texte, et d'autre part alimenter un dictionnaire qui augmente avec le temps.

Du coup, le logiciel se présente comme une sorte de système de gestion de bases de données textuelles. D'une part, il y a la base du texte traité, assortie des propriétés particulières à chaque mot. D'autre part, il y a le dictionnaire avec toutes les entrées (uniques car fondées sur la forme) et pour chaque entrée, les différentes propriétés. Une entrée comme *juge* peut avoir un descripteur *juger-verbe* et un deuxième *juge_nom*. Au premier descripteur, on peut associer par exemple la traduction en anglais, l'origine latine, la transcription phonétique (non disponible pour le moment mais envisagée pour l'avenir), et ainsi de suite.

Un des points forts de **MARQUEUR** est justement cette possibilité de désigner quelles propriétés on veut *coller* au mot, de le faire d'une manière souple et conviviale (l'aspect ergonomique est hautement respecté).

b) Utilisateurs ciblés

MARQUEUR peut être utilisé par des chercheurs de différents bords. Il peut intéresser le linguiste (lexicographe, terminologue, analyste de discours,...), comme tout autre analyste concerné par l'attribution de propriétés données à un texte donné, et par la confection d'un dictionnaire des mots de ce texte et de leurs propriétés respectives.

3. INSTALLATION

Créer un répertoire sur votre disque dur. Par exemple: C:\Marque

1. Copier tous les fichiers de la disquette de **MARQUEUR** dans ce répertoire.
2. Lancer le programme Install
3. Pour exécuter **MARQUEUR**, se positionner dans le répertoire *Marque* et taper « Marque » suivi de retour chariot.

4. UTILISATION***a) Démarrage***

L'utilisation de **MARQUEUR** est très intuitive. En général, un simple examen logique des choix proposés dans une fenêtre ou un menu est suffisant pour savoir ce qu'il faut faire.

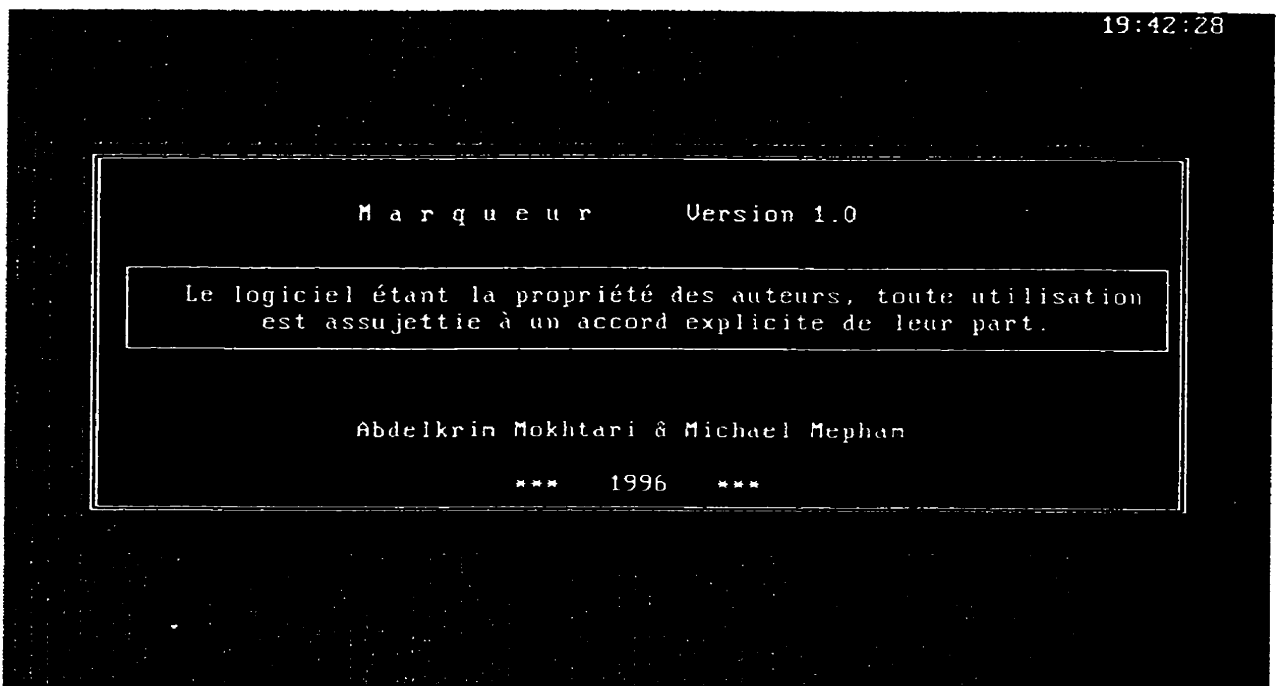


Figure 3 Fenêtre d'accueil de MARQUEUR

Vous pouvez lancer le programme **MARQUEUR** en tapant à partir du répertoire d'installation (par exemple: C:\Marque) le mot: « marque ». Faites ensuite un retour chariot. Vous verrez le premier écran s'afficher (voir Figure 3).

La Figure 4 montre le bureau de travail dans **MARQUEUR**. Tout en haut de ce bureau, se présente le menu principal.

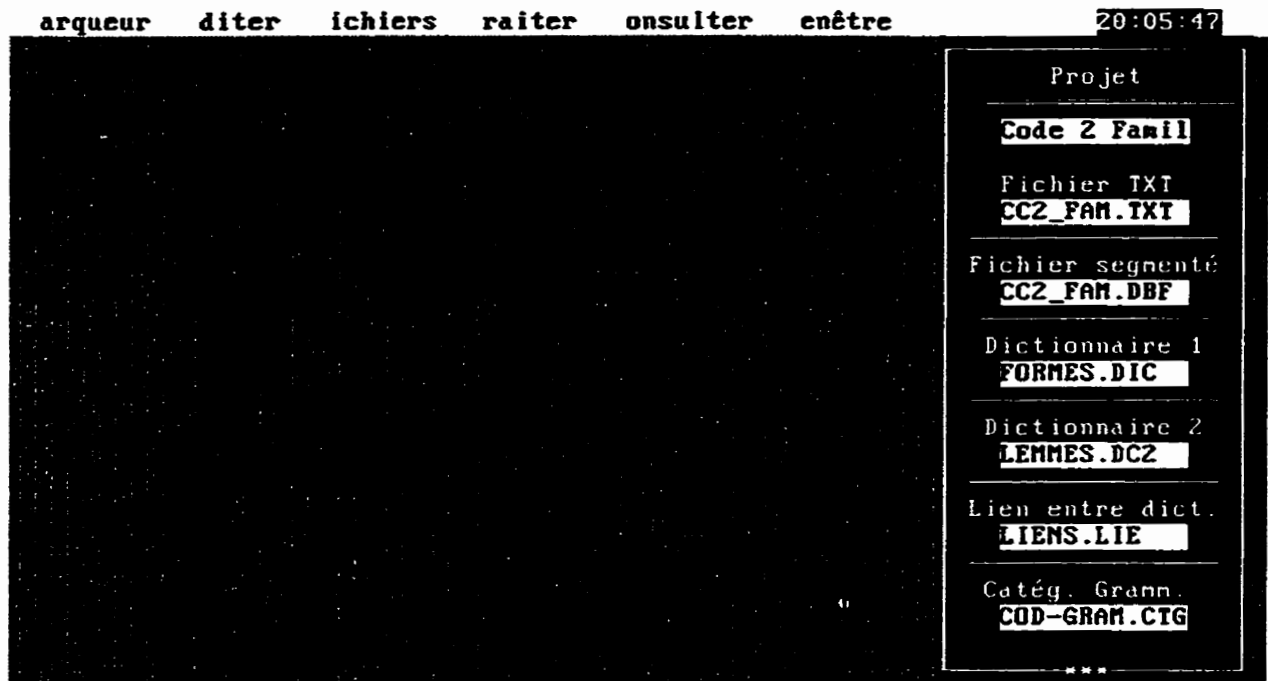


Figure 4 Menu principal

La colonne à droite dans le bureau indique, à titre d'exemple, un ensemble de fichiers ouverts utilisés dans le processus de marquage. Le tableau suivant énumère les fichiers nécessaires au bon fonctionnement de MARQUEUR.

Fichier	Extension	Commentaire
Projets.DBF	*.DBF	Ce fichier est utilisé par le logiciel pour recenser les noms de projet. À la différence des autres, ce fichier n'est pas affiché dans la colonne du bureau de MARQUEUR. En revanche, il est utilisé pour afficher le nom du projet en cours d'utilisation.
Fichier TXT	*.TXT	Doit être dans le format MS DOS avec sauts de ligne , à distinguer du fichier ASCII ou Texte (La différence qui nous intéresse ici est que le premier est formaté avec des lignes ne dépassant pas 256 caractères). En revanche, <u>l'extension doit être: TXT.</u> Il est recommandé de supprimer dans le traitement de texte utilisé les espaces qui peuvent se trouver entre la fin d'un mot et une marque de ponctuation. Exemple: « faculté . »
Fichier segmenté	*.DBF	Ce fichier est le résultat de la segmentation qui peut être réalisée à partir du menu Traiter.
Dictionnaire des Formes	*.DIC	Au commencement d'un projet, on peut créer ce fichier à partir des options de Fichiers
Dictionnaire des Lemmes	*.DC2	Au commencement d'un projet, on peut créer ce fichier à partir des options de Fichiers
Fichier des Liens	*.LIE	Au commencement d'un projet, on peut créer ce fichier à partir des options de Fichiers
Fichier des Catégories Grammaticales	*.CTG	Au commencement d'un projet, on peut créer ce fichier à partir des options de Fichiers

Les fichiers suivants sont essentiels à **MARQUEUR**, mais ils sont transparents à l'utilisateur. Ce dernier n'a pas à s'en occuper, à l'exception du fichier Comment.txt qui peut être utilisé en dehors de **MARQUEUR**.

Fichier	Taille	Utilité
Marque.exe	1 272 515	Fichier exécutable
Foxpro.int	43 187	Conversion des caractères
Comment.txt		Commentaire
Projets.DBF	670	Liste des projets
Projets.FPT	1 024	Mémo des projets
Infos.int	402	Fichier interne à MARQUEUR
Parametr.DBF		Paramètres du disque et du chemin, (interne à MARQUEUR)
Marque.TXT		Journal
Config.fp		Fichier de configuration

b) Création d'un projet

Pour marquer un texte, il est recommandé de commencer par créer un *projet* (dans Fichiers). Un projet est entendu ici comme l'ensemble des fichiers relatifs au marquage d'un texte. En effet, **MARQUEUR** a besoin en plus du texte à marquer, d'un jeu de dictionnaires et d'une table des catégories grammaticales. L'écran de la Figure 5 accessible à partir de Fichiers montre les boutons qui permettent de créer les différents fichiers.

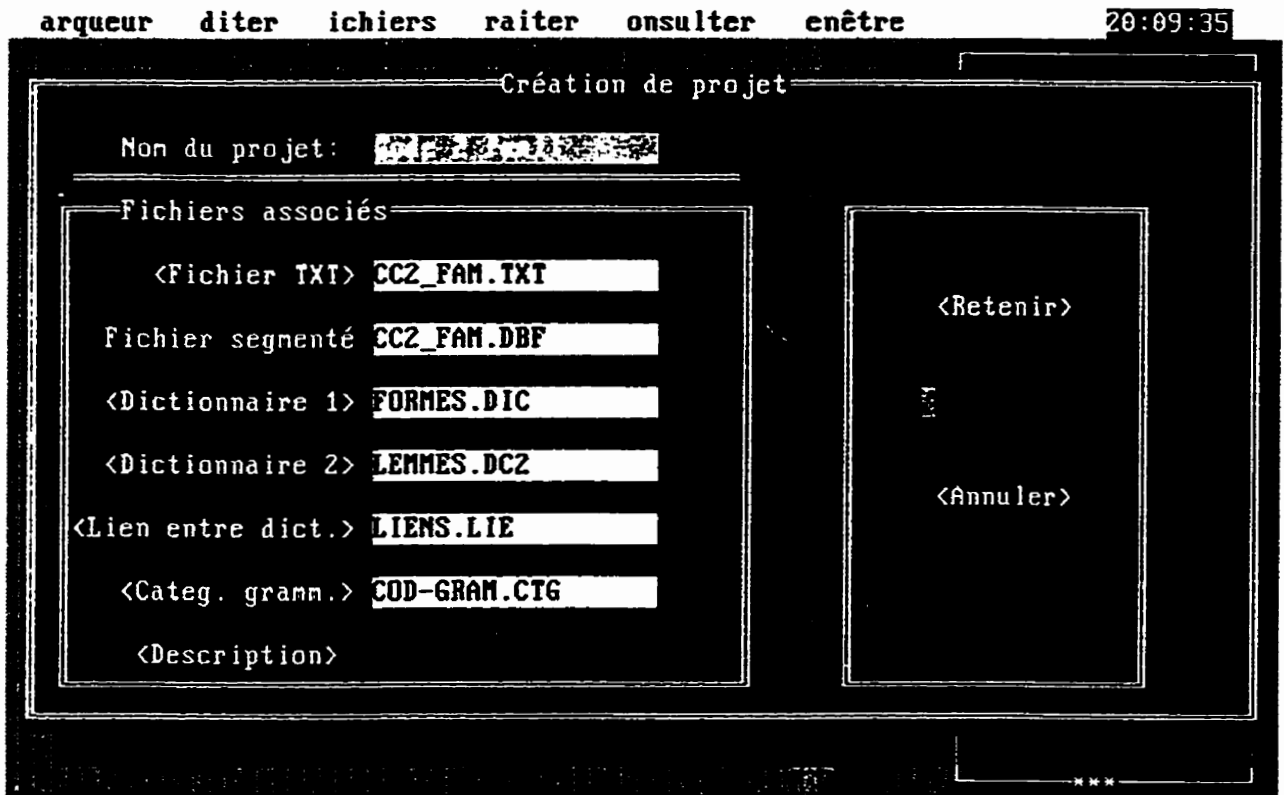


Figure 5 Création de projet

Le fichier segmenté (Extension DBF) est affiché automatiquement s'il en existe un sur le répertoire qui correspond au fichier TXT.

En cliquant sur un des boutons relatifs aux fichiers, il sera possible de sélectionner un fichier existant ou d'en créer un nouveau.

Le bouton Description vous permet de rattacher une description détaillée au nom du projet.

Retenir un projet revient à enregistrer les noms des différents fichiers comme un ensemble inséparable. En ouvrant le projet, vous vous épargnez la peine d'ouvrir séparément tous les fichiers indispensables au bon fonctionnement de MARQUEUR, mais surtout, il ne sera plus nécessaire de vous rappeler quel fichier fonctionne avec quel autre.

Retenir le projet revient aussi à l'ouvrir immédiatement. Si on veut éviter de l'utiliser tout de suite, on le retient, ensuite, on ouvre un autre projet en allant dans l'option appropriée dans Fichiers.

c) Ouverture d'un projet

Pour ouvrir un projet préalablement créé, il suffit d'atteindre la fenêtre représentée dans la Figure 6. Pour ce faire, sélectionner *Ouvrir un projet* dans *Fichiers*.

En réalité, cette fenêtre, combinée avec celle de la création de projet, vous permet de gérer différents projets: ouvrir celui qui vous convient pour la tâche présente, effacer celui dont vous n'avez plus besoin.

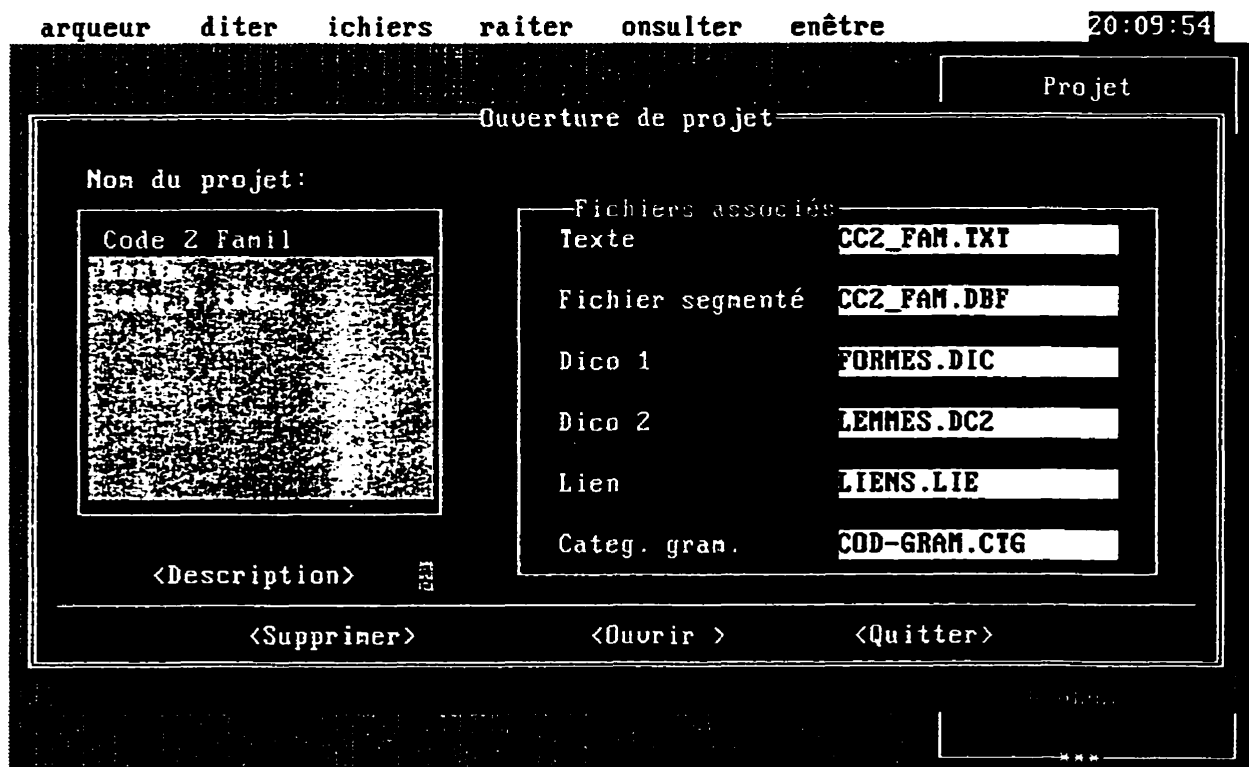


Figure 6 Ouverture de projet

Conseil: Utiliser le bouton *Description* pour vous rappeler les objectifs du projet, pour autant que vous avez pris la peine de remplir cette rubrique lors de la création du projet.

d) Segmentation du texte

L'étape qui suit l'ouverture du projet consiste en une segmentation du texte (fichier avec une TXT, mais enregistré dans le format **MS-DOS** avec **sauts de ligne**). Cette opération se fait en optant pour *Segmenter* dans *Fichiers*.

Une fois le texte segmenté, vous pourriez constater que la colonne de droite affiche maintenant le nom du fichier segmenté (même nom que le fichier TXT mais avec l'extension DBF).

Vous pourriez ensuite passer à un marquage automatique. Si les dictionnaires sont vides, cette opération n'apporte rien au marquage, mais elle est très avantageuse si vous utilisez un dictionnaire rempli d'une manière substantielle. En général, un dictionnaire de 1000 mots permet de marquer un texte inédit de 50 pages de plus de 50 %.

e) Marquage automatique

Le marquage automatique aboutit à l'affichage d'un ensemble d'informations dont l'exemple est représenté à la Figure 7.

arqueur diter ichiers raiter onsulter enêtre 20:11:35

Marquage terminé!

Fichier TXT
CCZ_FAM.TXT

Déroulement...
Marquage en cours... 100%

segmenté
T.DBF
naire 1
.DIC

Total des mots du texte	:	13613
Total des entrées du dictionnaire	:	1201
Occurrences présentes dans le dictionnaire	:	12209
- Occurrences univoques (traitées)	:	10449
- Occurrences ambiguës (non traitées)	:	1760
Mots absents du dictionnaire (non traités)	:	1404

Figure 7 Marquage automatique

f) Marquage interactif

Une fois le marquage automatique réalisé, on peut engager le marquage interactif.

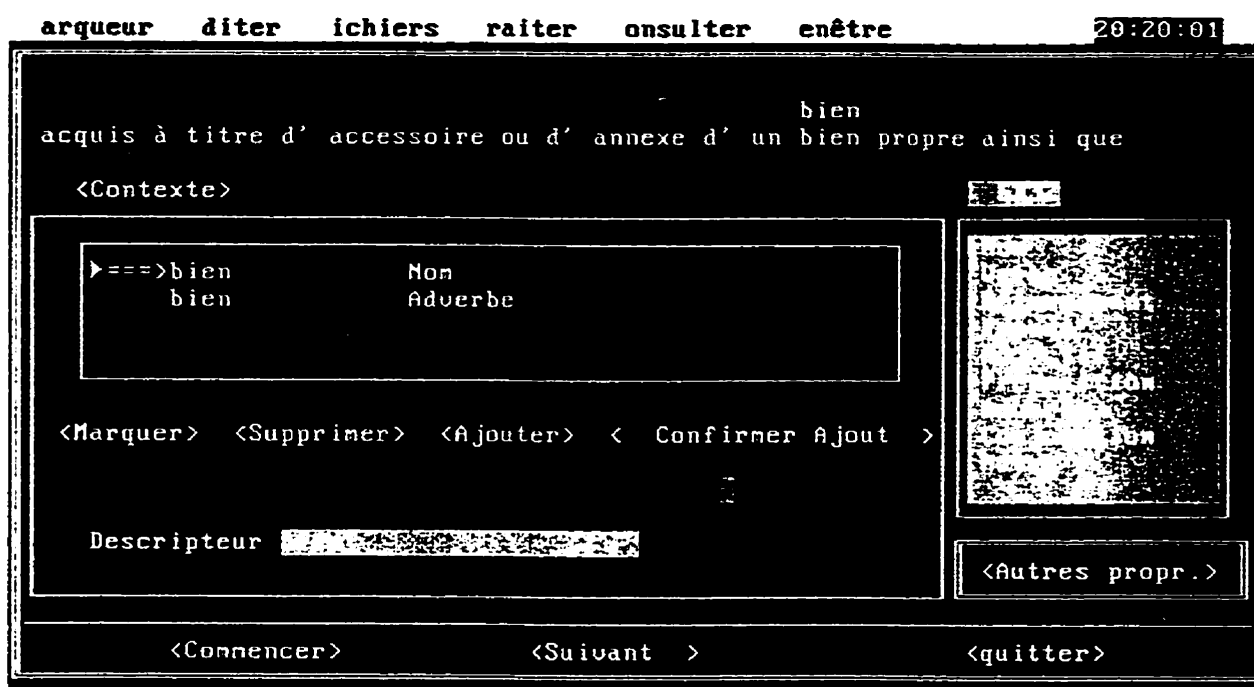


Figure 8 Marquage interactif

C'est dans la fenêtre de la Figure 8 que se fait le travail interactif. Les boutons et les champs de saisie vont s'activer ou se désactiver en fonction de la tâche à accomplir.

« Commencer »

C'est le bouton qui déclenche l'opération de marquage interactif. Si vous en êtes à vos premières séances de marquage, vous verrez rapidement une ligne de texte s'afficher avec un des mots répété sur une ligne séparée. Mais si vous en avez marqué beaucoup, il se peut que le logiciel prenne un peu de temps avant de vous donner la main (un compteur vous signale le progrès de l'opération).

Le mot écrit en blanc indique lequel des mots dans la ligne en rouge devrait être traité.

« Suivant »

Une fois que vous avez terminé les opérations de marquage concernant un mot, vous pouvez passer au mot suivant. Vous utilisez, pour ce faire, le bouton « *suivant* ».

« Quitter »

Vous pouvez quitter la fenêtre de marquage interactif en cliquant sur ce bouton.

« Marquer »

Ce bouton est actif quand le mot courant comprend plus d'un lemme. On peut d'ailleurs constater que dans ce cas la liste des lemmes et catégories est affichée.

Quand le bouton « Marquer » est activé, cela signifie que **MARQUEUR** a besoin de savoir quel duo lemme/catégorie vous voulez rattacher au mot courant. Pour répondre à ce besoin, il suffit de cliquer sur le bouton « Marquer », de sélectionner le duo désiré et de confirmer.

Remarque: Le bouton de confirmation affiche alors l'étiquette appropriée: «Confirmer Marque ».

Après confirmation, **MARQUEUR** indique le duo que vous avez choisi par un message sur la même ligne ou est affiché ce duo.

Cette annotation signifie que le duo des propriétés sélectionné ira se greffer au mot dans le texte (l'occurrence courante), mais il doit être entendu que le jeu des dictionnaires conserve l'ensemble des propriétés affichées dans la fenêtre, y comprises les propriétés qu'on vient de rattacher au mot du texte.

« Supprimer »

Ce bouton est actif quand le mot courant comprend au moins un duo lemme/catégorie. On peut d'ailleurs constater que quand la liste des duos lemmes/catégories se réduit à zéro, le bouton se désactive.

Quand le bouton « Supprimer » est activé, cela signifie que **MARQUEUR** vous donne la possibilité de « supprimer » un duo qui ne vous convient pas pour une raison ou une autre.

Pour faire une suppression, il suffit de cliquer sur le bouton « Supprimer », de sélectionner le duo qu'on veut effacer et de confirmer.

Remarque: Le bouton de confirmation adopte le label approprié: "Confirmer Supprimer".

« Ajouter »

Est actif quand il y a au moins un duo de propriétés affiché. Vous pouvez alors ajouter autant de duos que vous voulez.

Quand le mot est ambigu (du point de vue de l'association lemme/catégorie), les différents choix présents dans le dictionnaire sont affichés, mais il est possible qu'aucun choix ne vous convienne, vous pouvez alors en ajouter à votre convenance et marquer ensuite celui que vous préférez grâce au bouton Marquer.

Dans la logique de **MARQUEUR**, il est nécessaire d'ajouter au duo présent un ou plusieurs autres duos que vous considérez comme faisant partie du potentiel d'emploi du mot en question.

Par exemple, si **MARQUEUR** s'arrête sur le mot « ferme » parce qu'il ne l'a pas trouvé dans le dictionnaire, vous pouvez assigner un premier duo (fermer-verbe, si l'occurrence courante est un verbe), mais vous devez tout de suite envisager d'ajouter le duo « ferme-nom », le duo « ferme-adjectif » et enfin le duo « ferme-adverbe ». Le premier duo annotera l'occurrence du mot dans le texte, et sera également placé dans le dictionnaire. Les trois autres seront placés uniquement dans le dictionnaire.

Si vous attribuez un seul duo à l'occurrence présente, **MARQUEUR** traitera automatiquement toutes les occurrences ultérieures sans vous aviser.

Si vous cumulez plus d'un duo, **MARQUEUR** s'arrêtera sur chacune des occurrences suivantes. Ce sera fastidieux, mais vous pouvez veiller à retenir comme ambigüs uniquement les mots dont l'ambiguïté est pertinente par rapport à votre perspective, et donc accélérer le traitement en attribuant un seul duo.

Le bouton *Ajouter* active les champs de saisie du lemme et de la catégorie.

« Confirmer Assignment »

Ce bouton est actif seulement quand le mot courant est absent du dictionnaire des formes et que par conséquent il y a lieu de lui attribuer des propriétés.

Les champs de saisie des propriétés sont automatiquement activés dans une telle situation.

Faites votre choix avant de confirmer l'assignation.

g) Modification de la structure du dictionnaire

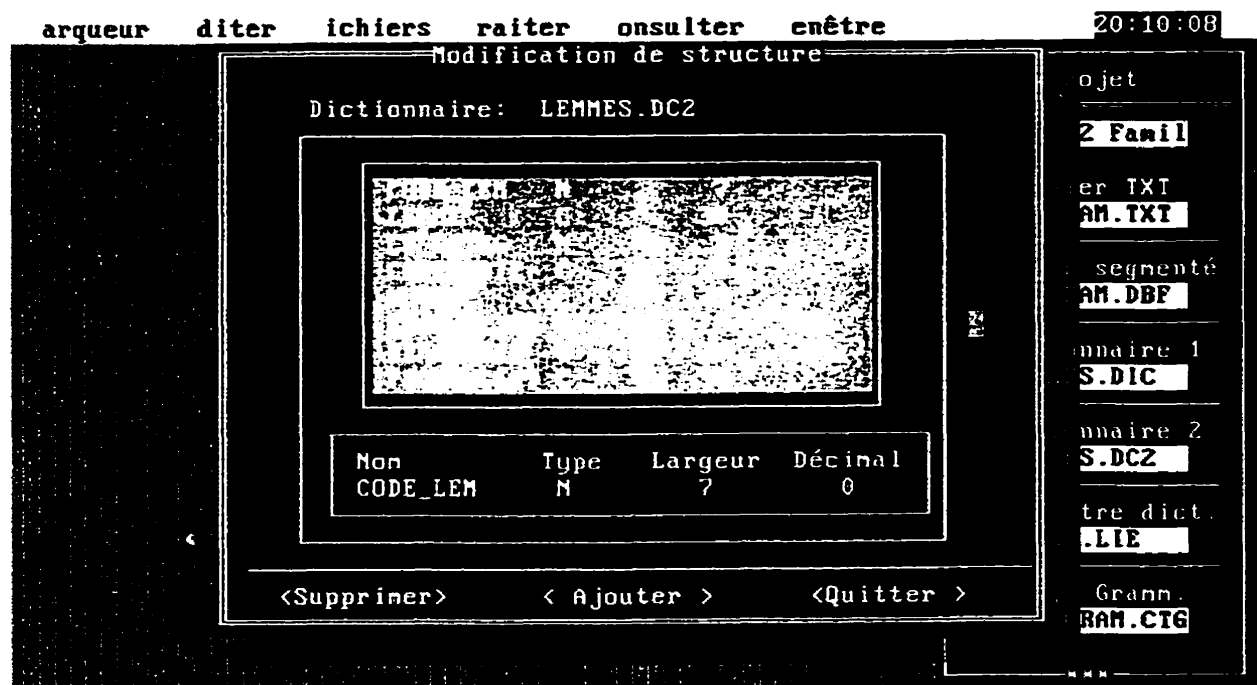


Figure 9 Modification de la structure du dictionnaire

La fenêtre de la figure Figure 9 sert d'interface pour modifier la structure du fichier dictionnaire. Pour y accéder, passer par *Fichiers*, ensuite *Modifier...*. Les champs du fichier sont affichés au milieu de la fenêtre. On peut faire défiler les noms de champs en utilisant l'ascenseur ou la ligne droite au dessous du dernier champ listé.

C'est ici que réside la puissance et l'intérêt de **MARQUEUR**. En effet, La latitude de modifier la structure du dictionnaire signifie que l'utilisateur est libre d'adapter **MARQUEUR** à ses véritables objectifs. De ce fait, il peut utiliser le logiciel pour des projets variés, et mettre dans chacun un dictionnaire différent.

Pour ajouter un champ, opter pour le bouton *Ajouter*. Des zones de saisie vont s'activer et l'on peut ainsi remplir le nom du champ additionnel (10 caractères maximum), le type (C pour caractère, N pour numérique et L pour logique, le choix se fait en appuyant sur la

barre d'espace), la dimension (maximum 10 pour le numérique, 255 pour le caractère, 1 pour le logique)

En cas d'erreur, utiliser la barre d'espace ou cliquer sur un champ actif pour faire disparaître le message d'erreur.

h) Propriétés additionnelles

La Figure 10 montre la fenêtre de saisie des champs additionnels. Le champ *Notes* reste cependant imposé à l'utilisateur.

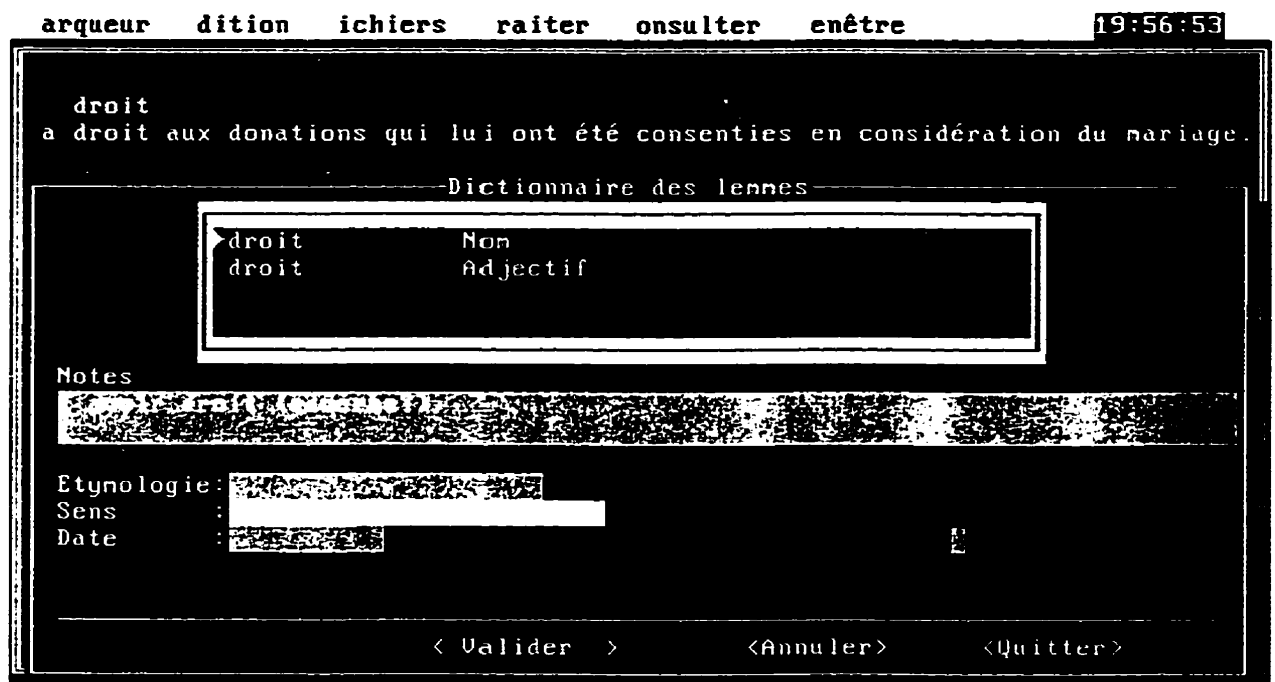


Figure 10 Saisie de propriétés additionnelles

i) Consultation des données

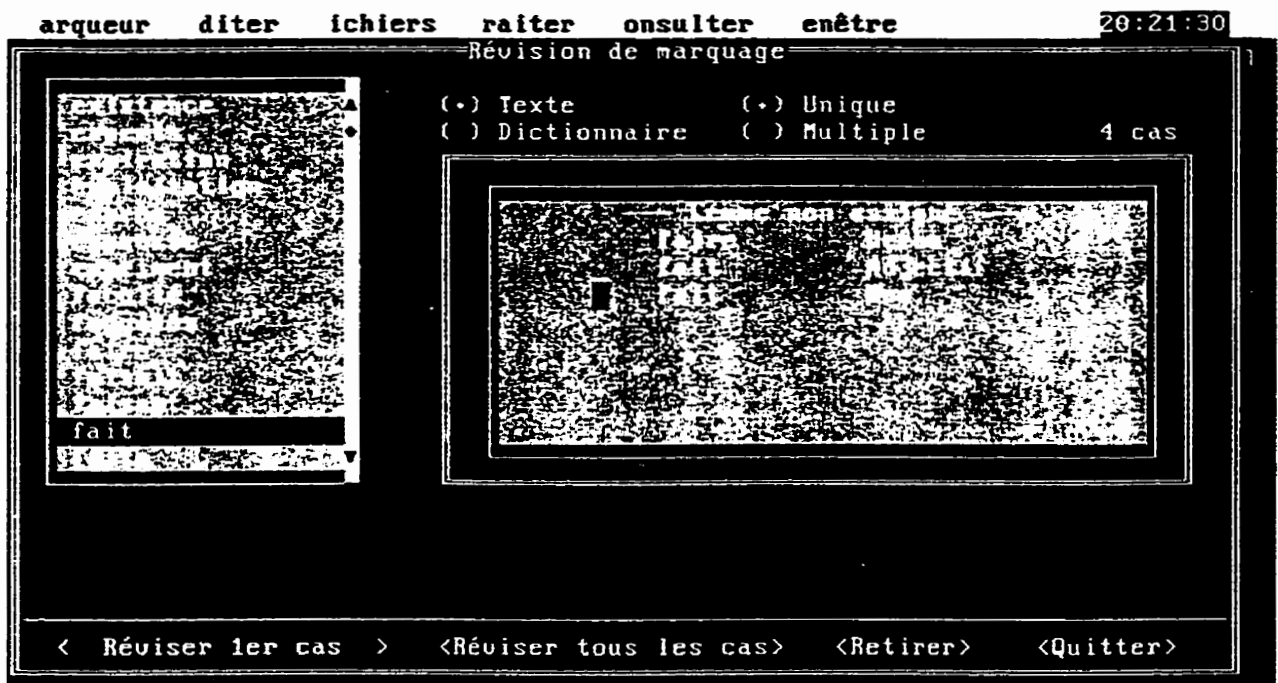


Figure 12 Révision de marquage, options: Texte, Unique.

Par contre, si vous sélectionnez Texte, Multiple, les occurrences du mot *fait* sont toutes affichées (sans le contexte cependant), ainsi qu'on le voit dans la Figure 13.

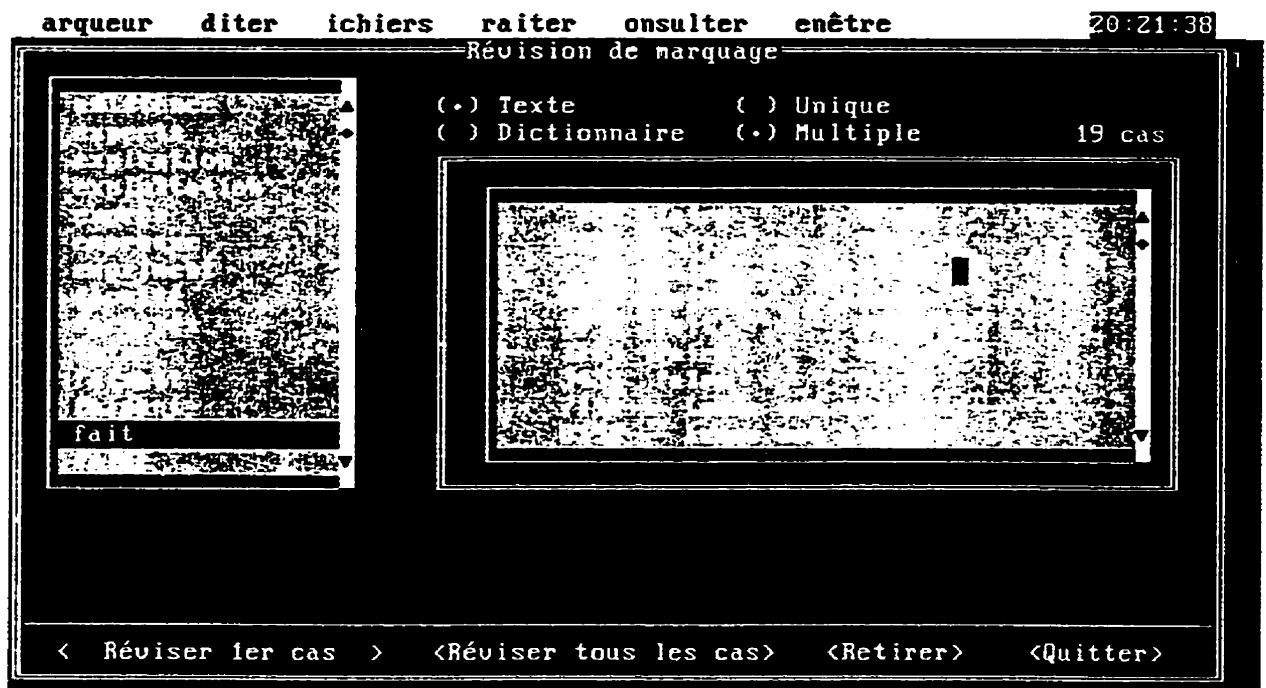


Figure 13 Révision, options: Texte, Multiple, sans visualisation du contexte

Tout en restant dans ces options, il est possible de voir le contexte particulier à chaque occurrence. Cette possibilité de visualiser le contexte offre en vérité un excellent moyen de repérer (manuellement) les différents sens d'un mot. Il est possible d'enregistrer le corpus dans un fichier texte. Il suffit d'activer *Capter* (Menu MARQUEUR), de sélectionner le contexte désiré et de placer le contenu dans le fichier Comment.TXT (qu'on peut ouvrir dans le menu MARQUEUR). La technique est similaire à Copier-Coller. Elle offre l'avantage de copier des tranches d'écran, qu'on soit positionné dans une zone de saisie ou non. Pour une illustration, voir la Figure 14.

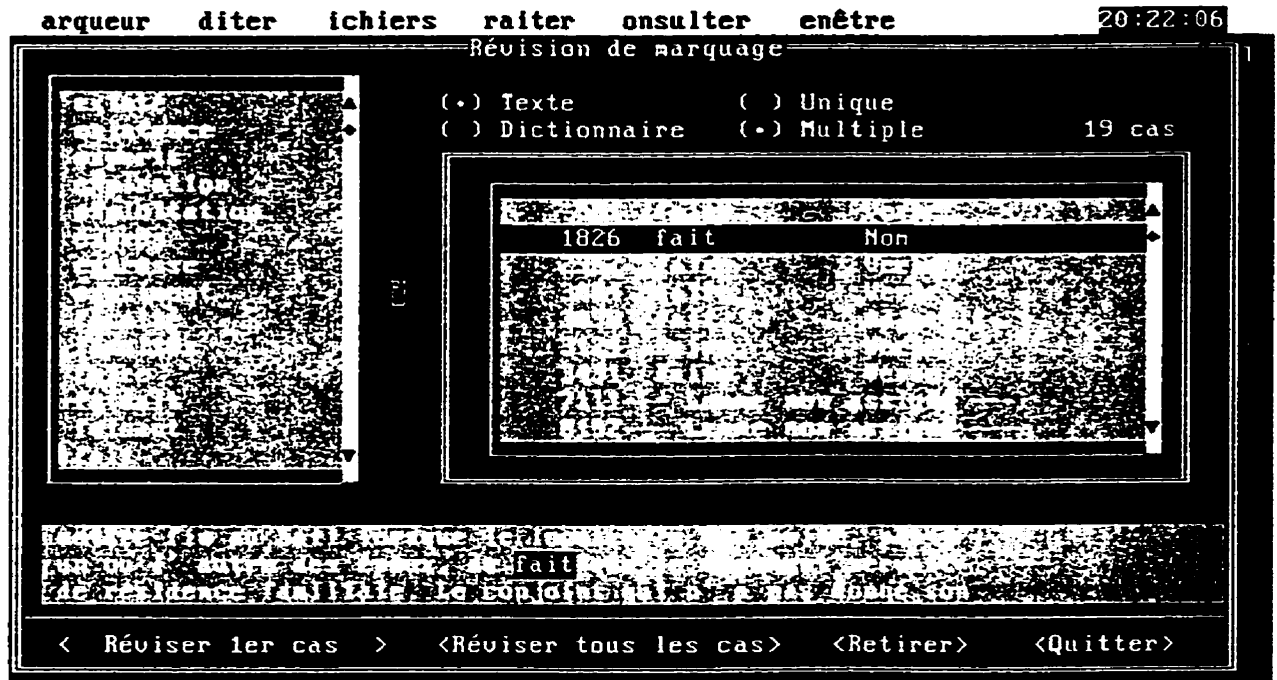


Figure 14 Révision, options: Texte, Multiple, avec visualisation du contexte

Il est encore possible de visualiser le contenu du dictionnaire par rapport au mot courant (voir la Figure 15).

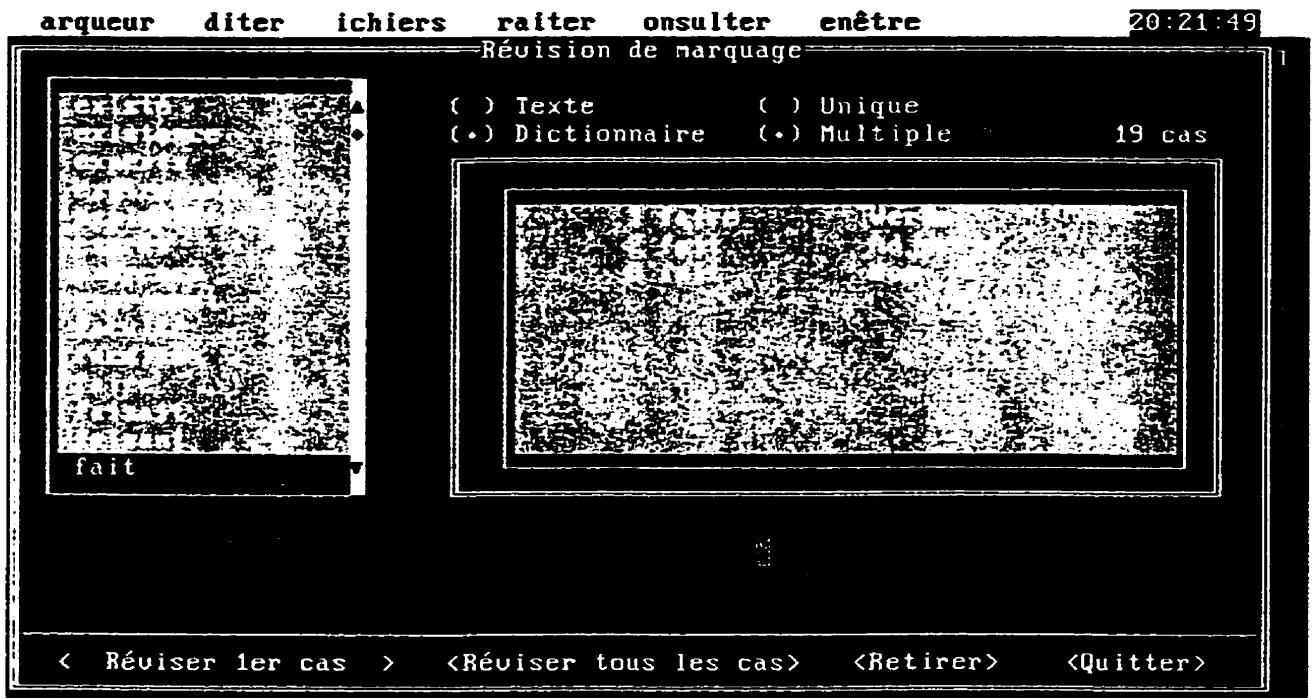


Figure 15 Révision: options: Dictionnaire.

Une fois que vous avez fini de consulter les différents affichages, vous pouvez sélectionner les occurrences que vous voulez réviser. Il suffit pour ce faire de double-cliquer sur l'occurrence voulue. Double-cliquez de nouveau pour annuler une sélection. Pour une sélection ou un retrait global, utilisez les boutons appropriés en bas de la fenêtre.

5. QUELQUES CONSEILS

Ne pas hésiter à considérer comme ambigus les cas suivants:

Entre catégories fortes, ou fortes et faibles

Nom (forte)	Verbe (forte)
demande	demande

Nom (forte)	Adverbe (faible)
pas	pas

Les cas marginaux ou rares:

Exemple: cas échéant

Adjectif et non verbe,

Négliger le verbe *échoir*.

Les cas des catégories non usuelles:

Il n'est pas nécessaire de retenir systematiquement toutes les ambiguïtés. Certaines ne sont pas pertinentes. Pas besoin, par exemple, de distinguer la fonction nominale d'un adjectif, si en tant que nom le mot n'est pas usuel.

Exemple: la troisième est... (Adj. et pas un nom)

Chaque fois que c'est lexicalisé, tenir compte de l'ambiguïté.

Le critère de distinction des catégories en vue de COLLOCATEUR doit guider le choix.

Les participes passés qui prennent la fonction adjectivale doivent être catégorisés comme adjectif afin d'obtenir des collocations nominales comme *témoin informé* ou *roi éclairé*.

Les ambiguïtés sans effets peuvent être négligées

Exemple: S' (conj. si et pronom s'aime) n'a pas besoin de désambiguïstation, en autant qu'il n'y a pas de collocations où on a besoin de « S' ».

A. GUIDE D'UTILISATION DE COLLOCATEUR

1. PRÉAMBULE

COLLOCATEUR a été initialement conçu par Michael Mepham. Ce dernier nous avait invité à contribuer, en collaboration avec Mario Toupin, un informaticien du C.T.I. (aujourd'hui SIT), à en raffiner la conception. Mario Toupin s'est ensuite occupé de l'analyse informatique et a produit un texte d'une centaine de pages. La tâche de la programmation nous a été confiée, mais nous n'avons pas travaillé seul. Mario Toupin nous a apporté un soutien informatique important. Il suivait de près l'avancement de notre programme et n'hésitait pas à écrire du code quand c'était nécessaire. M. Conrad Bourdon, également du C.T.I., suivait de son côté le projet et discutait régulièrement de son évolution.

Le résultat de cette collaboration a été une première version de COLLOCATEUR (1993). Cette version était assez riche quant à l'ensemble des fonctions offertes par l'application. Mais avec le temps, le menu proposé à l'époque est devenu vétuste, et la structure des fichiers devait être revue. C'est la raison pour laquelle nous avons cru bon préparer une nouvelle version qui cadre mieux avec les autres applications du système que nous présentons dans notre thèse. Cette nouvelle mouture est moins ambitieuse que la première quant à l'exhaustivité des fonctions offertes, mais elle est plus allégée et plus conviviale. Elle répond en tout état de cause à l'objectif que nous nous fixons: l'extraction automatique des collocations.

2. FONCTIONS

a) Description

COLLOCATEUR a été fondamentalement conçu pour extraire automatiquement des collocations. Le principe de fonctionnement est simple, mais productif. Il s'agit de retenir comme membre d'une collocation les mots pourvus d'une certaine catégorie grammaticale (l'utilisateur en décide) et placés à une position précise (choix également fait par l'utilisateur).

Si un utilisateur décide qu'un nom doit être retenu s'il est en position initiale, médiane ou finale, le programme retiendra, par exemple les mots suivants:

(1) Véhicule rouge.

(2) Il a acheté un véhicule rouge.

(3) un grand véhicule.

Pour retenir la première suite de mots (exemple (1)), il suffit de décider qu'en position finale, l'adjectif peut contribuer à la formation d'une collocation. Et il en va ainsi pour toutes les catégories grammaticales.

Il faudrait répondre maintenant à la question qu'on se pose à ce stade: comment le programme sait-il que tel mot est un nom ou un adjectif ou autre chose? La réponse est que, dans le contexte qui est le nôtre, le texte est traité préalablement par le logiciel MARQUEUR. C'est ce dernier qui permet d'assigner la catégorie grammaticale à chacun des mots du texte proposé au traitement.

Il faut ajouter que l'utilisateur a le loisir de changer ses paramètres à chaque extraction de façon à obtenir le type de collocations qu'il souhaite. On peut en effet disposer d'un jeu de fichiers paramètres et corrélativement d'un jeu de fichiers des collocations (résultats). En jouant sur ces possibilités, on obtient à chaque fois une taxonomie différente en traitant le même corpus.

b) Utilisateurs ciblés

COLLOCATEUR vise un public assez large d'utilisateurs. Il concerne tout chercheur intéressé par un inventaire plus ou moins systématique des collocations contenues dans un texte donné. En premier lieu, nous pensons aux linguistes interpellés par la question des expressions et des séquences plus ou moins figées. Dans cette catégorie d'utilisateurs, les analystes du discours, les lexicographes viennent en première place.

Les linguistes informaticiens peuvent trouver dans ce programme un outil efficace pour préparer des bases de données destinées elles-mêmes à d'autres applications.

L'inventaire des collocations repérées se présentant sous forme de base de données au standard DBF, l'exploitation de ces données sera d'une grande commodité. Malgré les bruits générés par le programme d'extraction, il reste que les résultats atteignent un niveau de validité respectable. Des corrections et un nettoyage manuel des résultats sont indispensables, mais il s'agit d'un travail peu contraignant.

3. INSTALLATION

L'installation du logiciel est très facile à effectuer. Il suffit de créer un nouveau répertoire et de copier tous les fichiers contenus dans la disquette du logiciel. Les données sont placées dans le même répertoire.

4. UTILISATION

a) Démarrage

Le programme étant une application DOS, il faudrait par conséquent se mettre sous le contrôle de ce système d'exploitation, se positionner dans le répertoire dans lequel le fichier a été installé et lancer le programme par le mot *Colloc*. Si vous avez déjà travaillé sur un projet, le programme ouvre en arrière plan tous les fichiers associés au dernier projet. Si en revanche c'est la première fois que vous exécutez le programme, celui-ci ne fait rien de spécial mais vous donne la main en vous proposant un menu dans les choix duquel vous trouvez *la création de projet*. La Figure 16 représente le bureau contenant d'une part le menu principal et d'autre part une colonne des fichiers et des paramètres actifs.

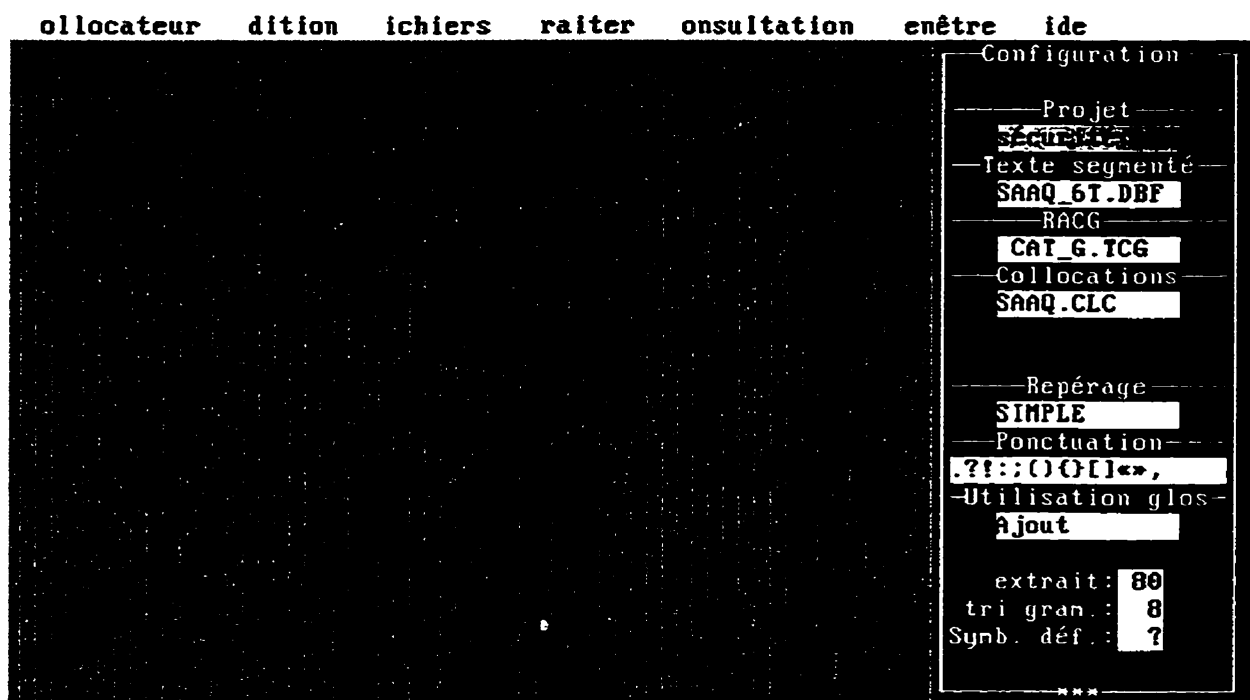


Figure 16 Bureau de MARQUEUR (menu principal et paramètres actifs)

b) Création d'un projet

Pour procéder à une extraction de collocations d'un texte, il est recommandé de commencer par créer un *projet* (dans Fichiers). Un projet est entendu ici comme l'ensemble des fichiers indispensables à l'extraction des collocations. En effet, COLLOCATEUR a besoin de disposer en plus du texte à traiter, d'un jeu de fichiers dont la table des catégories grammaticales. L'écran suivant accessible à partir de *Fichiers* montre les boutons qui permettent de créer ou d'ouvrir les différents fichiers.

En cliquant sur un des boutons relatifs aux fichiers, il sera possible de sélectionner un fichier existant ou d'en créer un nouveau.

Le bouton *Description* vous permet de rattacher une description détaillée au nom du projet. Voir la Figure 17.

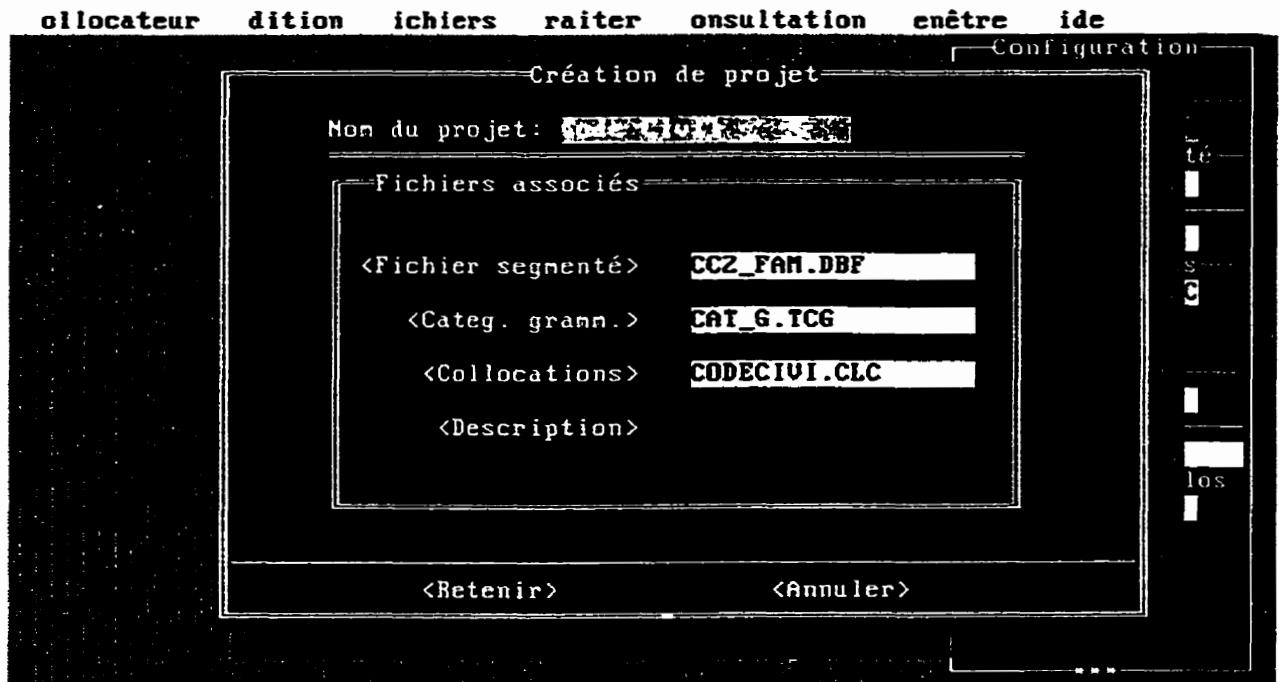


Figure 17 Création de projet

Retenir un projet revient à enregistrer les noms des différents fichiers comme appartenant à un ensemble indissociable. En ouvrant le projet, vous vous épargnez la peine d'ouvrir séparément tous les fichiers indispensables au bon fonctionnement de COLLOCATEUR, mais surtout, il ne sera plus nécessaire de vous rappeler quel fichier fonctionne avec quel autre.

Retenir le projet revient aussi à l'ouvrir immédiatement. Si on veut éviter de l'utiliser tout de suite, on le retient, ensuite, on ouvre un autre projet en allant dans l'option appropriée dans Fichiers.

c) Ouverture d'un projet

Pour ouvrir un projet préalablement créé, il suffit d'atteindre la fenêtre représentée dans la figure 3. Pour ce faire, sélectionner *Ouvrir un projet* dans *Fichiers*.

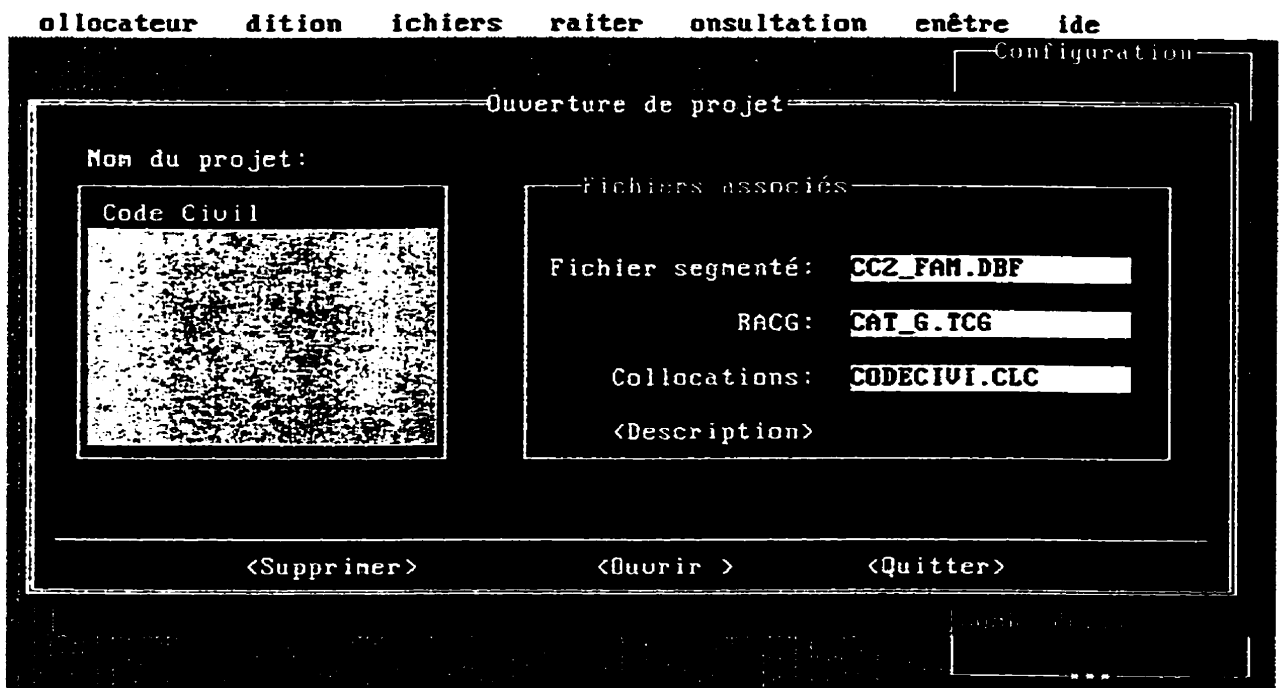


Figure 19 Ouverture de projet

En réalité, cette fenêtre, combinée avec celle de la création de projet, vous permet de gérer différents projets: ouvrir celui qui vous convient pour la tâche présente, effacer celui dont vous n'avez plus besoin.

Conseil: Utiliser le bouton *Description* pour vous rappeler les objectifs du projet, pour autant que vous avez pris la peine de remplir cette rubrique lors de la création du projet.

d) Extraction des collocations

Ainsi que nous l'avons spécifié plus haut (Cf. description, page 156), l'extraction des collocations se fait en fonction de paramètres contrôlés par l'utilisateur. Certains sont accessibles directement sur la fenêtre de la Figure 20. D'autres apparaissent à la demande en cliquant sur les boutons appropriés.

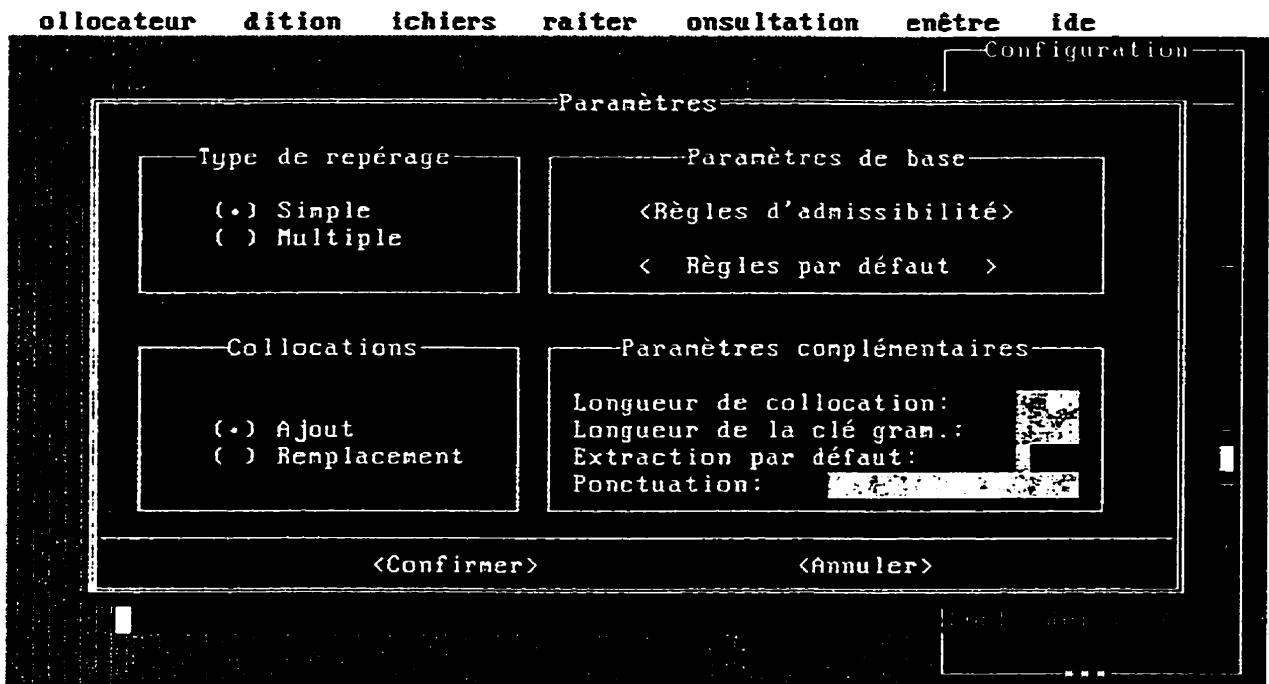


Figure 20 Paramètres

Dans ce qui suit, une description des boutons et des champs de la fenêtre des paramètres montre les différentes fonctions de paramétrage.

(1) Type de repérage

(a) Repérage simple

Le repérage simple signifie que lorsque le programme rencontre une séquence contenant plusieurs collocations, soit enchâssées, soit se chevauchant les unes par rapport aux autres, il ne retient qu'une seule collocation, à savoir la collocation la plus longue.

Exemple:

Véhicule de transport routier à Québec.

Bien que cette séquence contienne en son sein —si les règles d'admissibilité s'y prêtent— les collocations suivantes:

(1) *véhicule de transport*

(2) *véhicule de transport routier*

(3) *véhicule de transport routier à Québec*

(4) *transport routier*

(5) *transport routier à Québec*

COLLOCATEUR ne retiendra que la collocation (3) qui est la plus longue.

(b) Repérage multiple

En revanche, si le programme est paramétré pour un repérage multiple, il retiendra systématiquement toutes les collocations, en l'occurrence les collocations (1) à (5).

(2) Enregistrement des collocations

Les collocations recueillies par COLLOCATEUR sont placées dans le fichier des collocations retenu par l'utilisateur lors de la création du projet. L'option *Ajout* oblige le logiciel à concaténer les données d'une nouvelle extraction à celles retenues lors d'une extraction précédente. Cette option est utile quand on est intéressé par le regroupement de collocations émanant de différents textes. En dehors de cette situation, il serait préférable d'opter pour le remplacement. Une autre stratégie est de multiplier les projets et pour chacun un fichier des collocations différent.

(3) Paramètres de base

Dans les paramètres de base, nous incluons essentiellement les règles d'admissibilité et les règles par défaut que l'utilisateur doit attribuer à chaque catégorie. C'est en effet en fonction de ces règles que le programme COLLOCATEUR peut décider s'il doit retenir le mot rencontré comme membre de la collocation ou non. Bien entendu, ces règles

sont déterminées par rapport à la position du mot dans la chaîne examinée. Le programme a la charge de vérifier dans quelle position se trouve le mot.

(a) Règles d'admissibilité

Pour atteindre la fenêtre de ces paramètres, il suffit de cliquer sur le bouton *Règles d'admissibilité*. Voir la Figure 22.

collocateur dition ichiers raiter onsultation enêtre ide romse

Configuration

Cd_gramm		Def_cat	Paranètres			
			Abr_gram	Debut	Milieu	Fin
20	Nom	no	I	I	I	
30	Verbe	vb	F	F	F	
40	Déterminant	dt	F	I	F	
50	Pronom	pn	F	I	F	
60	Adjectif	aj	I	I	I	
70	Préposition	pr	F	I	F	
80	Adverbe	av	F	I	F	
90	Conjonction	cj	F	I	F	
	Autre	at	F	F	F	

Figure 22 Règles d'admissibilité.

(b) Règles par défaut

Dans COLLOCATEUR, nous venons de le dire, l'extraction se fait sur la base des règles d'admissibilité définies par l'utilisateur pour chaque catégorie. Par ailleurs, on sait maintenant que le texte d'où le logiciel extrait des collocations est un texte marqué. Or, pour une raison ou une autre, un mot peut se retrouver sans traitement. Dans cas là, précisément, il a été prévu d'utiliser des règles d'admissibilité par défaut, elles aussi placées sous le contrôle de l'utilisateur.

ollocateur dition ichiers raiter onsultation enêtre ide

Configuration

Paramètres

Type de repérage Paramètres de base

Valeurs par défaut - RACG
Préciser la règle d'admissibilité en usage
si aucune règle spécifique ne s'applique.

Catégorie grammaticale sans RACG	<input type="checkbox"/> Initiale	<Confirmer>
	<input type="checkbox"/> Médiane	
	<input type="checkbox"/> Finale	
Mot sans analyse grammaticale:	<input type="checkbox"/> Initiale	< Annuler >
	<input type="checkbox"/> Médiane	
	<input type="checkbox"/> Finale	

<Confirmer> <Annuler>

Figure 24 Règles par défaut

(4) Paramètres complémentaires**(a) Longueur de l'extrait**

Théoriquement, on ne connaît pas de limite à la longueur d'une collocation (une collocation peut être exagérément longue sans porter préjudice au grammairien pur; le bât peut par contre blesser au niveau de l'acceptabilité, du style..). COLLOCATEUR offre à l'utilisateur la possibilité d'imposer une limite.

(b) Longueur de la clé grammaticale

Pour chaque mot retenu dans une collocation, le programme enregistre la catégorie grammaticale sous forme d'abréviation. Se constitue ainsi une clé grammaticale (on peut l'appeler aussi *patron*). Celle-ci peut être plus ou moins longue, selon les besoins ou le désir de l'utilisateur.

Ainsi, pour l'exemple:

Véhicule de transport routier à Québec.

on peut obtenir: NoPpNoAJpAj

qui se lit:

Nom + Préposition + Nom + Adjectif + Préposition + Nom

Pour une clé grammaticale de longueur 3, on obtient: NoPpNo

qui naturellement se lit:

Nom + Préposition + Nom

L'exploitation de cette clé peut se faire en appliquant un filtre au fichier des collocations. Cette opération peut être effectuée dans n'importe quel logiciel qui manipule un

fichier au format DBF. Le filtre en question limite l'affichage ou l'impression des collocations à celles disposant de la clé souhaitée.

(c) Extraction par défaut

Il peut être intéressant de garder trace pour chacun des mots du texte ayant servi à l'extraction d'une collocation donnée de l'information indiquant le type de paramètres utilisé (règles d'admissibilité primordiales ou règles par défaut). Le programme COLLOCATEUR enregistre cette information mais seulement quand l'extraction s'est faite sur la base d'un paramètre par défaut (plutôt qu'une règle d'admissibilité primordiale) en utilisant un symbole choisi par l'utilisateur. Cette procédure permet à l'utilisateur de vérifier les mots qui n'ont pas reçu de marquage (les paramètres par défaut s'appliquent dans de telles situations).

(d) Ponctuation

L'extraction d'une collocation ne peut se faire indifféremment de la ponctuation. Généralement, une collocation ne contient pas de ponctuation, à l'exception de la virgule qui pose problème, mais l'utilisateur a toute la latitude d'imposer la frontière qui lui semble convaincante. Si celui-ci mène une étude sur un certain type d'expressions, et que la présence d'une virgule ne le gêne pas, il lui suffira de ne pas retenir la virgule comme frontière possible d'une collocation.

e) Consultation des données

Pour visualiser le contenu d'un fichier, opter pour un des choix du menu *Consulter*. Le premier exemple d'écran de consultation est présenté à la Figure 25, il s'agit du fichier marqué, utilisé dans le projet d'extraction par COLLOCATEUR.

Collocateur	Condition	Fichiers traités	Méthode de consultation	Identificateur	Source
LIURE	10	Non			LIURE
deuxième	50	Adjectif			deuxième
de	50	Préposition			de
la	50	Déterminant			la
premier	10	Non			premier
TITRE	10	Non			TITRE
premier	50	Adjectif			premier
de	50	Préposition			de
deuxième	10	Non			deuxième
chapitre	10	Non			chapitre
premier	50	Adjectif			premier
de	50	Préposition			de
deuxième	10	Non			deuxième
marriage	10	Non			marriage
et	50	Conjonction			et
de	50	Préposition			de
deuxième	50	Déterminant			deuxième
CELEBRATION	10	Non			CELEBRATION
365	50	Autre			365
le	50	Déterminant			le
marriage	10	Non			marriage

Figure 25 Consultation des données: le fichier marqué', extension DBF

La Figure 27 montre une page des collocations extraites à l'aide de COLLOCATEUR. Comme il s'agit d'un fichier DBF (malgré l'extension CLC.), son exploitation est commode.

collocateur dition ichiers raiter onsultation enêtre ide rouse

No collo	Extrait
1	LIURE DEUXIEME DE LA FAMILLE TITRE PREMIER DU MARIAGE CHAPITRE PREMI
2	celibataire compétent et en présence de deux témoins
3	homme et une femme
4	compétents. Il est écarté à cet égard
5	ministère des mariages compétents
6	professeurs ou professeurs adjoints de la Justice Supérieure désignés par
7	ministère des mariages habilités
8	compétents
9	ministère des mariages
10	ministère des mariages et leur compétence
11	ministère des mariages
12	ministère des mariages
13	ministère des mariages
14	ministère des mariages
15	ministère des mariages
16	ministère des mariages
17	ministère des mariages
18	ministère des mariages
19	ministère des mariages
20	ministère des mariages
21	ministère des mariages
22	ministère des mariages
23	ministère des mariages
24	ministère des mariages
25	ministère des mariages
26	ministère des mariages
27	ministère des mariages
28	ministère des mariages
29	ministère des mariages
30	ministère des mariages
31	ministère des mariages
32	ministère des mariages
33	ministère des mariages
34	ministère des mariages
35	ministère des mariages
36	ministère des mariages
37	ministère des mariages
38	ministère des mariages
39	ministère des mariages
40	ministère des mariages
41	ministère des mariages
42	ministère des mariages
43	ministère des mariages
44	ministère des mariages
45	ministère des mariages
46	ministère des mariages
47	ministère des mariages
48	ministère des mariages
49	ministère des mariages
50	ministère des mariages
51	ministère des mariages
52	ministère des mariages
53	ministère des mariages
54	ministère des mariages
55	ministère des mariages
56	ministère des mariages
57	ministère des mariages
58	ministère des mariages
59	ministère des mariages
60	ministère des mariages
61	ministère des mariages
62	ministère des mariages
63	ministère des mariages
64	ministère des mariages
65	ministère des mariages
66	ministère des mariages
67	ministère des mariages
68	ministère des mariages
69	ministère des mariages
70	ministère des mariages
71	ministère des mariages
72	ministère des mariages
73	ministère des mariages
74	ministère des mariages
75	ministère des mariages
76	ministère des mariages
77	ministère des mariages
78	ministère des mariages
79	ministère des mariages
80	ministère des mariages
81	ministère des mariages
82	ministère des mariages
83	ministère des mariages
84	ministère des mariages
85	ministère des mariages
86	ministère des mariages
87	ministère des mariages
88	ministère des mariages
89	ministère des mariages
90	ministère des mariages
91	ministère des mariages
92	ministère des mariages
93	ministère des mariages
94	ministère des mariages
95	ministère des mariages
96	ministère des mariages
97	ministère des mariages
98	ministère des mariages
99	ministère des mariages
100	ministère des mariages

Figure 27 Fichier des collocations extraites.

La Figure 29 montre, quant à elle, la liste des collocations contenant le mot droit, à l'exclusion des autres. Pour obtenir cette liste, il suffit d'ouvrir de fichier des collocations dans un logiciel SGBD (système gestionnaire de bases de données) acceptant d'extension DBF. (Remarquons qu'en l'occurrence notre fichier a l'extension CLC, mais dans sa structure interne, c'est un fichier DBF. Pour l'ouvrir, ne pas oublier l'extension CLC.); appliquer ensuite un filtre, du genre: SET FILTER TO "DROIT" \$ UPPER(EXTRAIT) et enfin commander la liste, par exemple par BROWSE. Naturellement, ce sont là des commandes xBase dont il est facile de trouver l'équivalent.

ystem ile dit atabase ecord rogram indow Ru rowse

Extrait	Extrait
69	les droits fixes par règlement du gouvernement
94	les droits et les devoirs des pères et mères à l'égard de leurs enf
97	la liquidation de leurs droits patrimoniaux
103	La liquidation des droits patrimoniaux
106	droit aux donations
123	le droit de l'époux de bonne foi à des aliments ou à une prestation
124	le droit
127	Le droit
129	plein droit à l'expiration de cette période
131	étatut le droit à des aliments
135	SECTION I DES DROITS ET DES DEVOIRS DES ÉPOUX
136	les mêmes droits et les mêmes obligations
137	les droits énoncés sous ce titre
137	l'exercice de leurs droits et l'accomplissement de leurs devoirs
138	un droit
138	un droit réel
140	L'époux exercent séparément leurs droits
213	atténue à ces autres droits
223	le titulaire arbitraire des droits et obligations
228	la garde d'un enfant au profit d'un usage de la justice familiale

Figure 29 Liste des collocations contenant le mot droit.

A. GUIDE D'UTILISATION DE LOCALEX

1. PRÉAMBULE

La raison d'être de **LOCALEX** est de permettre de mettre en œuvre la démarche de désambiguïsation lexicale automatique avancée dans le cadre de cette thèse. Cela dit, **LOCALEX** peut être vu comme un outil de recherche lexicale répondant à la double condition de la catégorie grammaticale et de la signification de l'unité linguistique en question.

Ce logiciel, fort didactique dans son interface, permet de visualiser les différentes étapes de cette démarche. On ne s'étonnera donc pas de voir que le menu dispose de la possibilité d'enrichir la base des lemmes des acceptions appropriées, et d'associer à ces acceptions les collocations jugées pertinentes.

On notera au passage que **LOCALEX** respecte le même style d'interface que les autres. Cette architecture vise l'harmonisation des trois programmes. Il était essentiel pour nous de produire un programme intuitif, afin que le chercheur-utilisateur ne soit pas trop perturbé par des aspects externes à la problématique qui l'occupe.

2. FONCTIONS

a) Description

Le principe de base implémenté dans **LOCALEX** consiste à comparer chacune des collocations associées à chacune des acceptions du mot recherché au contexte antérieur et ultérieur de l'occurrence de ce même mot dans le texte. C'est un traitement cyclique et séquentiel. Dès qu'une collocation est trouvée, le programme considère que l'acception correspondante est bien celle qui est valide dans le contexte particulier où on se trouve. Le cycle de recherche est repris et en cas de nouvelle occurrence, une nouvelle confrontation des collocations au contexte est effectuée. Selon les cas, le programme attribue une coloration différente au mot examiné.

correspondante est bien celle qui est valide dans le contexte particulier où on se trouve. Le cycle de recherche est repris et en cas de nouvelle occurrence, une nouvelle confrontation des collocations au contexte est effectuée. Selon les cas, le programme attribue une coloration différente au mot examiné.

b) Utilisateurs ciblés

Dans sa version actuelle, **LOCALEX** est surtout utile aux chercheurs qui désirent tester l'apport d'une collocation (ou d'un ensemble de collocations) dans le processus de désambiguïsation automatique.

3. INSTALLATION

L'installation du logiciel est très facile à effectuer. Il suffit de créer un nouveau répertoire et de copier tous les fichiers contenus dans la disquette du logiciel. Les données sont placées dans le même répertoire.

4. UTILISATION

a) Démarrage

Le programme étant une application DOS, il faudrait par conséquent se mettre sous le contrôle de ce système d'exploitation, se positionner dans le répertoire dans lequel les fichiers ont été installés et lancer le programme par le mot **LOCALEX**. Si vous avez déjà travaillé sur un projet, le programme ouvre en arrière plan tous les fichiers associés au dernier projet. Si en revanche, c'est la première fois que vous exécutez le programme, celui-ci ne fait rien de spécial mais vous donne la main en vous proposant un menu (voir Figure 30).

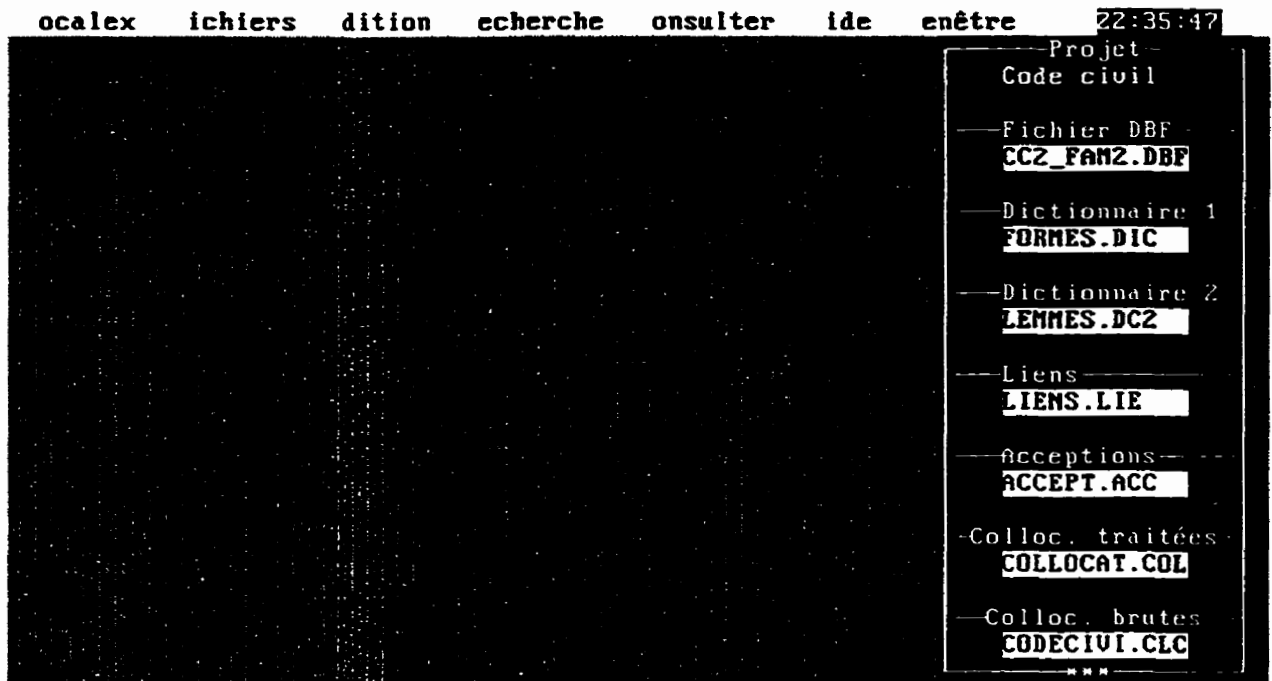


Figure 30 Menu principal

b) Création d'un projet

Pour procéder à la recherche d'une ou de plusieurs unités linguistiques dans le texte, il est recommandé de commencer par créer un *projet* (dans *Fichiers*). Un projet est entendu ici comme l'ensemble des fichiers relatifs à la recherche lexicale. En effet, en plus du texte à explorer, LOCALEX a besoin d'un jeu de fichiers. L'écran de la Figure 32 accessible à partir de *Fichiers* montre les boutons qui permettent d'ouvrir ou de créer les différents fichiers.

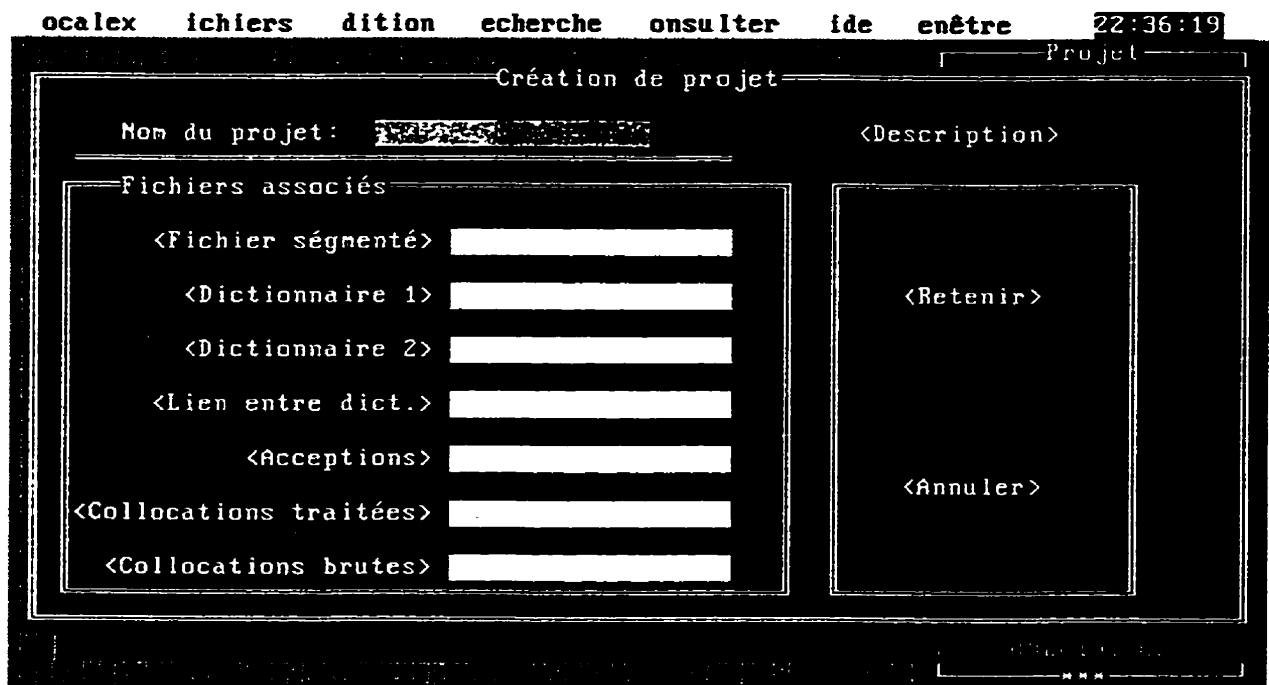


Figure 32 Création de projet

En cliquant sur un des boutons relatifs aux fichiers, il sera possible de sélectionner un fichier existant ou d'en créer un nouveau.

Le bouton *Description* vous permet de rattacher une description détaillée au nom du projet.

Retenir un projet revient à enregistrer les noms des différents fichiers comme appartenant à un ensemble indissociable. En ouvrant le projet, vous vous épargnez la peine d'ouvrir séparément tous les fichiers indispensables au bon fonctionnement de LOCALEX, mais surtout, il ne sera plus nécessaire de vous rappeler quel fichier fonctionne avec quel autre.

Retenir le projet revient aussi à l'ouvrir immédiatement. Si on veut éviter de l'utiliser tout de suite, on le retient pour qu'il soit sauvegardé, ensuite, on ouvre un autre projet en allant dans l'option appropriée dans *Fichiers*.

c) Ouverture d'un projet

Pour ouvrir un projet préalablement créé, il suffit d'atteindre la fenêtre représentée dans la Figure 34. Pour ce faire, sélectionner *Ouvrir un projet* dans *Fichiers*.

En réalité, cette fenêtre, combinée avec celle de la création de projet, vous permet de gérer différents projets: ouvrir celui qui vous convient pour la tâche présente, effacer celui dont vous n'avez plus besoin.

Conseil: Utiliser le bouton *Description* pour vous rappeler les objectifs du projet, pour autant que vous ayez pris la peine de remplir cette rubrique lors de la création du projet.

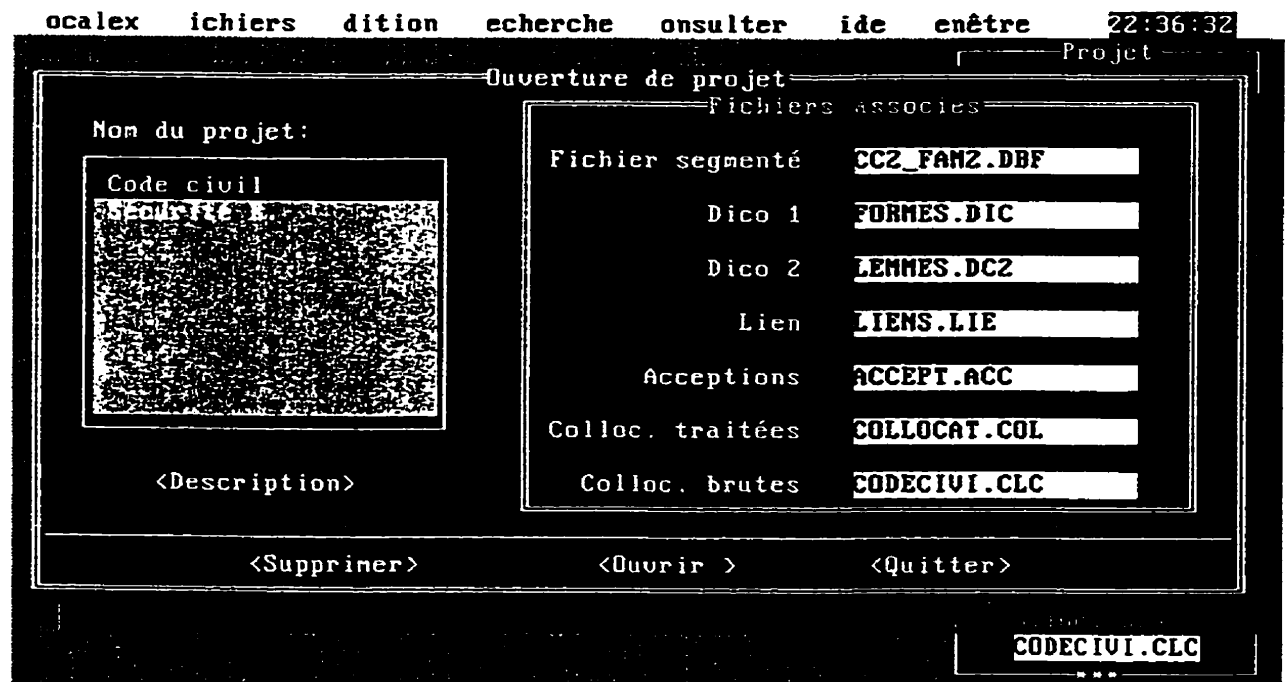


Figure 34 Ouverture de projet

d) Localisation automatique

Pour rechercher les occurrences d'un mot, il suffit de cliquer sur l'option *Recherche* du menu principal. La fenêtre qui s'ouvre vous permettra de saisir le mot à rechercher ainsi que

les paramètres de recherche, notamment le lemme et la catégorie du mot en question, mais aussi l'acception qui vous intéresse.

(1) Paramètres de la recherche

Ces différents paramètres sont recueillis en partant du bouton *mot clé*, comme on le voit à la Figure 36.

The screenshot shows a terminal-style interface for the LOCALEX application. At the top, a menu bar contains the following items: 'ocalex', 'ichiers', 'dition', 'echerche', 'onsulter', 'ide', 'enêtre', and a clock showing '0:15:37'. Below the menu bar, a dialog box titled '**** Grille de requête ****' is displayed. This dialog box contains several input fields:

- '< Mot clé >' with a text input field.
- 'Lemme:' with a text input field.
- 'Catégorie:' with a text input field.
- 'Acception:' with a text input field.
- 'Définition détaillée:' with a larger text input field.

 At the bottom of the dialog box, there are two buttons: '<Rechercher>' and '<Annuler>'. In the bottom right corner of the terminal window, there is a box labeled 'Champ lexical' containing the text 'LEXICAL.LEX' and three asterisks below it.

Figure 36 Paramètres

Le bouton *Mot clé* active une nouvelle fenêtre (Figure 38) dans laquelle vous aurez à sélectionner le mot objet de votre requête dans un index qui se présente sous forme de *liste déroulante*. Pour naviguer dans cette liste, vous utilisez l'ascenseur ou, encore plus commode, les barres horizontales au dessus et au dessous de la colonne des mots énumérés. Les touches *Home* et *End* vous permettent d'aller respectivement au début et à la fin de la liste. Une autre possibilité est de taper le mot dans le champ immédiatement au dessous de la liste et de valider par un retour chariot.

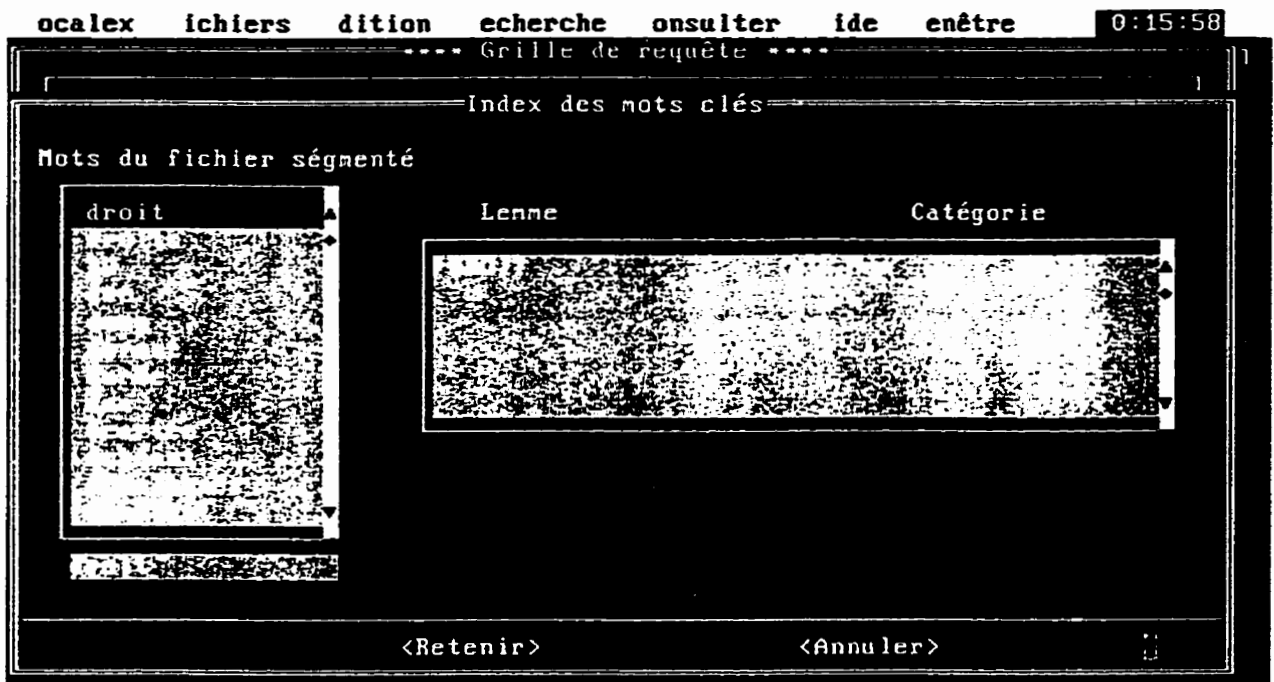


Figure 38 Sélection du mot à rechercher.

En sélectionnant un mot (d'une manière ou d'une autre), vous verrez s'afficher, dynamiquement, le lemme et la catégorie de ce mot. Si le mot dispose dans le texte de plus d'un lemme-catégorie, ils s'afficheront tous. C'est **LOCALEX** qui vérifie dans le texte toutes les occurrences du mot et répertorie les différents cas qui se présentent. Dans la figure donnée en illustration le mot *droit* figure peut-être plusieurs fois dans le texte mais dans tous les cas, c'est un nom. D'une certaine manière, le processus de désambiguïsation automatique commence déjà à ce stade.

Le choix du lemme et de la catégorie signifie que vous voulez localiser les occurrences du mot dont le lemme et la catégorie sont ceux que vous avez sélectionnés, à l'exclusion des autres. Dans l'exemple du mot *droit*, vous pouvez être intéressé par les occurrences de ce mot quand celui-ci est utilisé comme nom et non pas comme adjectif

Cette étape franchie, LOCALEX consulte sa base et vérifie si le mot a plusieurs acceptions. Si tel est le cas, il affichera dans une troisième fenêtre (Figure 40) la liste des acceptions rattachées au mot. Vous aurez à choisir!

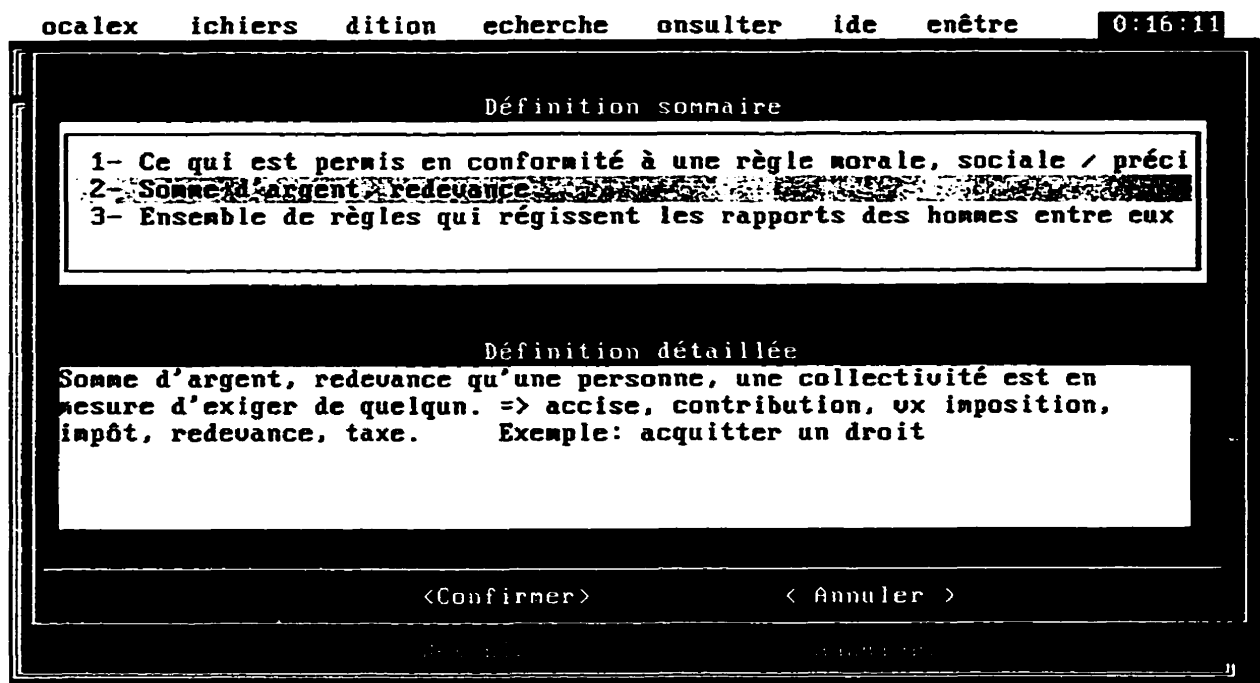


Figure 40 Sélection de l'acception désirée.

Le choix de l'acception signifie que vous désirez localiser les occurrences du mot ayant d'une part le lemme et la catégorie sélectionnés et d'autre part la signification voulue, à l'exclusion des autres occurrences.

(2) Recherche conditionnée

Les différents choix sont immédiatement reportés dans la fenêtre des paramètres initiale. L'illustration en est donnée à la Figure 42.

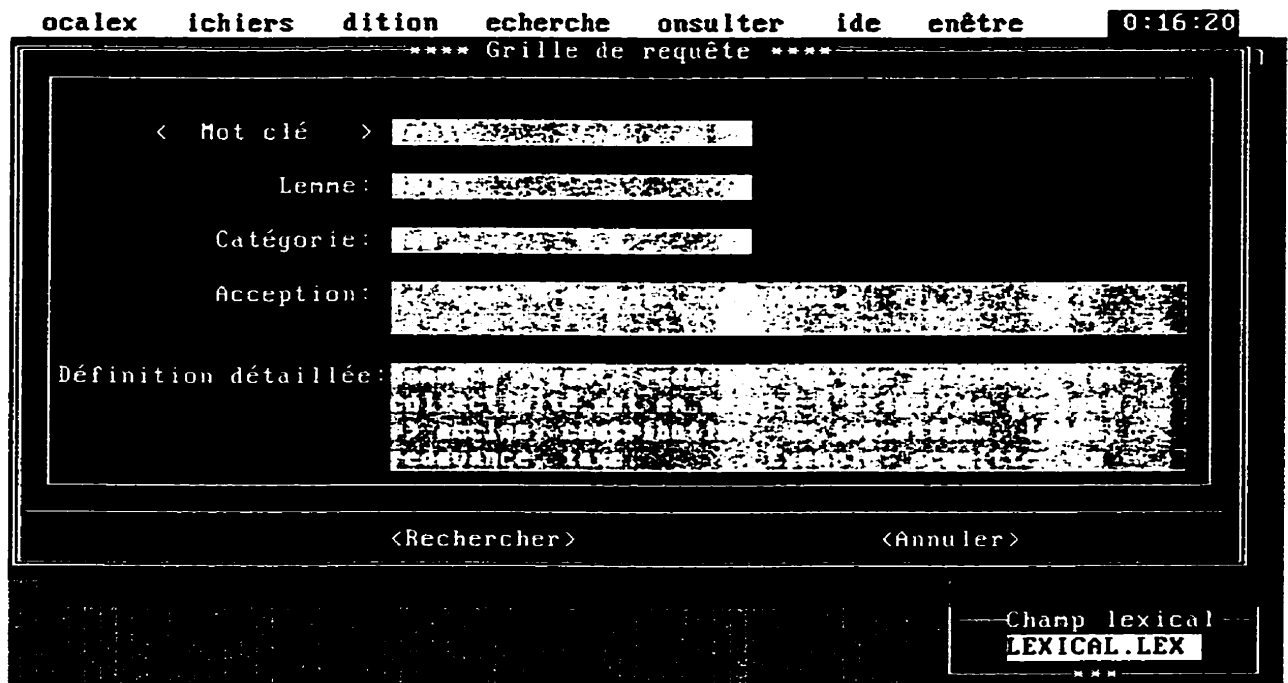


Figure 42 Récapitulation des différents paramètres de recherche dans la grille de requête.

Dès que vous lancez la *Recherche*, LOCALEX se met au travail. Pour chaque occurrence rencontrée, il active un module de vérification du lemme et catégorie, suivi du module de vérification des sens. Le premier type d'information est disponible dans le texte, puisque celui-ci est supposé être traité préalablement. Le deuxième type d'information est l'objet d'une investigation plus laborieuse. LOCALEX vérifie les différentes acceptions rattachées au lemme associé à l'occurrence rencontrée. Pour chacune de ces acceptions, il passe en revue chacune des collocations emmagasinées dans un fichier approprié (fichier des collocations traitées!). Si les différents items de la première collocation sont représentés dans le contexte gauche, LOCALEX considère que c'est l'acception sur laquelle il pointe à ce stade qui est attestée. Sinon, il refait le travail de comparaison pour la collocation suivante. Si après la *visite* de plusieurs mots à gauche (la limite peut être la ponctuation ou un nombre de mots

selon les présupposés théoriques retenus), il se heurte à un échec, **LOCALEX** continue la boucle pour examiner le contexte droit.

LOCALEX va imposer des couleurs différentes à chaque occurrence trouvée (voir la Figure 44). La signification de ces couleurs est présentée dans le **Erreur! Source du renvoi introuvable..**

Couleur	Signification
Rouge / Blanc	<i>1- C'est le bon duo lemme-catégorie, et la bonne acception.</i>
Bleu/Blanc	<i>2- C'est le bon duo lemme-catégorie, mais ce n'est pas la bonne acception.</i>
Noir/Blanc	<i>3- Ce n'est ni le bon duo lemme-catégorie, ni la bonne acception.</i>

*Tableau 1 Légende des couleurs imposées par **LOCALEX**.*

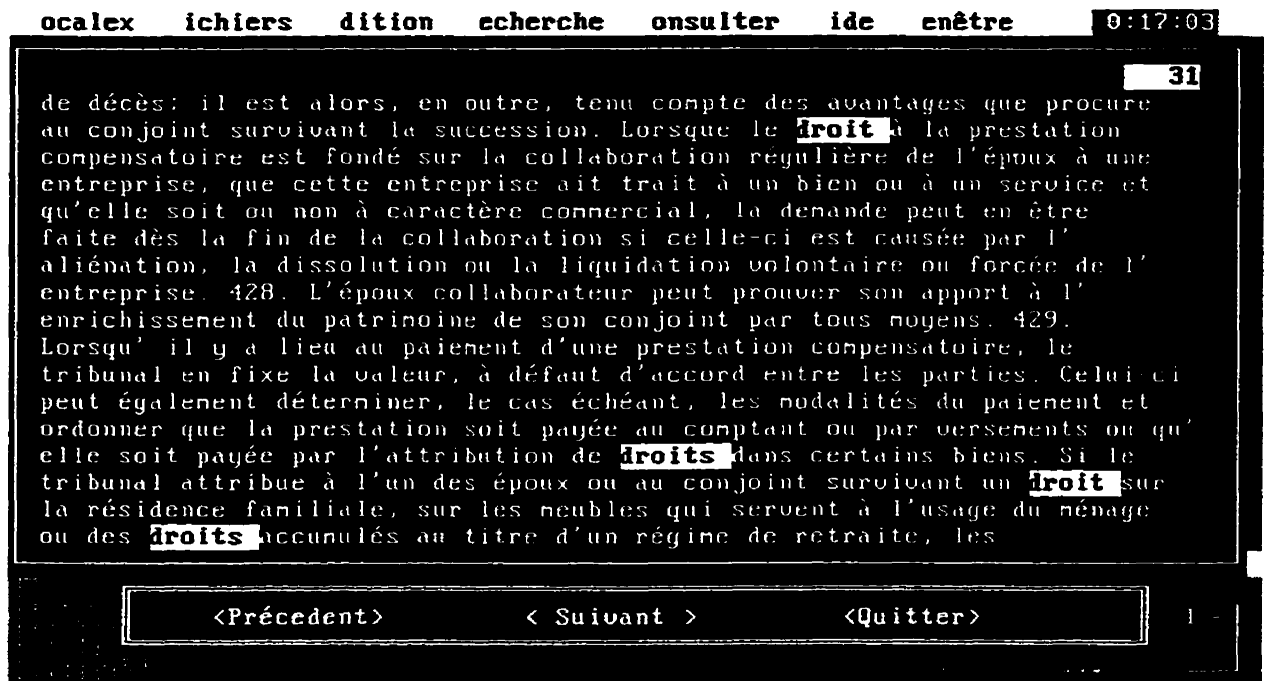


Figure 44 Affichage du texte et des occurrences repérées.

(3) Base des connaissances

LOCALEX étant un logiciel expérimental et loin de disposer d'une base complète, il se devait de fournir les outils pour enrichir progressivement sa base de connaissances. Ce que nous entendons par cette expression (habituellement utilisée en intelligence artificielle) est l'ensemble des données nécessaires à la désambiguïsation lexicale automatique. Outre le dictionnaire des formes, celui des lemmes et catégories, LOCALEX a besoin du fichier des acceptions reliées respectivement aux lemmes, et enfin du fichier des collocations associées respectivement aux acceptions.

C'est une tâche lourde et fastidieuse. Mais la conception que nous avons adoptée réduit avantageusement différentes étapes. Ainsi, pour ne citer qu'un exemple particulièrement saillant, les collocations sont enregistrées dans un format quasi naturel. La seule contrainte est de noter pour chaque item de la collocation le lemme plutôt que la forme.

L'interface d'enrichissement de la base est répartie en trois volets. Dans un premier temps, l'utilisateur devra sélectionner le lemme qu'il souhaite traiter. Cette opération préliminaire peut se faire dans la fenêtre représentée dans la Figure 46.



Figure 46 Sélection du lemme à traiter

La fenêtre suivante (Figure 48) propose de saisir (d'ajouter ou de modifier) la ou les acceptions que l'on veut rattacher au lemme choisi. (La détermination de ces acceptions ainsi que leur hiérarchisation reste à la discrétion de l'utilisateur, en fonction de ses objectifs).

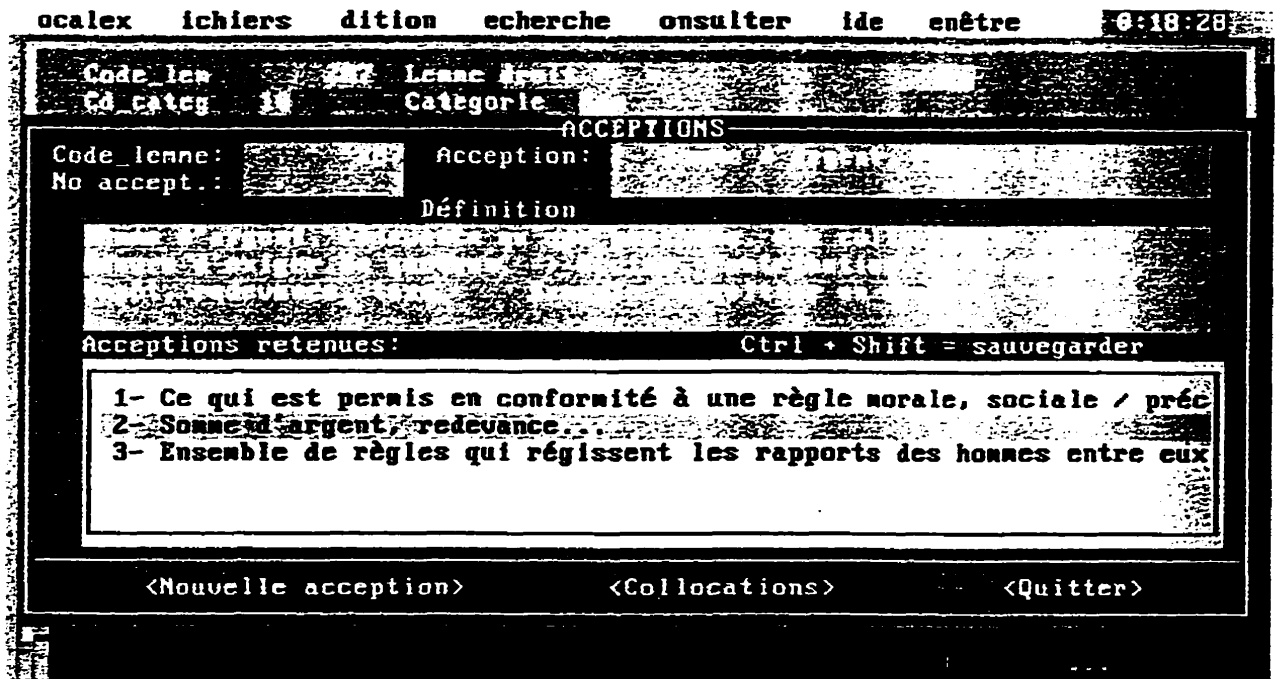


Figure 48 . Ajout d'acceptions et sélection de l'acception à traiter

En demandant (grâce au bouton *Collocations*) la fenêtre suivante (Figure 10), l'utilisateur pourra attribuer à l'acception choisie toutes les collocations qu'il juge pertinentes, il pourra puiser celles-ci dans des dictionnaires spécialisés ou, comme c'est le cas dans le cadre de notre thèse, les relever dans des textes appropriés. Précisément, le logiciel COLLOCATEUR, décrit plus haut dans l'annexe B, répond à ce type de besoins. Comme ce dernier produit un fichier de collocations, grande était la tentation d'intégrer ce fichier au projet de LOCALEX. On fera attention, le fichier des collocations provenant de COLLOCATEUR

porte l'extension **CLC** (c'est l'ensemble des **collocations brutes**, extraites d'un texte donné, et dont le nombre est généralement élevé, pour un texte de 60 pages, on a obtenu 2500 collocations brutes), alors que celui faisant partie de la base de **LOCALEX** porte l'extension **COL**. (C'est le sous-ensemble des collocations associées par l'utilisateur à une acception donnée, ces collocations sont dites **traitées** car effectivement l'utilisateur devrait les nettoyer de tous les bruits, observer quelques règles comme de retenir le lemme au lieu du mot-forme; pour chaque lemme, il faut en placer au moins une dizaine.) Le bouton *Collocations brutes* (voir Figure 50) permet d'afficher le fichier des collocations brutes (provenant de COLLOCATEUR) mais d'une manière sélective, de sorte que l'on ne voit que les collocations contenant le mot dont on est en train de traiter le lemme.

```
ocalex  fichiers  dition  echerche  consulter  ide  enêtre  22:37:55
```

Code_lem	287	Lemme	fixer
Cd_categ	3	Categorie	

ACCEPTIONS

Code_lemme:	287	Acception:	
No accept.:	3		

COLLOCATIONS

Code_lemme:	287	Collocation:	
No accept.:			
No Colloc.:	1		

Collocations retenues:

fixer
paierment
payer
acquitter
acquitté

<Nouvelle collocation> <Collocations brutes> <Quitter>

Figure 49 Saisie, ajout ou modification des collocations

```

ocalex  fichiers  dition  recherche  consulter  ide  enêtre  ro22:38:11
Code_len  Lemme
Cd_categ  Catégorie
ACCEPTIONS
Code_lenne:  Acception:
No accept.:
COLLOCATIONS
No_collo Extrait
les droits fixés par règlement du gouvernement
94 les droits et les devoirs des pères et mères à l' égard de leurs
97 la liquidation de leurs droits patrimoniaux
103 la liquidation des droits patrimoniaux
106 droit aux donations
123 le droit de l' époux de bonne foi à des aliments ou à une prestat
124 le droit
127 Le droit
129 plein droit à l' expiration de cette période
131 éteint le droit à des aliments
136 SECTION I DES DROITS ET DES DEVOIRS DES ÉPOUX
137 les mêmes droits et les mêmes obligations
141 ses droits civils sous ce nom

```

Figure 50 Liste de collocations brutes provenant de COLLOCATEUR.

e) Consultation des fichiers

Pour visualiser le contenu d'un fichier, opter pour un des choix du menu *Consulter*. Le premier exemple d'écran de consultation est présenté à la Figure 51, il s'agit du fichier du texte marqué.

ocalex fichiers dition echerche onsulter ide enêtre ro 0:14:58

Code_mot	Mot	Code_lem	Lemme	Categorie
1	LIURE	1	liure	Non
1	LIURE	2	liurer	Verbe
5	FAMILLE	8	famille	Non
7	PREMIER	11	premier	Adjectif
9	MARIAGE	14	mariage	Nom
10	CHAPITRE	15	chapitre	Nom
11	ET	16	et	Conjonction
12	SA	17	sa	Déterminant
13	CÉLÉBRATION	18	célébration	Nom
14	365	19	365	Autre
15	Le	20	le	Déterminant
16	doit	22	devoir	Verbe
17	être	23	être	Verbe
18	contracté	27	contracter	Verbe
19	publiquement	28	publiquement	Adverbe
21	un	30	un	Déterminant
22	célébrant	31	célébrant	Nom
22	célébrant	32	célébrant	Verbe
23	compétent	33	compétent	Adjectif
24	en	34	en	Préposition

Figure 52 Dictionnaire des formes et dictionnaire des lemmes reliés.

Le troisième exemple d'écran de consultation est présenté à la Figure 53, il s'agit du fichier des acceptions.

Localex Fichiers Edition Recherche Consulter Aide Fenêtre DXo 0:14:01

#16

Code_lemme	Num_sens	Sens
287	1	1- Ce qui est permis en conformité à une règle morale.
2985	1	côté opposé au côté gauche
147	1	requête
1202	1	maison
2401	1	construction
2231	1	accrochage
448	1	Moyen de transport
2402	1	usine
577	1	manière
680	1	pouvoir
680	2	institution
2137	1	Interdire
2508	1	nuisible
2508	2	mortel
2137	2	Limiter
448	2	Autre chose
287	3	2- Somme d'argent, redevance.
287	4	3- Ensemble de règles qui régissent les rapports des hom
719	1	1- préjudice
719	2	1- somme qui rémunère un créancier pour l'usage de son a

Figure 53 Fichier des acceptions

Le dernier exemple (Figure 54) montre un écran de consultation des collocations traitées.

Localex Fichiers Edition Recherche Consulter Aide Fenêtre DXo 0:14:13

#17

Code_lemme	Num_sens	Num_expr	Expression
287	1	1	de circuler
287	1	2	accorder
2401	1	1	de plaque
448	1	1	utilisation des
448	1	2	routier
440	1	3	automobile
448	1	4	affecté
448	1	5	de commerce
287	1	3	nettre
287	1	4	jouir
2137	1	1	paragraphe abrogé
2508	1	1	danger d'éclatement
2508	1	2	constituer un danger
2508	1	3	rendent apte à conduire ce véhicule sans constit
448	1	6	de promenade
448	1	7	de service
440	1	8	non conforme
448	1	9	notorisé
287	3	1	fixer
287	3	2	paient

Figure 54 Fichier des collocations

D. MOTS AMBIGUS ET LEUR CONTEXTE**1. Corpus Sécurité Routière du Québec****a) Le mot droit (s)**

Droit (nom)		
Signification 1	Signification 2	Signification 3
Ce qui est permis en conformité à une règle morale, sociale / précise. - Possibilité, pouvoir, qualité, autorisation, permission. - Faculté, habileté, prérogative, privilège.	Somme d'argent, redevance qu'une personne, une collectivité est en mesure d'exiger de quelqu'un. → accise, contribution, vx: imposition, impôt, redevance, taxe. Exemple: acquitter un droit.	Ensemble de règles qui régissent les rapports des hommes entre eux. → légalité, légitimité, justice, morale. → juridique.

Collocations emmagasinées

Signification 1	Signification 2	Signification 3
D: de propriétaire D: de jouir D: accordé D: mettre en circulation D: de circuler D: de propriété G: cession du G: exercice de D: de reprise D: de obtenir G: suspension du G: suspendre D: de personne D: et obligation G: accorder D: suspendu	D: fixé G: payer G: acquitter G: paiement des	D: commun D: romain D: écrit G: qui de

Occur	Contexte	Lclx	Expert humain	Explication de l'expert
150	un véhicule routier en vertu d'un titre assorti d'une condition ou d'un terme qui lui donne le <u>droit</u> (150) d'en devenir <u>propriétaire</u> , (...)	Rouge	Décision heureuse	Adéquation discontinue
165	(...) , ou en vertu d'un titre qui lui donne le <u>droit</u> (165) d'en jouir comme <u>propriétaire à charge de rendre</u> .	Rouge	Décision heureuse	Adéquation discontinue
2304	la demande de la Société ou d'un agent de la paix, la preuve de cette immatriculation; 4 le même <u>droit</u> 2304 est <u>accordé</u> aux étudiants, aux coopérants ou aux stagiaires du Québec qui séjournent au lieu du domicile de cet étudiant, coopérant ou stagiaire.	Rouge	Décision heureuse	Adéquation discontinue
2358	Pour obtenir l'immatriculation d'un véhicule routier et le <u>droit</u> (2358) de mettre ce véhicule en circulation, le <u>propriétaire</u> de ce dernier doit: 1 satisfaire aux conditions et aux formalités établies par règlement; 2 (paragraphe abrogé);	Rouge	Décision heureuse	Adéquation disloquée
2392	3 payer les frais fixés par règlement, les <u>droits</u> (2392) fixés par règlement et revalorisés, le cas échéant, conformément à l'article 151.4 de la Loi sur l'assurance automobile (chapitre A-25), la contribution d'assurance fixée en vertu des articles 151.1 et 151.2 de cette loi et revalorisée, le cas échéant,	Noir	décision heureuse	Inadéquation justifiée Le contexte ne contient, à juste titre, aucune collocation favorable à l'acception recherchée Signific. différente
3210	. <u>Droit</u> (3210) de <u>circuler</u> .	Rouge	décision heureuse	Adéquation continue
3218	Pour conserver le <u>droit</u> (3218) de circuler avec un véhicule routier immatriculé, le <u>propriétaire</u> de celui-ci, à moins d'en être exempté par règlement, doit, selon la fréquence prévue par règlement, payer à la Société les frais fixés par règlement,...	Rouge	décision heureuse	Adéquation disloquée
3255	(...) le propriétaire de celui-ci, à moins d'en être exempté par règlement, doit, selon la fréquence prévue par règlement, payer à la Société les frais fixés par règlement, les <u>droits</u> 3255	Noir	Décision heureuse	Inadéquation justifiée

Occur	Contexte	Lclx	Expert humain	Explication de l'expert
	fixés par règlement et revalorisés, le cas échéant, conformément à l'article 151.4			
3405	Il ne sera alors pas tenu de payer les droits 3405 et les (...)	Noir	Décision heureuse	Inadéquation justifiée
3543	Circulation. Il doit alors acquitter les droits 3543 et les frais, la contribution d'assurance et les frais supplémentaires prévus par règlement, conformément aux conditions et modalités prévues par règlement.	Noir	Décision heureuse	Inadéquation justifiée
4329	Lors de la cession du droit 4329 de propriété d'un véhicule routier entre des parties qui ne sont pas des commerçants, le cédant doit remettre à la Société le certificat et la plaque d'immatriculation délivrés pour ce véhicule après avoir endossé le certificat et le nouvel acquéreur doit demander une nouvelle immatriculation.	Rouge	Décision heureuse	Adéquation continue 2 eme colloc.
4454	Cession à un commerçant. 42. Lorsque la cession du droit 4454 de propriété d'un véhicule routier est faite à un commerçant, le cédant qui n'acquiert pas un nouveau véhicule doit remettre au commerçant le certificat d'immatriculation du véhicule, après l'avoir endossé, et transmettre à la Société la plaque d'immatriculation de ce véhicule.	Rouge	Décision heureuse	Adéquation continue 2 eme colloc.
4520	Si la cession du droit 4520 de propriété d'un véhicule routier est faite à ...	Rouge	Décision heureuse	Adéquation continue 2 eme colloc.
4597	Les articles 40 à 43 ne s'appliquent pas à la cession du droit 4597 de propriété d'un véhicule routier qui fait l'objet d'un contrat de location d'une durée d'au moins un an.	Rouge	Décision heureuse	Adéquation continue
4725	. 46. Sous réserve de l'article 17, la personne qui devient propriétaire d'un véhicule routier, par suite d'un décès, d'une donation, d'un partage, d'une liquidation, d'une faillite, de l'exercice	Rouge	Décision heureuse	Adéquation discontinue

Occur	Contexte	Lclx	Expert humain	Explication de l'expert
	d'un droit 4725 de reprise, d'une cession complète d' une entreprise ou d'une vente en justice, doit remettre à la Société le certificat d'immatriculation et lui demander une nouvelle immatriculation du véhicule.			
6191	règlement. Elle doit également pour obtenir un permis, sauf dans les cas prévus par règlement, payer à la Société les droits 6191 fixés par règlement et revalorisés, le cas échéant, conformément à l'article 151.4 de la Loi sur l'assurance automobile ainsi que la contribution d'assurance fixée en vertu des articles 151 et 151.2 de cette loi et revalorisée, le cas échéant, conformément à l'article 151.4 de cette loi.	Noir	Décision heureuse	Inadéquation justifiée Signific. différente
6805	76. Une personne dont le permis a été révoqué ou dont le droit 6805 d'en obtenir un a été suspendu par suite d'une déclaration de culpabilité pour une infraction à l'une des dispositions du Code criminel (...)	Rouge	Décision heureuse	Adéquation discontinue
6886	. Aucun permis ne peut lui être délivré à moins qu'il ne se soit écoulé, depuis la date de révocation du permis ou de suspension du droit: 6886 1 une période d'un an, si la	Rouge	Décision heureuse	Adéquation continue
7104	78. Une personne dont la classe de son permis autorisant la conduite d' un taxi a été révoquée ou dont le droit 7104 d'obtenir un permis de cette classe a été suspendu par suite d'une déclaration de culpabilité pour un acte criminel relié à l'exploitation du transport par taxi peut obtenir de la Société un permis d'une classe autre que celle autorisant la conduite d'un taxi.	Rouge	Décision heureuse	Adéquation continue LEMME
7180	inaptitude. . 79. Une personne dont le permis a été révoqué ou dont le droit 7180 d'en obtenir un a été suspendu par suite de l' accumulation de points d'inaptitude doit, pour en obtenir un, se	Rouge	Décision heureuse	Adéquation discontinue

Occur	Contexte	Lclx	Expert humain	Explication de l'expert
	conformer aux conditions et formalités établies par règlement. . Délai.			
7240	Aucun permis ne peut lui être délivré à moins qu'il ne se soit écoulé, depuis la date de révocation du permis ou de suspension du droit 7240 d'en obtenir un: 1 une période de trois mois, si la personne ne s'est vue imposer ni révocation ni suspension au cours des deux années qui précèdent cette révocation ou suspension; 2 une période	Rouge	Décision heureuse	Adéquation discontinue
7410	80.1. Une personne dont le permis a été révoqué ou dont le droit 7410 d'en obtenir un a été suspendu conformément à l'article 187.1 doit, pour en obtenir un, se conformer aux conditions et formalités établies par règlement.	Rouge	Décision heureuse	Adéquation discontinue
7468	. Aucun permis ne peut lui être délivré à moins qu'il ne se soit écoulé, depuis la date de révocation du permis ou de <u>suspension du droit</u> : 7468 1 une période de trois mois, si la personne ne s'est vue imposer ni révocation ni suspension au cours des deux années qui précèdent cette révocation ou suspension;	Rouge	Décision heureuse	Adéquation continue
7644	80.3. Une personne dont la classe de permis a été révoquée ou dont le droit 7644 d'obtenir une telle classe a été suspendu, conformément à l'article 187.2, doit pour obtenir cette classe se conformer aux conditions et formalités établies par règlement. . Période d'attente. .	Rouge	Décision heureuse	Adéquation continue
7711	. Cette classe de permis ne peut lui être délivrée à moins qu'il ne se soit écoulé depuis la date de la révocation de sa classe de permis ou de la suspension de son droit : 77111 une période <u>de</u> trois mois, si la <u>personne</u> ne s'est vue imposer aucune révocation ou suspension de classe pour une	Rouge	Décision heureuse	Erreur algo La ponctuation du mot recherchée n'est pas examinée disloquée

Occur	Contexte	Lcix	Expert humain	Explication de l'expert
	infraction à une même disposition, au cours des deux années qui précèdent cette révocation ou suspension; 2 une période de six mois, si ...			
8369	demandée; 3 fait l'objet d'une interdiction de conduire, d'une révocation ou d'une suspension de son permis ou de la classe de celui-ci ou d'une suspension de son droit 8369 d'obtenir un permis ou une classe de celui-ci en vigueur ou imposée mais non encore en vigueur, y compris en vertu d'une loi du Canada, d'une autre province ou d'un territoire du Canada ou d'une loi d'un État des États-Unis;	Rouge	Décision heureuse	Adéquation continue
8710	II: DROITS 8710 ET OBLIGATIONS DES NON-RÉSIDENTS .	Rouge	Décision heureuse	Adéquation continue
8770	1 il est titulaire d' un permis de conduire valide délivré par une autre autorité administrative <u>accordant</u> le même droit 8770 aux résidents du Québec; 2 le permis de conduire délivré par cette autre autorité administrative l' autorise à conduire un véhicule de la catégorie qu'il conduit au Québec;	Rouge	Décision heureuse	Adéquation discontinue LEMME
8823	3 il respecte les conditions dont son permis de conduire est assorti; 4 il ne fait pas l'objet d'une suspension du droit 8823 d'obtenir un permis délivré par la Société ou une classe de celui-ci. 1986, c. 91, a.	Rouge	Décision heureuse	Adéquation continue Autre collocation
9155	90. Le titulaire d'un permis de conduire valide délivré au Canada, par une autre autorité administrative peut, s'il s'établit au Québec, échanger sans examen ce permis contre un permis de conduire délivré par la Société, sur paiement des droits 9155 et des frais fixés par règlement et du montant fixé en vertu de l'article 151 de la Loi sur l'assurance automobile (chapitre A-25).	Noir	Décision heureuse	Inadéquation justifiée
9258	La Société peut, sur paiement des droits 9258 et des frais fixés par	Noir	Décision heureuse	Inadéquation justifiée

Occur	Contexte	Lcix	Expert humain	Explication de l'expert
	règlement ainsi que de la contribution d'assurance fixée en vertu des articles 151 et 151.2 de la Loi sur l'assurance automobile (chapitre A-25), délivrer un permis de conduire (...)			
9359	(...) et ne fait pas l'objet d'une interdiction de conduire, d'une révocation ou d'une suspension de son permis ou de la classe de celui-ci ou d'une suspension de son droit 9359 d'obtenir un permis ou une classe de celui-ci en vigueur ou imposée mais non encore en vigueur, y compris en vertu d'une loi du Canada, d'une autre province ou d'un territoire du Canada ou d'une loi d'un État des États-Unis. . Examen requis. . Toutefois, la Société 34	Rouge	Décision heureuse	Adéquation continue
9508	91. Le titulaire d'un permis de conduire valide délivré à l'extérieur du Canada peut, s'il s'établit au Québec, échanger ce permis contre un permis de conduire délivré par la Société à la condition de réussir l'examen de compétence et d'acquitter les droits 9508 et les frais fixés par règlement et le montant fixé en vertu de l'article 151 de la Loi sur l'assurance automobile (chapitre A-25). . Remise du permis. .	Noir	Décision heureuse	Inadéquation justifiée Signific. différente
10056	93.1. Le titulaire d'un permis de conduire doit, selon la fréquence prévue par règlement, payer à la Société les frais fixés par règlement, les droits 10056 fixés par règlement et revalorisés, le cas échéant, conformément à l'article 151.4 de la Loi sur l'assurance automobile ainsi que la contribution d'assurance fixée en vertu de l'article 151 de cette loi et revalorisée, le cas échéant, conformément à l'article 151.4 de	Noir	Décision heureuse	Inadéquation justifiée Signific. différente
11414	. Pour l'application des articles 105 et 106, sanction s'entend de la révocation ou de la suspension soit du	Rouge	Décision heureuse	Adéquation continue

Occur	Contexte	Lclx	Expert humain	Explication de l'expert
	permis autorisant la conduite d' un véhicule routier, soit d'une classe d'un tel permis, ainsi que de la suspension du droit 11414 <u>d'obtenir</u> un tel permis ou une telle classe. Sont aussi visées la révocation et la suspension prononcées par			2 eme colloc.
11877	111. La Société administre un système de points d'inaptitude établi par règlement, en vertu duquel elle révoque un permis ou suspend le droit 11877 <u>d'en obtenir</u> un. Attribution des points d'inaptitude. Les points d'inaptitude sont prescrits: 1 pour une infraction commise à l'	Rouge	Décision heureuse	Adéquation discontinue
12775	: 1 dans les deux ans qui précèdent la révocation ou la suspension qui donne lieu à la demande de permis restreint, le permis du requérant a déjà été révoqué ou suspendu ou son droit 12775 <u>d'obtenir</u> un permis a déjà été suspendu par suite d'une déclaration de culpabilité (...)	Rouge	Décision heureuse	Adéquation continue
12848	2 au moment de la présentation de la requête, le permis du requérant ou son droit 12848 <u>d'en obtenir</u> un fait l'objet d'une suspension en vigueur ou imposée et non encore en vigueur; 3 la sanction	Rouge	Décision heureuse	Adéquation discontinue
12977	; 6 le motif invoqué pour obtenir un permis restreint est relié à l'exploitation du transport par taxi et, au moment de la présentation de la requête, la classe du permis du requérant autorisant la conduite d' un taxi est révoquée ou son droit 12977 <u>d'obtenir</u> un permis de cette classe est suspendu par suite d'une déclaration de culpabilité pour un acte criminel relié à l'exploitation du transport par taxi;	Rouge	Décision heureuse	Adéquation continue
13180	. Si, lors de la réception par la Société d'une ordonnance de délivrer un permis restreint, l'article 121 est applicable à la personne visée dans cette ordonnance,	Rouge	Décision heureuse	Adéquation disloquée

Occur	Contexte	Lclx	Expert human	Explication de l'expert
	ou, si entre la date fixée pour la présentation de la requête et celle de la délivrance du permis restreint, le droit 13180 <u>de</u> cette personne <u>d'obtenir</u> un permis fait l'objet d'une suspension en vigueur ou imposée et non encore en vigueur, la Société doit en aviser le juge qui a rendu l'ordonnance.			
14503	. Quiconque contrevient au premier alinéa de l'article 105 commet une infraction et est passible d'une amende de 300 à 600 , si son permis ou son droit 14503 <u>d'en obtenir</u> un fait l'objet d'une révocation ou d' une suspension pour un motif autre que ceux visés à l'article 180.	Rouge	Décision heureuse	Adéquation discontinue
14574	. Quiconque contrevient au premier alinéa de l'article 105 commet une infraction et est passible d'une amende de 600 à 2 000 , si son permis ou son droit 14574 <u>d'en obtenir</u> un fait l'objet d'une révocation ou d' une suspension conformément à l'article 180.	Rouge	Décision heureuse	Adéquation discontinue
17487	. Droit 17487 suspendu.	Rouge	Décision heureuse	Adéquation continue
17513	Lorsque la personne n'est pas titulaire d' un permis d'apprenti-conducteur, d'un permis probatoire ou d'un permis de conduire, son droit 17513 <u>de l'obtenir</u> est suspendu.	Rouge	Décision heureuse	Adéquation discontinue
17565	Toute déclaration de culpabilité pour une infraction visée à l'article 180 entraîne une révocation du permis ou une suspension du droit 17565 <u>d'en obtenir</u> un distincte pour chaque déclaration.	Rouge	Décision heureuse	Adéquation discontinue
17633	. Toutefois, une déclaration de culpabilité pour plus d'une infraction prévue à l'article 253, au paragraphe (5) de l'article 254 ou aux paragraphes (2) ou (3) de l' article 255 du Code criminel (Lois révisées du Canada (1985), chapitre C-46) n'entraîne qu'une seule révocation du permis ou	Rouge	Décision heureuse	Adéquation discontinue

Occur	Contexte	Clx	Expert humain	Explication de l'expert
	une seule suspension du droit 17633 d'en obtenir un, lorsque les infractions se rapportent à un même événement.			
17672	. La révocation d'un permis ou la suspension du droit 17672 d'en obtenir un prévue à l'article 181 est maintenue même si la personne déclarée coupable d'une infraction	Rouge	Décision heureuse	Adéquation discontinue
17747	. Lorsqu' une personne est déclarée coupable d'un acte criminel relié l'exploitation du transport par taxi, la classe de son permis autorisant la conduite d'un taxi est révoquée et son droit 17747 d'obtenir un permis de cette classe est suspendu pour cinq ans, à moins qu' un pardon n'ait été obtenu.	Rouge	Décision heureuse	Adéquation continue
17848	. Lorsque la personne déclarée coupable d'un acte criminel relié à l'exploitation du transport par taxi n'est pas titulaire d'un permis d'une classe autorisant la conduite d'un taxi, son droit 17848 d'en obtenir un est suspendu pour cinq ans, à moins qu'un pardon n'ait été obtenu.	Rouge	Décision heureuse	Adéquation discontinue
17931	. Dès que le nombre de points d'inaptitude inscrits au dossier d'une personne est égal ou supérieur à celui prévu par règlement, la Société doit révoquer le permis de conduire de cette personne ou suspendre, si elle n' est plus titulaire d'un permis de conduire, son droit 17931 de l'obtenir (...)	Rouge	Décision heureuse	Adéquation discontinue
18020	(...), la Société doit révoquer le permis de conduire et le permis d'apprenti-conducteur de cette personne ou suspendre, si elle n'est plus titulaire d'un permis de conduire ou d'un permis d' apprenti-conducteur, son droit 18020 de l'obtenir .	Rouge	Décision heureuse	Adéquation discontinue
18117	En outre, si postérieurement à la date où l'infraction fut commise, cette personne cesse d'être titulaire d'un permis, elle doit suspendre son droit 18117 d'en obtenir un.	Rouge	Décision heureuse	Adéquation discontinue

Occur	Contexte	Lcix	Expert humain	Explication de l'expert
18217	VIII.1 de cette personne ou suspendre, si celle-ci n'est pas titulaire d'une telle classe, son droit 18217 <u>d'en obtenir</u> une.	Rouge	Décision heureuse	Adéquation discontinue
18235	. La classe ainsi révoquée ou dont le droit 18235 <u>de l'obtenir</u> a été suspendu doit correspondre à celle autorisant la conduite du véhicule automobile que conduisait la personne au moment où l'infraction fut commise.	Rouge	Décision heureuse	Adéquation discontinue
19365	(...) , si elle n'est pas titulaire d'un permis autorisant la conduite d'un cyclomoteur, d'un permis d'apprenti-conducteur ou d'un permis probatoire, son droit 19365 <u>de l'obtenir</u> .	Rouge	Décision heureuse	Adéquation discontinue
19448	(...) , la Société doit suspendre son permis d'apprenti-conducteur et son permis probatoire ou son permis de conduire ou suspendre, si elle n'est pas titulaire d'un permis d'apprenti-conducteur, d'un permis probatoire ou d'un permis de conduire, son droit 19448 <u>de l'obtenir</u> .	Rouge	Décision heureuse	Adéquation discontinue
19724	194. La Société doit suspendre le permis d'apprenti-conducteur et le permis probatoire ou le permis de conduire d'une personne ou suspendre, si elle n'est pas titulaire d'un permis d'apprenti-conducteur, d'un permis probatoire ou d'un permis de conduire, son droit 19724 <u>de l'obtenir</u> , lorsqu'elle reçoit l'avis prévu à l'article 364 du Code de procédure pénale (chapitre C-25.1).	Rouge	Décision heureuse	Adéquation discontinue
19797	. Droit 19797 <u>suspendu</u> .	Rouge	Décision heureuse	Adéquation continue
19826	195. Lorsque la période de validité d'un permis se termine avant la fin de la période de la suspension dont celui-ci faisait l'objet, le droit 19826 <u>d'obtenir</u> un permis est alors suspendu pour la durée de la période de suspension non expirée.	Rouge	Décision heureuse	Adéquation continue
19879	La Société doit suspendre le permis	Rouge	Décision	Adéquation

Occur	Contexte	Lcix	Expert humain	Explication l'expert
	restreint d'une personne si, après la date où il a été délivré, le <u>droit</u> 19879 de cette personne d' <u>obtenir</u> un permis fait l'objet d'une suspension en vigueur ou imposée et non encore en vigueur. Cette		heureuse	disloquée
19912	. Cette suspension demeure en vigueur tant que la suspension du <u>droit</u> 19912 d' <u>obtenir</u> un permis n'est pas levée.	Rouge	Décision heureuse	Adéquation continue
20066	vertu de cette loi, elle suspend le permis d'apprenti-conducteur et le permis probatoire ou le permis de conduire du propriétaire et du conducteur de ce véhicule ou suspend, s'ils ne sont pas titulaires d'un permis d'apprenti-conducteur, d'un permis probatoire ou d'un permis de conduire, leur <u>droit</u> 20066 de l' <u>obtenir</u> . En outre, elle doit interdire la remise en circulation de tout véhicule routier immatriculé au nom de l'une ou l'autre de ces personnes.	Rouge	Décision heureuse	Adéquation discontinue
20740	200. Sur réception d'un avis à cet effet du créancier concerné, la Société suspend le permis d'apprenti-conducteur et le permis probatoire ou le permis de conduire ou le <u>droit</u> 20740 de les <u>obtenir</u> et interdit la remise en circulation du véhicule routier immatriculé au nom du débiteur qui n'a pas satisfait, dans le délai d'exécution, à un jugement qui prononce au Canada une condamnation définitive:	Rouge	Décision heureuse	Adéquation discontinue
21660	78 . <u>Droit</u> 21660 <u>suspendu</u> .	Rouge	Décision heureuse	Adéquation continue
21700	Lorsque la période de validité d'un permis d'école de conduite ou d'un permis d'enseignement se termine avant la fin de la période de la suspension dont ce permis faisait l'objet, le <u>droit</u> 21700 d' <u>obtenir</u> un tel permis est alors suspendu pour la durée de la période de suspension non	Rouge	Décision heureuse	Adéquation continue

Occur	Contexte	Lcix	Expert humain	Explication de l'expert
	expirée.			
21878	. La Société doit suspendre une licence de commerçant ou de recycleur ou le droit 21878 d'obtenir (...)	Rouge	Décision heureuse	Adéquation continue
22242	. Droit 22242 suspendu.	Rouge	Décision heureuse	Adéquation continue
22277	Lorsque la période de validité d'une licence de commerçant ou de recycleur se termine avant la fin de la période de la suspension dont cette licence faisait l'objet, le droit 22277 d'obtenir une telle licence est alors suspendu pour la durée de la période de suspension non expirée.	Rouge	Décision heureuse	Adéquation continue
23332	: 1 à l'avant, de deux feux de gabarit jaunes, placés à la même hauteur et à au plus 150 mm des extrémités supérieures droite 23332 et gauche du véhicule;	Bleu	Décision heureuse	Mauvaise catégorie
23356	; 2 à l'arrière, de deux feux de gabarit rouges, placés à au plus 150 mm des extrémités supérieures droite 23356 et gauche du véhicule;	Bleu	Décision heureuse	Mauvaise catégorie
24673	. Lorsqu' une motocyclette est équipée d'une caisse adjacente, cette dernière doit être munie d'un feu rouge à l'arrière, placé le plus près possible de l'extrémité droite 24673 de la caisse.	Bleu	Décision heureuse	Mauvaise catégorie
26685	3: . Lorsque le rétroviseur intérieur est inutilisable, un rétroviseur doit être fixé à l'extérieur droit 26685 du véhicule. . Remorque ou semi-remorque.	Bleu	Décision heureuse	Mauvaise catégorie
26748	; 2 l'autre à l'extérieur droit 26748 du véhicule automobile (...)	Bleu	Décision heureuse	Mauvaise catégorie
26764	(...) si celui qui est fixé à l'intérieur ou à l'extérieur droit 26764 du véhicule est inutilisable.	Bleu	Décision heureuse	Mauvaise catégorie
				75

Tableau 36

b) Condition

Condition (nom)	
Signification 1	Signification 2
État dont l'existence indispensable pour qu'un autre état existe. Exemples : remplir les conditions exigées C'est une condition nécessaire. Clause, stipulation.	Ensemble de faits dont dépend qqch.

Collocations emmagasinées

Signification 1	Signification 2
G: assortir de	D: atmosphérique
G: satisfaire à	
G: satisfaire aux	
G: assorti de	
D: établir	
D: et formalités	
G: et modalités	
D: attacher	
D: de obtention	
D: conformer aux	
D: au permis	
G: supprimer	
D: de délivrance	
G: respecter	
D: prévu par règlement	
D: attaché	
D: établi	
G: selon	
G: à la	
G: fixer	
D: fixé	
D: fixer	
G: remplir	
G: conformément aux	

Occur.	Contexte	JLclx	Expert humain	Explication
141	2. Les dispositions du présent code qui s'appliquent au propriétaire d'un véhicule routier sont également applicables à l'égard de toute personne qui acquiert ou possède un véhicule routier en vertu d'un titre assorti d'une condition 141 ou d'un terme qui lui donne le droit d'en devenir propriétaire, ou en vertu d'un titre qui lui donne le droit d'en jouir comme propriétaire à charge de rendre.	Rouge	Décision heureuse	Adéquation discontinue
1105	9. La Société peut, avec l'approbation du ministre des Transports, nommer, aux conditions 1105 qu'elle établit, des personnes qu'elle autorise à effectuer pour son compte la perception des sommes prévues à l'un des articles 21 ou 31.1 ainsi que toute opération qu'elle indique	Rouge	Décision heureuse	Adéquation discontinue
1251	. La Société délivre, dans les cas et aux conditions 1251 prévus par règlement, l'une ou plusieurs des pièces suivantes: un certificat d'immatriculation, une plaque d'immatriculation,	Rouge	Décision heureuse	Adéquation continue
1385	10.1, l'immatriculation de certaines catégories de véhicules routiers prévues par règlement peut s'effectuer par l'inscription, dans le registre de la Société, des renseignements prévus par règlement relativement à ces catégories de véhicules et aux personnes qui satisfont aux conditions 1385 prévues par règlement pour obtenir cette immatriculation.	Rouge	Décision heureuse	Adéquation continue 2 ème colloc.
	11. Sur paiement des frais exigibles établis par règlement, la Société délivre une vignette d'identification à toute personne handicapée, propriétaire d'un véhicule routier et titulaire d'un permis de conduire, qui en fait la demande et satisfait aux	Rouge	Décision heureuse	Adéquation discontinue

Occur.	Contexte	Leix	Expert humain	Explication
	normes, conditions 1550 et formalités établies par règlement. . Vignette d'identification.			
1570	. La Société peut également, dans les cas et aux conditions 1570 <u>prévus par règlement</u> , délivrer une vignette d'identification:	Rouge	Décision heureuse	Adéquation continue
2029	18. Est exemptée de l'immatriculation, la remorque ou la semi-remorque louée pour une période n'excédant pas 12 mois, lorsque sont <u>réunies</u> les conditions 2029 suivantes: 1 le contrat de location a été conclu à l'	Rouge	Décision heureuse	Adéquation discontinue
2126	1986, c. 91, a. 18. . Non-résident. . 19. Sont exemptés de l'immatriculation, pour une période de six mois consécutifs depuis leur arrivée au Québec, le véhicule de promenade, la remorque ou la semi-remorque d'un non-résident, lorsque sont <u>réunies</u> les conditions 2126 suivantes: 1 le véhicule est immatriculé	Rouge	Décision heureuse	Adéquation discontinue
2238	20. Est exempté de l'immatriculation, un véhicule routier acquis en dehors du Québec par un étudiant, coopérant ou stagiaire étranger qui séjourne au Québec, pendant la durée de ses études ou de son stage, lorsque sont <u>réunies</u> les conditions 2238 suivantes: 1 le véhicule	Rouge	Décision heureuse	
2374	: 1 <u>satisfaire aux conditions</u> 2374 et aux formalités établies par règlement;	Rouge	Décision heureuse	Adéquation continue
2861	IV: CONDITIONS 2861 <u>ATTACHÉES</u> À L'IMMATRICULATION.	Rouge	Décision heureuse	Adéquation continue
3560	circulation. Il doit alors acquitter les droits et les frais, la contribution d'assurance et les frais supplémentaires prévus par règlement, conformément aux	Rouge	Décision heureuse	Adéquation continue

Occur.	Contexte	Lcbx	Expert humain	Explication
	conditions 3560 et modalités prévues par règlement.			
3785	34. Aucune plaque ou vignette qui peut être confondue avec une plaque d'immatriculation ou une vignette de contrôle délivrée par la Société ou par une autre autorité administrative compétente ne peut être fixée sur un véhicule routier ou apposée sur une plaque, sauf s'il s'agit d'une plaque ou d'une vignette requise en vertu d'une autre loi en vigueur au Québec ou d'une plaque fixée conformément aux normes et aux conditions 3785 établies en vertu du paragraphe 19 de l'article 618.	Rouge	Décision heureuse	Adéquation discontinue 2 ème colloc
5749	. Toutefois, la Société peut délivrer un permis sans la photographie ou la signature du titulaire ou sur support papier <u>selon</u> la catégorie et la classe du permis ainsi que les conditions 5749 et circonstances déterminées par règlement.	Rouge	Décision heureuse	Adéquation discontinue
5761	1995, c. 6, a. 2. . conditions. 5761. 64.	Noir	Indécidable	Contexte nul
5782	La Société peut, dans les cas prévus et selon les critères établis par règlement, <u>assortir</u> un permis <u>de conditions. 5782</u>	Rouge	Décision heureuse	Adéquation discontinue
6131	. conditions 6131 d'obtention.	Rouge	Décision heureuse	Adéquation continue
6148	. 69. Pour obtenir ou pour renouveler un permis, une personne doit <u>satisfaire aux conditions 6148</u> et aux formalités établies par règlement et, sauf dans les cas prévus par règlement, payer à la Société les frais fixés par règlement.	Rouge	Décision heureuse	Adéquation continue
6327	Perception. . 69.1. La Société peut, avec l'approbation du ministre des Transports, nommer, aux conditions 6327 qu'elle établit, des	Rouge	Décision heureuse	Adéquation discontinue
6408	70. Pour faire changer la classe de son permis ou pour lui ajouter une	Rouge	Décision heureuse	Adéquation continue

Occur.	Contexte	Lcbx	Expert humain	Explication
	autre classe, une personne doit <u>satisfaire aux conditions</u> 6408 d'obtention d'un permis de la classe qu'elle demande. 1986, c. 91, a. 70. .			2 ème colloc
6626	24 73. La Société peut exiger d'une personne qui demande l'obtention ou le renouvellement d'un permis, d'en faire changer la classe ou de lui en ajouter une autre ou de faire <u>supprimer</u> une <u>condition</u> 6626 y apparaissant, qu'elle se soumette à un examen médical ou optométrique et qu'elle lui remette un rapport de cet examen dans un délai qui ne peut excéder 90 jours. . Examen. . La Société peut requérir que l'examen soit	Rouge	Décision heureuse	Adéquation discontinue
6708	. La Société peut également exiger que la personne qui demande de faire <u>supprimer</u> une <u>condition</u> 6708 apparaissant sur son permis se soumette à un examen de compétence. 1986, c. 91, a. 73; 1987, c. 94, a. 9; 1990,	Rouge	Décision heureuse	Adéquation discontinue
6851	76. Une personne dont le permis a été révoqué ou dont le droit d'en obtenir un a été suspendu par suite d'une déclaration de culpabilité pour une infraction à l'une des dispositions du Code criminel (Lois révisées du Canada (1985), chapitre C-46) visées à l'article 180 doit, pour en obtenir un, se conformer aux <u>conditions</u> 6851 et <u>formalités</u> établies par règlement. . Restriction. . Aucun permis ne peut lui être délivré à	Rouge	Décision heureuse	Adéquation continue 2 ème colloc
7205	79. Une personne dont le permis a été révoqué ou dont le droit d'en obtenir un a été suspendu par suite de l'accumulation de points d'inaptitude doit, pour en obtenir un, se conformer aux <u>conditions</u> 7205 et <u>formalités</u> établies par règlement.	Rouge	Décision heureuse	Adéquation continue 2 ème colloc
7431	187.1 doit, pour en obtenir un, se	Rouge	Décision	Adéquation

Occur.	Contexte	Lclx	Expert humain	Explication
	conformer aux conditions 7431 et formalités établies par règlement. . Période d'attente. . Aucun permis ne		heureuse	continue 2 ème colloc
7666	187.2, doit pour obtenir cette classe se conformer aux conditions 7666 et formalités établies par règlement.	Rouge	Décision heureuse	Adéquation continue 2 ème colloc
8135	2 . condition 8135 au permis.	Rouge	Décision heureuse	Adéquation continue
8147	. 82. La Société peut refuser de <u>supprimer</u> une condition 8147 apparaissant sur un permis si son titulaire: 1 refuse de se soumettre à un examen médical ou optométrique	Rouge	Décision heureuse	Adéquation discontinue
82113	; 2 selon un rapport médical ou optométrique, est toujours atteint de la maladie, de la déficience ou se trouve encore dans la situation qui a amené la Société à <u>assortir</u> son permis de cette condition ; 82113	Rouge	Décision heureuse	Adéquation discontinue
8281	. La Société doit refuser de délivrer un permis, d'en changer la classe ou de lui en ajouter une autre, si la personne qui en fait la demande: 1 ne <u>satisfait pas aux conditions</u> 8281 de délivrance du permis de la classe demandée;	Rouge	Décision heureuse	Adéquation discontinue
8421	États-Unis; 4 ne <u>satisfait pas aux conditions</u> 8421 et aux formalités visées aux articles 76, 79, 80.1 et 80.3; 5 ne satisfait pas	Rouge	Décision heureuse	Adéquation discontinue
8599	; 2 qu'elle peut conduire un véhicule routier correspondant au permis de la classe demandée, en <u>respectant</u> des conditions 8599 reliées à son état fonctionnel, lesquelles la rendent apte à conduire ce véhicule sans constituer un danger pour la sécurité du public.	Rouge	Décision heureuse	Adéquation discontinue
8804	; 3 il <u>respecte</u> les conditions 8804 dont son permis de conduire est assorti;	Rouge	Décision heureuse	Adéquation discontinue

Occur.	Contexte	Lcix	Expert humain	Explication
9497	91. Le titulaire d'un permis de conduire valide délivré à l'extérieur du Canada peut, s'il s'établit au Québec, échanger ce permis contre un permis de conduire délivré par la Société à la condition 9497 de réussir l'examen de compétence et d'acquitter les droits et les frais fixés par règlement et le montant fixé en vertu de l'article 151 de la Loi sur l'assurance automobile (chapitre A-25).	Rouge	Décision heureuse	Adéquation continue
9968	III: CONDITIONS 9968 ATTACHÉES AUX PERMIS Signature.	Rouge	Décision heureuse	Adéquation discontinue 2 ème colloc
10326	(...), doit alors acquitter ces sommes ainsi que les frais supplémentaires prévus par règlement, conformément aux conditions 10326 et aux modalités prévues par règlement	Rouge	Décision heureuse	Adéquation discontinue
10487	. Sauf dans les cas et aux conditions 10487 prévues par règlement:	Rouge	Décision heureuse	Adéquation continue
10843	2 39 conditions. 10843.	Noir	Indécidable	
10854	98. Le titulaire d'un permis doit respecter les conditions 10854 dont son permis est assorti.	Rouge	Décision heureuse	Adéquation discontinue
11619	108. La Société doit aviser le titulaire d'un permis sur support plastique comportant sa photographie dont les classes ou les conditions 11619 ne correspondent pas à celles qui lui sont fixées de le remplacer dans le délai qu'elle indique.	Rouge	Décision heureuse	Adéquation discontinue
13562	société. Le requérant ainsi que la corporation ou la société pour laquelle il agit doivent également satisfaire aux conditions 13562 et aux formalités établies par règlement.	Rouge	Décision heureuse	Adéquation continue 2 ème colloc
13825	. Cette personne doit également satisfaire aux conditions 13825 et aux formalités établies par règlement.	Rouge	Décision heureuse	Adéquation continue 2 ème colloc

Occur.	Contexte	Léx.	Expert humain	Explication
13900	. 132. Le titulaire d'un permis d'école de conduite et le titulaire d'un permis d'enseignement doivent <u>respecter</u> les conditions 13900 établies par règlement se rattachant à leur permis. 1986, c. 91, a. 132. . <u>Registre</u>	Rouge	Décision heureuse	Adéquation discontinue
15137	151. Pour faire le commerce de véhicules routiers, une personne doit être titulaire d'une licence de commerçant délivrée par la Société, sur paiement des frais fixés et aux conditions 15137 <u>et formalités</u> établies par règlement.	Rouge	Décision heureuse	Adéquation continue
15331	153. Pour faire le commerce de véhicules routiers mis au rancart, de pièces provenant de ces véhicules ou de carcasses de véhicules, une personne doit être titulaire d'une licence de recycleur délivrée par la Société, sur paiement des frais fixés et aux conditions 15331 <u>et formalités</u> établies par règlement.	Rouge	Décision heureuse	Adéquation continue
15976	volée; 2 obtenir de la Société un permis pour vendre publiquement ce véhicule, délivré sur paiement des frais fixés et aux conditions 15976 <u>et formalités</u> établies par règlement.	Rouge	Décision heureuse	Adéquation continue
16091	. Les cautionnements prévus au présent titre doivent être fournis pour le montant fixé par règlement et selon la forme, les modalités et aux conditions 16091 <u>établies</u> par règlement. . <u>Règlement.</u>	Rouge	Décision heureuse	Adéquation continue
16106	. Il est mis fin à ces cautionnements aux conditions 16106 <u>établies</u> par règlement. 1986, c. 91, a. 160. . <u>Cautionnement.</u>	Rouge	Décision heureuse	Adéquation continue
16131	. 160.1. Une association de commerçants ou de recycleurs peut, selon la forme, les conditions 16131 <u>et les modalités</u> établies par règlement du gouvernement, se porter caution pour ses membres. Elle doit alors déposer	Rouge	Décision heureuse	Adéquation discontinue

Occur.	Contexte	Lcix	Expert humain	Explication
16175	3 58 conditions. 16175.	Noir	Indécidable	
16196	Le titulaire d'une licence ou d'un permis délivrés en vertu du présent titre doit <u>respecter</u> les conditions 16196 établies par règlement se rattachant à sa licence ou son permis.	Rouge	Décision heureuse	Adéquation discontinue
16239	. 161.1. Le titulaire d'une licence de commerçant qui est autorisé à effectuer l'immatriculation des véhicules routiers doit <u>respecter</u> les conditions 16239 établies par la Société en matière de transaction d'immatriculation et d'utilisation de l'immatriculation temporaire. 1987,	Rouge	Décision heureuse	Adéquation discontinue
16302	. La Société doit refuser de délivrer une licence ou un permis si le commerçant, le recycleur ou la personne visée à l'article 158 ne <u>satisfait pas aux conditions</u> 16302 de délivrance de la licence ou du permis.	Rouge	Décision heureuse	Adéquation discontinue
19220	public; 2 qu'il peut conduire un véhicule routier correspondant au permis ou à la classe en cause, en <u>respectant</u> des conditions 19220 reliées à son état fonctionnel, lesquelles le rendent apte à conduire ce véhicule sans constituer un danger pour la sécurité du public. 1990, c.	Rouge	Décision heureuse	Adéquation discontinue
20348	197. La Société doit lever la suspension imposée à la personne visée à l'article 196 et l'interdiction de remettre en circulation tout véhicule routier immatriculé à son nom, si celle-ci <u>remplit</u> l'une des conditions 20348 suivantes: 1 elle fournit à la Société une garantie	Rouge	Décision heureuse	Adéquation discontinue
21251	2. conditions 21251 non <u>respectées</u> .	Rouge	Décision heureuse	Adéquation discontinue
21278	. La Société peut suspendre un permis d'école de conduite ou un permis d'enseignement si son titulaire ne <u>remplit</u> plus les conditions 21278 se rattachant à ce	Rouge	Décision heureuse	Adéquation discontinue

Occur.	Contexte	Leix	Expert humain	Explication
	permis, s'il est déclaré coupable d'une infraction visée à l'article 146.2 à moins qu'un pardon n'ait été obtenu, ou s'il néglige ou refuse de suspendre			
21756	: 1 une licence de commerçant ou de recycleur si son titulaire ne <u>respecte</u> plus les conditions 21756 se rattachant à cette licence;	Rouge	Décision heureuse	Adéquation discontinue
22374	. La Société peut également apposer un numéro d'identification sur un véhicule routier sur paiement des frais fixés et aux conditions 22374 <u>établies</u> par règlement.	Rouge	Décision heureuse	Adéquation continue
23815	. Les feux et les réflecteurs prescrits au présent chapitre peuvent être combinés à <u>la condition</u> 23815 de satisfaire aux exigences du présent chapitre.	Rouge	Décision heureuse	Adéquation continue
24043	. Le ministre des Transports peut autoriser, aux conditions <u>établies par règlement</u> , l'installation et l'utilisation de phares blancs à l'arrière de certaines catégories ou sous-catégories de véhicules routiers.	Rouge	Décision heureuse	Adéquation continue
24150	munis de feux rouges clignotants ou pivotants. En outre, ils peuvent être munis de phares blancs clignotants alternatifs, dans les cas et aux conditions <u>prévus par règlement</u> . . Véhicules de police. . Les véhicules de police peuvent être munis de feux bleus clignotants ou pivotants.	Rouge	Décision heureuse	Adéquation continue
24216	. Les véhicules de service, les véhicules d'équipement, les véhicules utilisés pour le déneigement ou pour l'entretien des chemins, les véhicules pour lesquels les conditions 24216 <u>de délivrance</u> d'un permis spécial de circulation l'	Rouge	Décision heureuse	Adéquation continue
24321	228. Lorsque les conditions 24321 <u>de délivrance</u> d'un permis spécial de circulation exigent qu'un véhicule routier soit muni d'un feu jaune	Rouge	Décision heureuse	Adéquation continue

Occur.	Contexte	Lcbx	Expert human	Explication
	clignotant ou pivotant, ce feu peut y être fixé en permanence.			
24378	Toutefois, ce feu ne peut être utilisé que lors du transport d'un bien qui requiert la délivrance d'un permis spécial de circulation <u>conformément aux conditions</u> 24378 qui apparaissent à ce permis.	Rouge	Décision heureuse	Adéquation continue
24841	Les phares blancs prescrits au présent chapitre doivent être solidement fixés au véhicule et ajustés de façon à donner, dans des conditions 24841 atmosphériques normales et sur une route horizontale, un éclairage permettant au conducteur du véhicule de distinguer une personne ou un objet à une distance de 150 mètres. .	Rouge	Décision malheureuse	Conflit avec l'item "fixé"
27346	9 274.1. Le propriétaire d'un véhicule routier affecté au transport de personnes handicapées doit se conformer aux normes, conditions 27346 et <u>modalités</u> d'utilisation, de garde, d'entretien, de salubrité et de sécurité relatives à son véhicule. . Normes d'utilisation. . Il doit	Rouge	Décision heureuse	Adéquation continue
				68

Tableau 37

c) Examen

Examen (nom)	
Signification 1	Signification 2
Action de considérer, d'observer avec attention. Considération, étude, inspection, investigation, observation, recherche, analyse, etc... <i>Spécialement: examen médical</i>	Epreuve destinée à déterminer l'aptitude d'un candidat... Examen et concours. Examen écrit, oral. Examen partiel. Composition, interrogation.

Collocations enmagasinées

Signification 1	Signification 2
G:remettre pour	G:réussir
D: médical	G:requérir
D:optométrique	
D:par optométriste	
G:rapport de	
D:de compétence	
D:du local	
D:des registres	
D:de société	
D:par société	

Occur.	Contexte	Lcix	Expert humain	Explication
3918	Examen. 3918. 36.	Noir	Indécidable	Contexte nul
3941	. La personne qui conduit un véhicule routier doit, à la demande d'un agent de la paix, lui remettre pour examen 3941 les pièces visées à l'article 35.	Rouge	décision heureuse	Adéquation continue
5958	Examens. 5958.	Noir	Indécidable	Contexte nul
5978	67. Pour obtenir un permis, une personne doit, à moins d'en être exemptée par règlement, réussir les examens 5978 de compétence de la Société, qui en établit les formalités, les modalités et le contenu pour chacune des classes de permis	Rouge	décision heureuse	Adéquation continue Pb de réussir les examens
6265	. Toutefois, lorsqu' une personne ne demande son permis d' apprenti-conducteur que pour la durée de la séance pratique de son examen 6265 de compétence , elle n'est pas tenue de verser les sommes prévues au premier alinéa.	Rouge	décision heureuse	Adéquation continue
6635	24 73. La Société peut exiger d'une personne qui demande l'obtention ou le renouvellement d'un permis, d'en faire changer la classe ou de lui en ajouter une autre ou de faire supprimer une condition y apparaissant, qu' elle se soumette à un examen 6635 médical ou optométrique (...)	Rouge	décision heureuse	Adéquation continue
6648	(...) et qu'elle lui remette un <u>rapport de cet examen 6648</u> dans un délai qui ne peut excéder 90 jours.	Rouge	décision heureuse	Adéquation discontinue
6659	. Examen. 6659.	Noir	Indécidable	Contexte nul
6667	La Société peut requérir que l' examen 6667 soit fait <u>par un optométriste</u> ou un médecin qu'elle désigne nommément ou par un médecin dont elle détermine la spécialité.	Rouge	décision heureuse	Adéquation discontinue
6690	. Examen 6690 de compétence.	Rouge	décision heureuse	Adéquation continue
6717	. La Société peut également exiger que la personne qui demande de faire	Rouge	décision heureuse	Adéquation continue

Occur.	Contexte	Lclx	Expert humain	Explication
	supprimer une condition apparaissant sur son permis se soumettre à un examen 6717 de compétence .			
6757	75. Pour obtenir un permis de conduire, une personne doit réussir les examens 6757 de compétence visés à l'article 67 si elle n'a plus l'autorisation de conduire depuis trois ans ou plus.	Rouge	décision heureuse	Adéquation continue
7929	Demande: 1 refuse de se soumettre à un examen médical ou optométrique (...)	Rouge	décision heureuse	Adéquation continue
7943	(...) ou omet de lui remettre le rapport d'un tel examen ; 7943	Rouge	décision heureuse	Adéquation discontinue
8062	Demandée; 4 refuse de se soumettre à un examen 8062 de compétence ou y subit un échec;	Rouge	décision heureuse	Adéquation continue
8162	Titulaire: 1 refuse de se soumettre à un examen 8162 médical ou optométrique (...)	Rouge	décision heureuse	Adéquation continue
8176	(...) ou omet de lui remettre le rapport d'un tel examen ; 8176	Rouge	décision heureuse	Adéquation discontinue
8219	; 3 refuse de se soumettre à un examen 8219 de compétence ou y subit un échec.	Rouge	décision heureuse	Adéquation continue
9140	90. Le titulaire d'un permis de conduire valide délivré au Canada, par une autre autorité administrative peut, s'il s'établit au Québec, échanger sans examen 9140 ce permis contre un permis de conduire délivré par la Société , sur paiement des droits et des frais fixés par règlement et du montant fixé en vertu de l'article 151 de la Loi sur l'assurance automobile (chapitre A-25).	Rouge	Indécidable	Adéquation disloquée Contexte pauvre
9182	. Examen 9182 requis.	Noir	Indécidable	Contexte pauvre
9197	. Toutefois, la Société peut exiger que ce titulaire se soumette à un examen 9197 lorsque le permis échangé est un permis autorisant la conduite d'un véhicule de commerce, d'un véhicule d'urgence, d'un taxi, d'	Noir	Indécidable	Contexte pauvre

Occur.	Contexte	Lcix	Expert humain	Explication
	un autobus ou d'un minibus.			
9407	34 . Examen 9407 requis.	Noir	Indécidable	Contexte pauvre
9422	. Toutefois, la Société peut exiger que ce titulaire se soumette à un examen 9422 lorsque le permis demandé est un permis autorisant la conduite d'un véhicule de commerce, d'un véhicule d'urgence, d'un taxi, d'un autobus ou d'un minibus.	Noir	Indécidable	Contexte pauvre
9459	. Examen. 9459. 91.	Noir	Indécidable	Contexte nul
9501	. Le titulaire d'un permis de conduire valide délivré à l'extérieur du Canada peut, s'il s'établit au Québec, échanger ce permis contre un permis de conduire délivré par la Société à la condition de réussir l' examen 9501 de compétence et d'acquitter les droits et les frais fixés par règlement et le montant fixé en vertu de l'article 151 de la Loi sur l'assurance automobile (chapitre A-25).	Rouge	Décision heureuse	Adéquation continue
9677	(...) les personnes suivantes, si elles n'ont pas la citoyenneté canadienne et qu'elles n'exercent aucune entreprise, charge ou emploi au Québec autre que leur fonction auprès du gouvernement ou de l'organisme qu'elles représentent, peuvent obtenir de la Société, sans examen , 9677 un permis de conduire correspondant à celui dont elles sont titulaires, (...)	Noir	Indécidable	Contexte pauvre
11016	101. Le permis d'apprenti-conducteur autorisant la conduite d'une motocyclette ne permet à son titulaire de circuler sur les chemins publics que lors d'un cours pour la conduite d'une motocyclette approuvé par la Société et en présence du titulaire du permis d'enseignement ainsi que lors d'un examen 11016 de compétence de la Société.	Rouge	Décision heureuse	Adéquation continue
11061	102. Les personnes visées dans les	Rouge	Décision	Adéquation

Occur	Contexte	Lclx	Expert humain	Explication
	articles 97 et 100 doivent, à la demande d'un agent de la paix, <u>remettre</u> leur permis <u>pour</u> examen. 11061.	-	heureuse	discontinue
11691	<u>Examen</u> . 11674. 109.	Noir	Indécidable	
	La Société peut exiger que le titulaire d'un permis se soumette à un <u>examen</u> 11691 visé aux articles 67 ou 73 dans les cas suivants:	Noir	Indécidable	Contexte pauvre
11744	; 3 il n'a pas subi d' <u>examen</u> 11744 depuis dix ans; 4 elle a des motifs raisonnables de vérifier son état de	Noir	Indécidable	Contexte pauvre
13781	. <u>Examens</u> 13781 <u>de</u> <u>compétence</u> . .	Rouge	Décision heureuse	Adéquation continue
13800	. Pour obtenir ou renouveler un permis d'enseignement, une personne doit avoir réussi les <u>examens</u> 13800 <u>de</u> <u>compétence</u> de la Société qui en établit les formalités, les modalités et le contenu.	Rouge	Décision heureuse	Adéquation continue
13951	. 1986, c. 91, a. 133. . <u>Examen</u> 13951 <u>du</u> <u>local</u> et des registres.	Rouge	Décision heureuse	Adéquation continue
15020	. <u>Examen</u> 15020 <u>de</u> la <u>Société</u> .	Rouge	Décision heureuse	Adéquation discontinue
	Quiconque utilise sciemment, dans l'exploitation d'une école de conduite, tout ou partie de l' <u>examen</u> 15042 préparé par la <u>Société</u> pour évaluer la compétence de ceux qui désirent obtenir un permis de conduire, commet une infraction et est passible d'une amende de 600 à 2000 .	Rouge	Décision heureuse	Adéquation discontinue
18719	: I refuse de se soumettre à un <u>examen</u> 18719 <u>médical</u> ou optométrique (...)	Rouge	Décision heureuse	Adéquation continue
18733	(...) ou omet de lui remettre le rapport d'un tel <u>examen</u> ; 18733	Rouge	Décision heureuse	Adéquation discontinue
18872	; 4 refuse de se soumettre à un <u>examen</u> 18872 de compétence ou y subit un échec;	Rouge	Décision heureuse	Adéquation continue

Tableau 38 Le mot examen, Sens 1

d) Remettre

Remettre (verbe)	
Signification 1	Signification 2
Mettre de nouveau Faire passer dans un autre état ou un état antérieur. Exemple : remettre en circulation.	Mettre en la possession ou au pouvoir de qqn. Exemple : remettre un papier au destinataire.

Collocations enmagasinées

Signification 1	Signification 2
D: en circulation	D: pour examen
	D: pièces
	D: certificat
	D: rapport
	D: permis
	D: licence

Occur.	Contexte	Lcix	Expert humain	Explication
	. Les articles 188, 189, 196 à 202 s'appliquent compte tenu des adaptations nécessaires, sous réserve que la Société annule l'immatriculation de la personne qui l'a obtenue conformément au présent article au lieu d'interdire de remettre 1495 un véhicule routier en circulation.	Noir	décision heureuse	Inadéquation justifiée Le contexte ne contient, à juste titre, aucune collocation favorable à l'acception recherchée Signific. différente
	. Lorsque le propriétaire n'a pas payé les sommes prévues au premier alinéa à la date d'échéance ou lorsqu'il a avisé la Société qu'il renonce à circuler avec ce véhicule conformément au deuxième alinéa, nul ne peut, à compter de la date d'échéance ou de la date à laquelle la Société a reçu l'avis de renonciation, selon le cas, et sans autre avis, remettre 3499 le véhicule routier en circulation.	Noir	décision heureuse	Inadéquation justifiée Le contexte ne contient, à juste titre, aucune collocation favorable à l'acception recherchée Signific. différente
	. Remise en circulation. . Le propriétaire peut demander à la Société, pendant la durée correspondant au paiement des sommes visées au premier alinéa, l'autorisation de remettre 3532 ce véhicule routier en circulation. Il	Noir	décision heureuse	Inadéquation justifiée Le contexte ne contient, à juste titre, aucune collocation favorable à l'acception recherchée Signific. différente
	36. La personne qui conduit un véhicule routier doit, à la demande d'un agent de la paix, lui remettre 3939 pour examen les pièces visées à l'article 35.	Rouge	décision heureuse	Adéquation continue
	. Remise. . L'agent doit remettre 3955 ces pièces à leur détenteur dès qu'il les a examinées.	Rouge	décision heureuse	Adéquation discontinue
	immatriculation. Sur la demande motivée de l'agent de la paix, le	Rouge	décision heureuse	Adéquation discontinue

Occur.	Contexte	Lcx	Expert humain	Explication
	propriétaire doit lui remettre 4245 immédiatement son <u>certificat</u> et sa plaque d'immatriculation.			
	. Nul ne peut remettre 4276 en circulation un véhicule routier à l'égard duquel une décision de la Société rendue en vertu de l'un des articles 188, 189, 196 à 202 est en vigueur.	Noir	décision heureuse	Inadéquation justifiée Le contexte ne contient, à juste titre, aucune collocation favorable à l'acception recherchée Signific. différente
	. Lors de la cession du droit de propriété d'un véhicule routier entre des parties qui ne sont pas des commerçants, le cédant doit remettre 4348 à la Société le <u>certificat</u> et la plaque d'immatriculation délivrés pour ce véhicule après avoir endossé le certificat et le nouvel acquéreur doit demander une nouvelle immatriculation.	Rouge	décision heureuse	Adéquation discontinue
	. Lors de l'échange de véhicules routiers entre des parties qui ne sont pas des commerçants, chaque propriétaire doit remettre 4412 à la Société le <u>certificat</u> d'immatriculation endossé qui a été délivré pour son véhicule et demander	Rouge	décision heureuse	Adéquation discontinue
	42. Lorsque la cession du droit de propriété d'un véhicule routier est faite à un commerçant, le cédant qui n'acquiert pas un nouveau véhicule doit remettre 4476 au commerçant le <u>certificat</u> d'immatriculation du véhicule, après l'avoir endossé, et transmettre à la Société la plaque d'immatriculation de ce véhicule.	Rouge	décision heureuse	Adéquation discontinue
	. Si la cession du droit de propriété d'un véhicule routier est faite à un	Rouge	décision heureuse	Adéquation discontinue

Occur.	Contexte	Lcb	Expert humain	Explication
	commerçant, le cédant qui acquiert un nouveau véhicule doit conserver la plaque d'immatriculation, remettre 4545 le <u>certificat</u> d'immatriculation au commerçant, après l'avoir endossé, et demander à la Société la délivrance d'un certificat pour son nouveau véhicule.			
	. Le locataire d'un véhicule routier loué pour une période d'au moins un an doit remettre 4649 à la Société, à la fin de son contrat de location, le <u>certificat</u> d'immatriculation et la plaque d'immatriculation	Rouge	décision heureuse	Adéquation discontinue
	46. Sous réserve de l'article 17, la personne qui devient propriétaire d'un véhicule routier, par suite d'un décès, d'une donation, d'un partage, d'une liquidation, d'une faillite, de l'exercice d'un droit de reprise, d'une cession complète d'une entreprise ou d'une vente en justice, doit remettre 4742 à la Société le <u>certificat</u> d'immatriculation et lui demander une nouvelle immatriculation du véhicule.	Rouge	décision heureuse	Adéquation discontinue
	: 1 refuse de se soumettre à un examen médical ou optométrique ou omet de lui remettre 7937 le <u>rapport</u> d'un tel examen;	Rouge	décision heureuse	Adéquation discontinue
	: 1 refuse de se soumettre à un examen médical ou optométrique ou omet de lui remettre 8170 le <u>rapport</u> d'un tel examen; 2 selon un rapport médical ou optométrique, est toujours atteint de la maladie, de la déficience ou se trouve encore dans la situation qui a amené la Société à assortir son permis de cette condition;	Rouge	décision heureuse	Adéquation discontinue
	. Toutefois, le titulaire d'un permis de conduire valide délivré ailleurs qu'aux États-Unis n'est pas tenu de remettre 9558 le <u>permis</u> délivré dans	Rouge	décision heureuse	Adéquation discontinue

Occur.	Contexte	Lclx	Expert humain	Explication
	son pays d'origine.			
	. Les personnes visées dans les articles 97 et 100 doivent, à la demande d'un agent de la paix, remettre 11057 leur permis <u>pour examen</u> .	Rouge	décision heureuse	Adéquation discontinue
	. L'agent doit remettre 11070 ce <u>permis</u> à son titulaire dès qu'il l'a examiné.	Rouge	décision heureuse	Adéquation discontinue
	. Sur la demande motivée de l'agent de la paix, la personne doit lui remettre 11551 immédiatement son <u>permis</u> .	Rouge	décision heureuse	Adéquation discontinue
	. Un commerçant ou un recycleur qui vend un véhicule routier monté par un recycleur doit remettre 15805 à l'acheteur un <u>certificat</u> de vérification mécanique indiquant que le véhicule satisfait aux exigences du présent code.	Rouge	décision heureuse	Adéquation discontinue
	. Sur la demande motivée de l'agent de la paix, la personne doit lui remettre 16409 immédiatement sa <u>licence</u> .	Rouge	décision heureuse	Adéquation discontinue
18282	II: INTERDICTION DE REMETTRE 18282 UN VÉHICULE ROUTIER EN CIRCULATION, SUSPENSION DES <u>PERMIS</u> ET DES LICENCES SECTION I:	Rouge	décision malheureuse	pseudo
18298	INTERDICTION DE REMETTRE 18298 UN VÉHICULE ROUTIER EN CIRCULATION, SUSPENSION DES <u>PERMIS</u> D'APPRENTIS-CONDUCTEURS, DES PERMIS PROBATOIRES ET DES PERMIS DE CONDUIRE .	Rouge	décision malheureuse	pseudo
	Interdiction de remettre 18321 en circulation.	Noir	décision heureuse	Inadéquation justifiée Le contexte ne contient, à juste titre, aucune collocation favorable à

Occur	Contexte	Lclx	Expert humain	Explication
				l'acception recherchée Signific. différente
	. 188. La Société peut interdire de remettre 18331 un véhicule routier en circulation dans les cas suivants:	Noir	décision heureuse	Inadéquation justifiée Le contexte ne contient, à juste titre, aucune collocation favorable à l'acception recherchée Signific. différente
	. La Société doit interdire de remettre 18553 un véhicule routier en circulation dans les cas suivants: 1 lorsque la Commission des	Noir	décision heureuse	Inadéquation justifiée Le contexte ne contient, à juste titre, aucune collocation favorable à l'acception recherchée Signific. différente
	: 1 refuse de se soumettre à un examen médical ou optométrique ou omet de lui remettre 18727 le rapport d'un tel examen; 2 selon un rapport médical ou optométrique, est atteint d'une maladie, d'une déficience ou se trouve dans une situation qui, suivant les normes	Rouge	décision heureuse	Adéquation discontinue
	. Les suspensions prévues au premier alinéa ne doivent pas être imposées ou doivent être annulées ou l'interdiction de remettre 20118 un véhicule routier en circulation ne doit pas être prononcée ou doit être annulée lorsqu' il est démontré à la satisfaction de la Société qu'au moment de l'accident, le véhicule	Noir	décision heureuse	Inadéquation justifiée Le contexte ne contient, à juste titre, aucune collocation favorable à l'acception recherchée

Occur.	Contexte	Lcix	Expert humain	Explication
	non assuré était légalement stationné,			Signific. différente
	. À l'égard du conducteur et du propriétaire, les suspensions prévues au premier alinéa ne doivent pas être imposées ou doivent être annulées ou l'interdiction de remettre 20234 un véhicule routier en circulation ne doit pas être prononcée ou doit être annulée lorsqu' il est démontré, à la satisfaction de la Société, qu'au moment de l'accident, le conducteur était propriétaire d'un véhicule routier pour lequel il détenait le	Noir	décision heureuse	Inadéquation justifiée Le contexte ne contient, à juste titre, aucune collocation favorable à l'acception recherchée Signific. différente
	. La Société doit lever la suspension imposée à la personne visée à l' article 196 et l'interdiction de remettre 20332 en circulation tout véhicule routier immatriculé à son nom, si celle-ci remplit l'une des conditions suivantes:	Noir	décision heureuse	Inadéquation justifiée Le contexte ne contient, à juste titre, aucune collocation favorable à l'acception recherchée Signific. différente
	. Lorsque le créancier ayant conclu une entente visée au premier alinéa avise la Société qu'il y a eu interruption des versements, celle-ci doit remettre 20445 en vigueur la suspension et l'interdiction qu'elle avait levées à la suite de cette entente.	Noir	décision heureuse	Inadéquation justifiée Le contexte ne contient, à juste titre, aucune collocation favorable à l'acception recherchée Signific. différente
	. La Société doit lever la suspension imposée à la personne visée à l' article 200 et l'interdiction de remettre 20904 en circulation le véhicule routier immatriculé au nom de cette personne dans les cas suivants:	Noir	décision heureuse	Inadéquation justifiée Le contexte ne contient, à juste titre, aucune collocation favorable à l'acception

Occur.	Contexte	Lcix	Expert humain	Explication
				recherchée Signific. différente
	. Dans le cas visé au paragraphe 3, la Société doit, sur réception d'un avis du créancier indiquant l'interruption des versements, remettre 21071 en vigueur la suspension et l'interdiction qu'elle avait levées à la suite de l'entente.	Noir	décision heureuse	Inadéquation justifiée Le contexte ne contient, à juste titre, aucune collocation favorable à l'acception recherchée Signific. différente
	; 1990, 202. Lorsque la Société a effectué un paiement pour satisfaire à un jugement, elle doit lever la suspension imposée à la personne visée à l'article 200 et l'interdiction de remettre 21139 en circulation le véhicule routier immatriculé au nom de cette personne dans les cas suivants:	Noir	décision heureuse	Inadéquation justifiée Le contexte ne contient, à juste titre, aucune collocation favorable à l'acception recherchée Signific. Différente
				35

e) *Objet*

Objet (nom)	
Signification 1	Signification 2
<p>CONCRET: Chose solide ayant unité et indépendance...</p> <p>Chose; Familier : bidule, machin, truc.</p> <p>Instrument, outil.</p> <p>Affaire, article.</p> <p>Bagatelle, broutilie.</p> <p>Colifichet.</p>	<p>ABSTRAIT: Matière sur laquelle porte un droit, une obligation, un contrat.</p> <p>Exemple : l'objet d'un litige, d'un procès.</p>

Collocations enmagasinées

Signification 1	Signification 2

G:libre de tout	G:faire
D:inanimé	

Occur.	Contexte	Ech.	Expert humain	Explication
37	objet. 16. 1. Le présent code régit l'utilisation des véhicules et la circulation des piétons sur les chemins publics.	Noir	?	
37	. Le présent code régit l'utilisation des véhicules et la circulation des piétons sur les chemins publics. . objet. 37	Noir	?	
1960	propriété est transférée par suite d'un décès, si ce véhicule fait l'objet 1960 d'un autre transfert de propriété dans un délai de sept jours à compter du moment où l'héritier ou le légataire en a obtenu la délivrance ou l'a reçu en paiement de son legs.	Rouge	Décision heureuse	Adéquation discontinue
3614	. La plaque d'immatriculation doit être libre de tout objet 3614 ou de toute matière pouvant en empêcher la lecture. Elle doit, en outre,	Noir	Décision heureuse	Inadéquation justifiée Le contexte ne contient, à juste titre, aucune collocation favorable à l'acceptation recherchée Signific. différente
4607	propriété d'un véhicule routier qui fait l'objet 4607 d'un contrat de location d'une durée d'au moins un an. 1986, c. 91, a. 44. . Remise à	Rouge	Décision heureuse	Adéquation discontinue
8341	17 44. Les articles 40 à 43 ne s'appliquent pas à la cession du droit de 30 demandée; 3 fait l'objet 8341 d'une interdiction de conduire, d'une révocation ou d'une suspension de son permis ou de la classe de celui-ci ou d'une suspension de son droit d'obtenir un permis ou une classe de celui-ci en vigueur ou imposée mais non encore en vigueur, y compris en vertu d'une loi du Canada, d'une autre province ou	Rouge	Décision heureuse	Adéquation discontinue

Occur.	Contexte	Lclx	Expert humain	Explication
	d'un territoire du Canada ou d'une loi d'un État des États-Unis; 4 ne satisfait pas aux			
8818	; 4 il ne fait pas l'objet 8818 d'une suspension du droit d'obtenir un permis délivré par la Société ou une classe de celui-ci. 1986,	Rouge	Décision heureuse	Adéquation discontinue
9331	, (...) si ce titulaire s'établit au Québec et ne fait pas l'objet 9331 d'une interdiction de conduire, d'une révocation ou d'une suspension de son permis ou de la classe de celui-ci ou d'une suspension de son droit d'obtenir un permis ou une classe de celui-ci en vigueur ou imposée mais non encore en vigueur, (...)	Rouge	Décision heureuse	Adéquation discontinue
11201	40 105. Une personne ne peut conduire un véhicule routier lorsqu'elle fait l'objet 11201 d'une sanction même si elle est titulaire d'un permis de conduire international. . Permis restreint. . Toutefois, le titulaire	Rouge	Décision heureuse	Adéquation discontinue
11297	Permis de conduire international. . 106. Le propriétaire, le locataire ou la personne qui a le contrôle d'un véhicule routier ne peut laisser conduire ce véhicule par une autre personne faisant l'objet 11297 d'une sanction, même si cette dernière est titulaire d'un permis de conduire international. . Permis restreint. . Il peut toutefois laisser	Rouge	Décision heureuse	Adéquation discontinue
11465	107. Une personne dont le permis ou une classe de celui-ci fait l'objet 11465 d'une suspension ou d'une révocation doit, sur demande de la Société, lui retourner son permis à la date d'entrée en vigueur de la suspension ou de la révocation ou	Rouge	Décision heureuse	Adéquation discontinue

Occur.	Contexte	Lcix	Expert humain	Explication
	à toute autre date ultérieure fixée par la Société. . Confiscation. . La Société peut demander à un agent de la			
12855	obtenu; 2 au moment de la présentation de la requête, le permis du requérant ou son droit d'en obtenir un fait l'objet 12855 d'une suspension en vigueur ou imposée et non encore en vigueur; 3 la sanction	Rouge	Décision heureuse	Adéquation discontinue
13190	(...), si entre la date fixée pour la présentation de la requête et celle de la délivrance du permis restreint, le droit de cette personne d'obtenir un permis fait l'objet 13190 d'une suspension en vigueur ou imposée et non encore en vigueur, la Société doit en aviser le juge qui a rendu l'ordonnance.	Rouge	Décision heureuse	Adéquation discontinue
14510	52 143. Quiconque contrevient au premier alinéa de l'article 105 commet une infraction et est passible d'une amende de 300 à 600 , si son permis ou son droit d'en obtenir un fait l'objet 14510 d'une révocation ou d'une suspension pour un motif autre que ceux visés à l'article 180. 1986,	Rouge	Décision heureuse	Adéquation discontinue
14581	c. 91, a. 143; 1990, c. 4, a. 212. . Infraction et peine. . 144. Quiconque contrevient au premier alinéa de l'article 105 commet une infraction et est passible d'une amende de 600 à 2 000 , si son permis ou son droit d'en obtenir un fait l'objet 14581 d'une révocation ou d'une suspension conformément à l'article 180.	Rouge	Décision heureuse	Adéquation discontinue
16332	163. Une personne dont la licence fait l'objet 16332 d'une	Rouge	Décision heureuse	Adéquation discontinue

Occur.	Contexte	Lex	Expert humain	Explication
	suspension doit, sur demande de la Société, retourner sa licence à la date d'entrée en vigueur de la suspension ou à toute autre date ultérieure fixée par la Société. . Confiscation. . La Société peut			
16887	171. Le conducteur d'un véhicule routier qui est impliqué dans un accident avec un animal pesant plus de 25 kg, un véhicule routier inoccupé ou un autre objet 16887 inanimé doit, lorsque le propriétaire du bien endommagé ou une personne qui le représente ne peut être rejoint sur les lieux de l'accident ou à proximité, communiquer sans délai avec le poste de police le plus près afin de rapporter l'accident et de fournir les renseignements prévus à l'article 170.	Noir	Décision heureuse	Inadéquation justifiée Le contexte ne contient, à juste titre, aucune collocation favorable à l'acception recherchée Signific. différente
	71 195. Lorsque la période de validité d'un permis se termine avant la fin de la période de la suspension dont celui-ci faisait l'objet, 19824 le droit d'obtenir un permis est alors suspendu pour la durée de la période de suspension non expirée. 1986, c. 91, a. 195; 1990, c. 83, a.	Rouge	Décision heureuse	Adéquation discontinue
19889	85. . Suspension obligatoire. . 195.1. La Société doit suspendre le permis restreint d'une personne si, après la date où il a été délivré, le droit de cette personne d'obtenir un permis fait l'objet 19889 d'une suspension en vigueur ou imposée et non encore en vigueur. Cette	Rouge	Décision heureuse	Adéquation discontinue
21698	206. Lorsque la période de validité d'un permis d'école de conduite ou d'un permis d'enseignement se termine avant la fin de la période de la	Rouge	Décision heureuse	Adéquation discontinue

Occur.	Contexte	Lcix	Expert humain	Explication
	suspension dont ce permis faisait l'objet, 21698 le droit d'obtenir un tel permis est alors suspendu pour la durée de la période de suspension non expirée. 1986, c. 91, a. 206. SECTION III: LICENCES DE			
22275	80 209. Lorsque la période de validité d'une licence de commerçant ou de recycleur se termine avant la fin de la période de la suspension dont cette licence faisait l'objet, le droit d'obtenir une telle licence est alors suspendu pour la durée de la période de suspension non expirée. 1986, c. 91, a. 209. TITRE VI: RÈGLES CONCERNANT LES VÉHICULES	Rouge	Décision heureuse	Adéquation discontinuée
24862	89 235. Les phares blancs prescrits au présent chapitre doivent être solidement fixés au véhicule et ajustés de façon à donner, dans des conditions atmosphériques normales et sur une route horizontale, un éclairage permettant au conducteur du véhicule de distinguer une personne ou un objet 24862 à une distance de 150 mètres. Éclairage.	Noir	Décision heureuse	Inadéquation justifiée Le contexte ne contient, à juste titre, aucune collocation favorable à l'acception recherchée Signific. différente
24862	ou un objet 24862 à une distance de 150 mètres. Éclairage. Toutefois, l'éclairage doit permettre au conducteur d'un cyclomoteur de distinguer une personne ou un objet 24888 à une distance de 90 mètres et au conducteur d'une bicyclette à une distance de 10 mètres. 1986, c.	Noir	Décision heureuse	Inadéquation justifiée Le contexte ne contient, à juste titre, aucune collocation favorable à l'acception recherchée Signific. Différente
				24

Tableau 39 Objet

*f) Application***Signification retenue :**

Application (nom)	
Signification 1	Signification 2
Action de mettre une chose sur une autre de manière qu'elle la recouvre...	Action de faire porter sur quelque chose
Placage: l'application d'un enduit sur un mur. Pose: application de ventouses.	Mise en pratique : mettre en application une loi

Collocations emmagasinées :

Signification 1	Signification 2
D: de force	G: Champ de
	D: de loi
	D: du code
	D: des articles
	D: de article
	D: du chapitre
	D: du article
	D: du titre

Occur	Contexte	Lclx	Expert humain	Explication
12	CHAMP D'APPLICATION 12 ET DEFINITIONS	Rouge	Décision heureuse	Adéquation continue
691	un véhicule automobile exploité en vertu d'un permis délivré en application 691 de la Loi sur le transport par taxi (chapitre T-11.1); . véhicule automobile ; . véhicule automobile :	Rouge	Décision heureuse	Adéquation discontinue
953	Pour l' application 953 du présent code , le mot <u>personne</u> comprend également une société. 1986, c. 91, a. 5. TITRE I:	Rouge	Décision heureuse	Adéquation discontinue
1447	application 1447 de aa. 188 et 189, 196 à 202. . Les articles	Noir	Décision malheureuse	Inadéquation fautive: Carence du contexte : le contexte est pauvre.
1883	16. Pour l' application 1883 des articles 14 et 15, un agriculteur est une personne physique membre d'une association accréditée en vertu de la <u>Loi</u> sur les producteurs agricoles (chapitre P-28) ou une personne propriétaire ou locataire d'une ferme et dont l'agriculture est la principale activité.	Rouge	Décision heureuse	Adéquation disloquée
11374	106.1. Pour l' application 11374 des articles 105 et 106, sanction s'entend de la révocation ou de la suspension soit du permis autorisant la conduite d'un véhicule routier, soit d'une classe d'un tel permis, ainsi que de la suspension du droit d'obtenir un tel permis ou une telle classe. Sont aussi visées la révocation et la suspension prononcées par	Rouge	Décision heureuse	Adéquation continue
11803	application. 11803. 110.	Noir	Décision heureuse	Inadéquation fautive: b- Carence du contexte le contexte est pauvre.

Occur	Contexte	Lcbx	Expert humain	Explication
12703	120. Lorsqu' une requête lui est signifiée conformément à l'article 119, la Société doit transmettre au tribunal, avant la date fixée pour la présentation de la requête, tout renseignement qu'elle détient à l'égard du requérant relativement à l' application 12703 de l'article 121.	Rouge	Décision heureuse	Adéquation discontinue
13108	123. Une décision rendue en application 13108 des articles 118 à 121 est finale et sans appel. 1986, c. 91, a. 123. . Révision d'une	Rouge	Décision heureuse	Adéquation continue
13304	105. 1986, c. 91, a. 125; 1990, c. 83, a. 56. . application. 13304.	Noir	?	Inadéquation fautive: b- Carence du contexte le contexte est pauvre.
16639	167. Pour l' application 16639 du présent titre , un accident est un événement au cours duquel un dommage est causé par un véhicule routier en mouvement. 1986, c. 91, a. 167. . Conducteur impliqué. . 168. Le	Rouge	Décision heureuse	Adéquation discontinue
20591	. Pour l' application 20591 du présent article , la Société peut exiger un rapport de l'évaluation des dommages découlant de l'accident et déterminer sous quelle forme une garantie peut lui être fournie.	Rouge	Décision heureuse	Adéquation discontinue
21521	. Pour l' application 21521 du présent article , il ne doit pas être tenu compte d'une déclaration de culpabilité pour laquelle un pardon a été obtenu.	Rouge	Décision heureuse	Adéquation discontinue
22198	. Pour l' application 22198 du présent article , il ne doit pas être tenu compte d'une déclaration de culpabilité pour laquelle un pardon a été obtenu.	Rouge	Décision heureuse	Adéquation discontinue
24253	. Pour l' application 24253 du	Rouge	Décision	Adéquation

Occur	Contexte	Lcix	Expert humain	Explication
	présent <u>article</u> , un véhicule de service est un véhicule automobile agencé pour l'approvisionnement, la réparation ou le remorquage des véhicules routiers et un véhicule d'équipement est un véhicule automobile servant au transport de l'équipement qui y est fixé en permanence et comportant un espace pour le chargement.		heureuse	discontinue
25307	241. Pour l' <u>application</u> 25307 du présent <u>chapitre</u> , les mots véhicule automobile ne comprennent pas la motocyclette et le cyclomoteur. 1986, c.91, a. 241. . Véhicule routier. . 242. Sous réserve des articles 243 à	Rouge	Décision heureuse	Adéquation discontinue
25509	244. Les remorques et les semi-remorques qui font partie d'un ensemble de véhicules routiers et dont la masse, charge comprise, est de 1 300 kg ou plus ou dont la masse, charge comprise, excède la moitié de la masse nette du véhicule automobile qui les tire doivent être munies d'un système de freins indépendant permettant l' <u>application</u> 25509 d'une force de freinage sur chaque roue portante.	Noir	Décision heureuse	Inadéquation justifiée Le contexte ne contient, à juste titre, aucune collocation favorable à l'acception recherchée
26165	93 253. Pour l' <u>application</u> 26165 <u>des articles</u> 251 et 252, un détecteur de radar de vitesse est tout appareil ou ensemble d'appareils qui peut être utilisé pour aviser le conducteur d'un véhicule routier de la présence d'un radar de vitesse ou pour nuire au fonctionnement normal d' un tel radar.	Rouge	Décision heureuse	Adéquation continue
26330	véhicule routier. <u>application</u> . 26330. En outre des chemins	Noir	?	Inadéquation fautive: b- Carence du contexte le contexte est pauvre.

Tableau 40 Le mot **application**

g) Avis

Avis (nom)	
Signification 1	Signification 2
Ce que l'on pense, ce que l'on exprime sur un sujet. Jugement, estimation, opinion, pensée, point de vue, sentiment.	Ce que l'on porte à la connaissance de qqn. Annonce, communication, information, message, note, notification, nouvelle, renseignement. Préavis. Indication, ordre.

Collocations emmagasinées

Signification 1	Signification 2
	G:réception de
	D:total des points
	G:recevoir
	D:visé
	G:sans autre
	D:transmis
	D:à société
	G:faire parvenir

Occur	Contexte	Lcbx	Expert humain	Explication
2532	. Toutefois la Société peut interdire à ce propriétaire de mettre son véhicule en circulation sans autre avis 2532 s'il ne paie pas les sommes prévues au paragraphe 3 du premier alinéa ou s'il est débiteur de la Société à l'égard des sommes <u>visées</u> dans l'article 31.1	Rouge	Décision heureuse	Adéquation disloquée
3489	. Lorsque le propriétaire n'a pas payé les sommes prévues au premier alinéa à la date d'échéance ou lorsqu'il a avisé la Société qu'il renonce à circuler avec ce véhicule conformément au deuxième alinéa, nul ne peut, à compter de la date d'échéance ou de la date à laquelle la Société a <u>reçu</u> l'avis 3489 de renonciation, selon le cas, (...)	Rouge	Décision heureuse	Adéquation discontinue
3498	. Lorsque le propriétaire n'a pas payé les sommes prévues au premier alinéa à la date d'échéance ou lorsqu'il a avisé la Société qu'il renonce à circuler avec ce véhicule conformément au deuxième alinéa, nul ne peut, à compter de la date d'échéance ou de la date à laquelle la Société a <u>reçu</u> l'avis 3489 de renonciation, selon le cas, et sans autre avis, 3498 remettre le véhicule routier en circulation.	Rouge	Décision heureuse	Adéquation disloquée
8029	; 3 selon un rapport médical ou optométrique est atteinte d'une maladie, d'une déficience ou se trouve dans une situation non visées dans les normes médicales et optométriques établies par règlement mais qui, d'après l'avis 8029 d'un membre du Comité consultatif médical et optométrique, sont incompatibles avec la conduite d'un véhicule routier correspondant au permis de la classe demandée;	Noir	Décision heureuse	Inadéquation justifiée 2eme signif
8653	84. La Société doit refuser de délivrer un permis à une personne à	Rouge	Décision heureuse	Adéquation discontinue

Occur	Contexte	Lcix	Expert humain	Explication
	l'égard de laquelle elle a <u>reçu</u> l'avis 8653 prévu à l'article 364 du Code de procédure pénale (chapitre C-25.1). . Acquittement.			
8684	Code de procédure pénale (chapitre C-25.1). . Acquittement. . Ce refus de délivrer est maintenu tant que la Société n'a pas <u>reçu</u> l'avis 8684 prévu à l'article 365 de ce code.	Rouge	Décision heureuse	Adéquation discontinue
10143	. A défaut de paiement au cours de cette période, le titulaire ne peut, à compter du premier jour suivant la date d'expiration de cette période et sans autre avis 10143 de la Société, conduire un véhicule routier.	Rouge	Décision heureuse	Adéquation continue
12180	. Avis 12180 du <u>total des points</u> d'incapacité.	Rouge	Décision heureuse	Adéquation discontinue
12224	. Dès que le nombre total de points d'incapacité inscrits au dossier d'une personne a atteint celui prescrit, la Société doit lui <u>faire parvenir</u> , à la dernière adresse qui lui a été fournie, un avis 12224 l'informant du nombre de points inscrits à son dossier et lui rappelant ses pouvoirs de révocation et de suspension.	Rouge	Décision heureuse	Adéquation discontinue Exemple modèle
12267	115. Le défaut par la Société de donner l'avis 12267 <u>visé</u> à l'article 114 n'entraîne pas la nullité.	Rouge	Décision heureuse	Adéquation continue
12280	. n'entraîne pas la nullité d'un avis 12280 <u>transmis</u> ultérieurement et ne l'empêche pas d'exercer ultérieurement un pouvoir ou un devoir en vertu de la présente section, du chapitre III et du titre V.	Rouge	Décision heureuse	Adéquation continue
16947	. Avis 16947 à la <u>Société</u> . . 172.	Rouge	Décision heureuse	Adéquation discontinue
18829	; 3 selon un rapport médical ou optométrique, est atteint d'une maladie, d'une déficience ou se trouve dans une situation non visées dans les normes médicales et	Noir	Décision heureuse	Inadéquation justifiée 2eme signif

Occur	Contexte	Lcbx	Expert humain	Explication
	optométriques établies par règlement mais qui, d'après l'avis 18829 d'un membre du Comité consultatif médical et optométrique, sont incompatibles avec la conduite d'un véhicule routier correspondant à l'un des permis ou à l'une des classes de permis qu'il possède;			
19732	194. La Société doit suspendre le permis d'apprenti-conducteur et le permis probatoire ou le permis de conduire d'une personne ou suspendre, si elle n'est pas titulaire d'un permis d'apprenti-conducteur, d'un permis probatoire ou d'un permis de conduire, son droit de l'obtenir, lorsqu' elle reçoit l'avis 19732 prévu à l'article 364 du Code de procédure pénale (chapitre C-25.1).	Rouge	Décision heureuse	Adéquation discontinue
19762	Cette suspension demeure en vigueur tant que la Société n'a pas reçu l'avis 19762 prévu à l'article 365 de ce code.	Rouge	Décision heureuse	Adéquation discontinue
20706	. Avis 20706 du créancier.	Noir	Indécidable	Ambigu
20715	. Sur réception d'un avis 20715 à cet effet du créancier concerné, la Société suspend le permis d'apprenti-conducteur et le permis probatoire ou le permis de conduire ou le droit de les obtenir et interdit la remise en circulation du véhicule routier immatriculé au nom du débiteur qui n'a pas satisfait, dans le délai d'exécution, à un jugement qui prononce au Canada une condamnation définitive.	Rouge	Décision heureuse	Adéquation discontinue
21063	. Dans le cas visé au paragraphe 3, la Société doit, sur réception d'un avis 21063 du créancier indiquant l'interruption des versements, remettre en vigueur la suspension et l'interdiction qu'elle avait levées à la suite de l'entente.	Rouge	Décision heureuse	Adéquation discontinue

Occur	Contexte	Lcbx	Expert human	Explication
				19

Tableau 41 le mot Avis

h) Remise

Remise (nom)	
Signification 1	Signification 2
Mettre de nouveau Faire passer dans un autre état ou un état antérieur. Exemple : remettre en circulation.	Mettre en la possession ou au pouvoir de qqn Exemple: remettre un papier au destinataire.

Collocations emmagasinées

Signification 1	Signification 2
D: en circulation	D: à société
D: en vigueur	D: du permis
	D: du certificat
	D: de permis
	D: de appareil

Occur.	Contexte	Lclx	Expert humain	Explication
3506	. Remise 3506 en circulation.	Noir	Décision heureuse	Inadéquation justifiée Le contexte ne contient, à juste titre, aucune collocation favorable à l'acception recherchée Signific. différente
3950	Remise. 3950. L'agent doit remettre ces pièces à leur détenteur	Noir	?	
4192	. Remise 4192 du certificat et de la plaque.	Rouge	Décision heureuse	Adéquation continue
4627	. Remise 4627 à la Société.	Rouge	Décision heureuse	Adéquation discontinue
9535	. Remise 9535 du permis	Rouge	Décision heureuse	Adéquation continue
11033	. Remise 11033 du permis.	Rouge	Décision heureuse	Adéquation continue
11063	. Remise 11063 du permis.	Rouge	Décision heureuse	Adéquation continue
11447	. Remise 11447 à la Société.	Rouge	Décision heureuse	Adéquation discontinue
11556	Remise 11556 à la Société. .	Rouge	Décision heureuse	Adéquation discontinue
11567	. La Société peut également exiger la remise 11567 de tout autre permis délivré par une autre autorité administrative.	Rouge	Décision heureuse	Adéquation discontinue
20076	. En outre, elle doit interdire la remise 20076 en circulation de tout véhicule routier immatriculé au nom de l'une ou l'autre de ces personnes.	Noir	Décision heureuse	Inadéquation justifiée Le contexte ne contient, à juste titre, aucune collocation favorable à l'acception recherchée Signific. différente
20747	. Sur réception d'un avis à cet effet du créancier concerné, la Société suspend le permis d'apprenti-conducteur et le permis probatoire ou	Noir	Décision heureuse	Inadéquation justifiée Le contexte ne contient, à juste

Occur.	Contexte	Lcbr.	Expert humain	Explication
	le permis de conduire ou le droit de les obtenir et interdit la remise 20747 en circulation du véhicule routier immatriculé au nom du débiteur qui n'a pas satisfait, dans le délai d'exécution, à un jugement qui prononce au Canada une condamnation définitive.			titre, aucune collocation favorable à l'acceptation recherchée Signific. différente
21045	. Remise 21045 en vigueur.	Noir	Décision heureuse	Inadéquation justifiée Le contexte ne contient, à juste titre, aucune collocation favorable à l'acceptation recherchée Signific. différente
25209	. Remise 25209 à la Société.	Rouge	Décision heureuse	Adéquation discontinue
26383	. Remise 26383 de l'appareil à la Société.	Rouge	Décision heureuse	Adéquation discontinue
				16

i) Pièce

Pièce (nom)	
Signification 1	Signification 2
Tout écrit servant à établir un droit, à faire la preuve d'un fait. Exemples : pièces justificatives. Pièces d'identité.	Chacun des éléments dont l'assemblage forme un tout organisé. Exemples : pièces d'une machine.

Collocations enmagasinées

Signification 1	Signification 2
D:visé	D:de véhicule
D:viser	D:majeur
G:délivrer	D:détaché

Occur.	Contexte	Lclx	Expert humain	Explication
1260	. La Société délivre, dans les cas et aux conditions prévus par règlement, l'une ou plusieurs des pièces 1260 suivantes:	Noir	décision heureuse	Inadéquation justifiée Le contexte ne contient, à juste titre, aucune collocation favorable à l'acception recherchée Signific. différente
3943	. La personne qui conduit un véhicule routier doit, à la demande d'un agent de la paix, lui remettre pour examen les pièces 3943 visées à l'article 35.	Noir	décision heureuse	Inadéquation justifiée Le contexte ne contient, à juste titre, aucune collocation favorable à l'acception recherchée Signific. différente
3957	. L'agent doit remettre ces pièces 3957 à leur détenteur dès qu'il les a examinées.	Noir	décision heureuse	Inadéquation justifiée Le contexte ne contient, à juste titre, aucune collocation favorable à l'acception recherchée Signific. différente
15300	. Pour faire le commerce de véhicules routiers mis au rancart, de pièces 15300 provenant de ces véhicules ou de carcasses de véhicules, une personne doit être titulaire d'une licence de recycleur délivrée par la Société, sur paiement des frais fixés et aux conditions et formalités établies par règlement.	Rouge	décision heureuse	Adéquation discontinue
15431	; 2 au propriétaire d'un véhicule routier volé, qui a été démantelé ou	Rouge	décision heureuse	Adéquation disloquée

Occur.	Contexte	Lcix	Expert humain	Explication
	vendu en pièces 15431 détachées par un recycleur, le remboursement du prix de ce véhicule, évalué au moment du vol, sur réclamation en justice;			
15555	: 1 le numéro d'identification de tout véhicule routier qu'il reçoit, la description de ses pièces 15555 majeures et leur identification par le numéro d'identification du véhicule ou par un numéro qui y fait référence;	Rouge	décision heureuse	Adéquation discloquée
15594	; 2 la description et l'identification, par le numéro d'identification du véhicule d'où elle provient, de toute pièce 15594 majeure qu'il reçoit détachée d'un véhicule;	Rouge	décision heureuse	Adéquation discontinue
15615	; 3 la date d'acquisition de tout véhicule routier et de toute pièce 15615 majeure ainsi que les noms et adresse de la personne de qui il les a reçus; 4 la date de vente d'un véhicule ainsi que les nom et adresse de l'acheteur.	Rouge	décision heureuse	Adéquation continue
15650	. pièces 15650 majeures.	Rouge	décision heureuse	Adéquation continue
15656	. Sont considérées comme pièces 15656 majeures d'un véhicule: le moteur, la transmission, le pont arrière, la traction avant, le capot, la calandre, les ailes, les pare-chocs, le panneau de lunette arrière, les panneaux latéraux, le couvercle du coffre et les portes.	Rouge	décision heureuse	Adéquation discontinue
15749	. Le recycleur doit, à la demande d'un agent de la paix ou d'un employé de la Société spécialement désigné à cette fin, lui permettre de vérifier, à toute heure raisonnable, son registre ainsi que les véhicules routiers et les pièces 15749 majeures qu'il a en sa possession.	Rouge	décision heureuse	Adéquation continue
				12

Tableau 42 Le mot pièce(s)

j) Charge

Charge (nom)		
Signification 1	Signification 2	Signification 3
Ce que porte ou peut porter une personne, un animal, un véhicule.. Fardeau, poids	Fonction dont qqn a tout le soin; responsabilité publique. Dignité, emploi, fonction, ministère, office, place, poste.	Ce qui met dans la nécessité de faire des frais, des dépenses. A charge, être à la charge de qqn, vivre à ses dépens.

Collocations enmagasinées

Signification 1	Signification 2	Signification 3
D: comprise	G: exercer	G: être à leur
		G: être à la

Occur	Contexte	Lcix	Expert humain	Explication
172	2. Les dispositions du présent code qui s'appliquent au propriétaire d'un véhicule routier sont également applicables à l'égard de toute personne qui acquiert ou possède un véhicule routier en vertu d'un titre assorti d'une condition ou d'un terme qui lui donne le droit d'en devenir propriétaire, ou en vertu d'un titre qui lui donne le droit d'en jouir comme propriétaire à charge 172 de rendre.	Bleu	Décision heureuse	Mauvaise catégorie
345	: la surface de terrain ou d'un ouvrage d'art dont l'entretien est à la charge 345 d'une municipalité, d'un gouvernement ou de l'un de ses organismes, et sur une partie de laquelle sont aménagées une ou plusieurs chaussées ouvertes à la circulation publique des véhicules routiers et, le cas échéant, une ou plusieurs voies cyclables, à l'exception: 1 des	Rouge	Décision heureuse	Adéquation continue
9652	(...) les personnes suivantes, si elles n'ont pas la citoyenneté canadienne et qu'elles n'exercent aucune entreprise, charge 9652 ou emploi au Québec autre que leur fonction auprès du gouvernement ou de l'organisme qu'elles représentent, peuvent obtenir de la Société, sans examen, un permis de conduire (...)	Noir	Décision heureuse	Inadéquation justifiée Le contexte ne contient, à juste titre, aucune collocation favorable à l'acceptation recherchée Signific. différente
9845	service; 6 les conjoints des personnes visées aux paragraphes 1 à 3 et 5 et leurs enfants majeurs qui sont financièrement à leur charge 9845 et qui résident avec eux. 1986, c. 91, a. 92; 1988, c. 41, a. 87; 1988,	Rouge	Décision heureuse	Adéquation discontinue
25470	244. Les remorques et les semi-remorques qui font partie d'un ensemble de véhicules routiers et dont la masse, charge 25470 comprise, est de 1 300 kg ou plus	Noir	Décision heureuse	Inadéquation justifiée Le contexte ne contient, à juste titre, aucune

				collocation favorable à l'acception recherchée Signific. différente
25483	ou dont la masse, charge comprise, excède la moitié de la masse nette du véhicule automobile qui les tire doivent être munies d'un système de freins indépendant permettant l'application d'une force de freinage sur chaque roue portante. . Véhicule remorqueur. . Le	Noir	Décision heureuse	Inadéquation justifiée Le contexte ne contient, à juste titre, aucune collocation favorable à l'acception recherchée Signific. Différente
				7

k) Porter

Porter (nom)	
Signification 1	Signification 2
avoir sur soi Porter une bague, un insigne.	Se porter comme Se porter garant.

Collocations emmagasinées

Signification 1	Signification 2
D:marque	G:se
D:inscription	D:caution

Occur.	Contexte	Lch	Expert humain	Explication
2905	. Ce certificat doit porter 2905 la marque d'identification de la Société ou la signature d'une personne autorisée par celle-ci. 1986, c.	Rouge	décision heureuse	Adéquation discontinue
3591	32. Une plaque d'immatriculation ne peut porter 3591 une inscription autre que celles déterminées par la Société. . Plaque	Rouge	décision heureuse	Adéquation discontinue
9993	. Tout permis doit porter 9993 la marque d'identification de la Société ou la signature d'une personne autorisée par celle-ci. 1986, c.	Rouge	décision heureuse	Adéquation discontinue
16141	160.1. Une association de commerçants ou de recycleurs peut, selon la forme, les conditions et les modalités établies par règlement du gouvernement, se porter 16141 caution pour ses membres. Elle doit alors déposer une somme en garantie auprès d'une société de fiducie.	Noir	décision heureuse	Inadéquation justifiée Le contexte ne contient, à juste titre, aucune collocation favorable à l'acceptation recherchée Signific. différente

2. Corpus: Code Civil du Québec

a) Droit

Occur.	Contexte	Lclx	Expert humain	Explication
605	. Le greffier ou son adjoint procède à la célébration du mariage selon les règles prescrites par le ministre de la Justice et perçoit des futurs époux, pour le compte du ministre des Finances, les droits 605 fixés par règlement du gouvernement.	Noir	Décision heureuse	Inadéquation justifiée Le contexte ne contient, à juste titre, aucune collocation favorable à l'acception recherchée Signific. différente
845	. Elle laisse subsister les droits 845 et les devoirs des pères et mères à l'égard de leurs enfants.	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
889	. Il est procédé notamment à la liquidation de leurs droits 889 patrimoniaux qui sont alors présumés avoir existé, à moins que les époux ne conviennent de reprendre chacun leurs biens.	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
945	Si un seul des époux était de bonne foi, il peut, à son choix, reprendre ses biens ou demander la liquidation des droits patrimoniaux qui lui résultent du mariage.	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
965	Sous réserve de l'article 386, l'époux de bonne foi a droit 965 aux donations qui lui ont été consenties en considération du mariage.	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
1130	; en prononçant la nullité, il statue sur le droit 1130 de l'époux de bonne foi à des aliments ou à une prestation compensatoire.	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
1152	La nullité du mariage éteint le droit 1152 qu'avaient les époux de se réclamer des aliments, à moins que, sur demande, le tribunal, au moment où il prononce la nullité, n'ordonne à l'un des époux de verser des aliments à l'autre (...)	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
1207	(...) ou, s'il ne peut statuer équitablement sur la question en raison des circonstances, ne réserve	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.

Occur.	Contexte	Lclx	Expert humain	Explication
	le droit 1207 d'en réclamer.			
1212	Le droit 1212 de réclamer des aliments ne peut être réservé que pour une période d'au plus deux ans;	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
1235	il est éteint de plein droit 1235 à l'expiration de cette période.	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
1253	. Lorsque le tribunal a accordé des aliments ou réservé le droit 1253 d'en réclamer,	Rouge	Décision heureuse	Adéquation disloquée
1269	. Lorsque le tribunal a accordé des aliments ou réservé le droit 1253 d'en réclamer, il peut toujours, postérieurement à l'annulation du mariage, déclarer éteint le droit 1269 à des aliments.	Rouge	Décision heureuse	Adéquation disloquée
1299	SECTION I DES DROITS 1299 ET DES DEVOIRS DES ÉPOUX 392.	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
1313	. Les époux ont, en mariage, les mêmes droits 1313 et les mêmes obligations.	Rouge	Décision heureuse	Adéquation discontinue
1346	; il exerce ses droits 1346 civils sous ce nom.	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
1623	. Si les époux ne parviennent pas à s'accorder sur l'exercice de leurs droits 1623 et l'accomplissement de leurs devoirs, les époux ou l'un d'eux peuvent saisir le tribunal qui statuera dans l'intérêt de la famille, après avoir favorisé la conciliation des parties.	Rouge	Décision heureuse	Adéquation discontinue
1805	. L'époux locataire de la résidence familiale ne peut, sans le consentement écrit de son conjoint, sous-louer, céder son droit, 1805 ni mettre fin au bail lorsque le locateur a été avisé, par l'un ou l'autre des époux, du fait	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
1895	. L'époux propriétaire d'un immeuble de moins de cinq logements qui sert, en tout ou en partie, de résidence familiale ne peut, sans le consentement écrit de son conjoint, l'aliéner, le grever d'un droit 1895 réel ni en louer la partie	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.

Occur.	Contexte	Lclx	Expert humain	Explication
	réservée à l'usage de la famille.			
2097	. L'époux autrement titulaire de droits 2097 qui lui confèrent l'usage de la résidence familiale ne peut non plus en disposer sans le consentement de son conjoint.	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
2177	. L'époux qui n'a pas consenti à l'acte pour lequel son consentement était requis peut, sans porter atteinte à ses autres droits, 2177 réclamer des dommages-intérêts de son conjoint ou <u>de toute autre personne</u> qui, par sa faute, lui a causé un préjudice.	Rouge	Décision heureuse	Adéquation disloquée
2255	. L'attribution lie le locateur dès que le jugement lui est signifié et libère, pour l'avenir, le locataire originaire des droits 2255 et <u>obligations</u> résultant du bail.	Rouge	Décision heureuse	Adéquation continue
2320	. Il peut également attribuer à l'époux auquel il <u>accorde</u> la garde d'un enfant un droit 2320 d'usage de la résidence familiale.	Rouge	Décision heureuse	Adéquation disloquée
2355	. L'attribution du droit 2355 d'usage ou <u>de propriété</u> se fait, à défaut d'accord entre les parties, aux conditions que le tribunal détermine et notamment, s'il y a lieu, moyennant une soulte payable au comptant ou par versements.	Rouge	Décision heureuse	Adéquation discontinue
2416	. L'attribution judiciaire d'un droit 2416 <u>de propriété</u> est assujettie aux dispositions relatives à la vente.	Rouge	Décision heureuse	Adéquation continue
2433	. Le jugement qui attribue un droit 2433 d'usage ou <u>de propriété</u> équivaut à titre et en a tous les effets.	Rouge	Décision heureuse	Adéquation discontinue
2482	Le mariage emporte constitution d'un patrimoine familial formé de certains biens des époux sans égard à celui des deux qui détient un droit 2482 <u>de propriété</u> sur ces biens.	Rouge	Décision heureuse	Adéquation continue
2514	: les résidences de la famille ou les droits 2514 qui en confèrent l'usage,	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.

Occur.	Contexte	Lcix	Expert humain	Explication
	les meubles qui les garnissent ou les orment et (...)			
2548	(...) qui servent à l'usage du ménage, les véhicules automobiles utilisés pour les déplacements de la famille et les droits 2548 accumulés durant le mariage au titre d'un régime de retraite.	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
2615	. Sont toutefois exclus du patrimoine familial, si la dissolution du mariage résulte du décès, les gains visés au deuxième alinéa ainsi que les droits 2615 accumulés au titre d'un régime de retraite régi ou établi par une loi qui (...)	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
2636	(...) <u>accorde</u> au conjoint survivant le droit 2636 à des prestations de décès.	Rouge	Décision heureuse	Adéquation discontinue
3603	. Les époux ne peuvent renoncer, par leur contrat de mariage ou autrement, à leurs droits 3603 dans le patrimoine familial.	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
3684	. La renonciation doit être inscrite au registre des droits 3684 personnels et réels mobiliers.	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
3707	. A défaut d'inscription dans un délai d'un an à compter du jour de l'ouverture du droit 3707 au partage, l'époux renonçant est réputé avoir accepté.	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
3815	. Le partage des droits 3815 accumulés par l'un des époux au titre d'un régime de retraite régi ou établi par une loi est effectué conformément, s'il en existe, aux règles d'évaluation et de dévolution édictées par cette loi.	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
3858	Toutefois, le partage de ces droits 3858 ne peut en aucun cas avoir pour effet (...)	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
3874	(...) de priver le titulaire original de ces droits 3874 de plus de la moitié (...)	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
3885	(...) de la valeur totale des droits	Noir	Décision	Inadéquation fautive:

Occur.	Contexte	Lcblx	Expert humain	Explication
	3885 qu'il a accumulés avant ou pendant le mariage, (...)		malheureuse	Carence de la base.
3901	(...) ni de conférer au bénéficiaire du droit 3901 (...)	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
3906	(...) au partage plus de droits 3906 qu'en possède, (...)	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
3920	(...) en vertu de son régime, le titulaire original de ces droits. 3920	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
3934	Entre les époux ou pour leur bénéfice, et nonobstant toute disposition contraire, ces droits, 3934 ainsi que ceux accumulés au titre d'un autre régime de retraite, sont cessibles et saisissables pour le partage du patrimoine familial.	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
4066	. Lorsque le droit 4066 à la prestation compensatoire est fondé sur la collaboration régulière de l'époux à une entreprise, que cette entreprise ait trait à un bien ou à un service et qu'elle soit ou non à caractère commercial, (...)	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
4210	. Celui-ci peut également déterminer, le cas échéant, les modalités du paiement et ordonner que la prestation soit payée au comptant ou par versements ou qu'elle soit payée par l'attribution de droits 4210 dans certains biens.	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
4228	Si le tribunal attribue à l'un des époux ou au conjoint survivant un droit 4228 sur la résidence familiale, (...)	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
4245	(...) sur les meubles qui servent à l'usage du ménage ou des droits 4245 accumulés au titre d'un régime de retraite, les dispositions des sections II et III sont applicables.	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
4916	. Un avis de tout contrat de mariage doit être inscrit au registre des droits 4916 personnels et réels mobiliers sur la réquisition du notaire instrumentant.	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
4931	.-De l'exercice des droits 4931 et	Noir	Décision	Inadéquation fautive:

Occur.	Contexte	Lcix	Expert humain	Explication
	pouvoirs résultant du régime matrimonial 443.		malheureuse	Carence de la base.
4955	Chacun des époux peut donner à l'autre mandat de le représenter dans l'exercice des droits 4955 et pouvoirs que le régime matrimonial lui attribue.	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
5050	. Ce mandat cesse de plein droit 5050 dès que le conjoint est pourvu d'un tuteur ou d'un curateur.	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
5337	; 4 Les droits 5337 ou avantages qui lui échoient à titre de titulaire subrogé ou à titre de bénéficiaire déterminé d'un contrat ou d'un régime de retraite, d'une autre rente ou d'une assurance de personnes;	Rouge	Décision heureuse	Adéquation disloquée
5572	. Le droit 5572 d'un époux à une pension alimentaire, à une pension d'invalidité ou à quelque autre avantage de même nature, lui reste propre, mais sont acquêts tous les avantages pécuniaires qui en proviennent et qui sont échus ou perçus au cours du régime ou qui sont payables, à son décès, à ses héritiers et ayants cause.	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
5662	. Sont également propres à l'époux le droit 5662 de réclamer des dommages-intérêts et l'indemnité reçue en réparation d'un préjudice moral ou corporel	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
5685	La même règle s'applique au droit 5685 et à l'indemnité découlant d'un contrat d'assurance ou de tout autre régime d'indemnisation, mais aucune récompense n'est due en raison des primes ou sommes payées avec les acquêts. 455.	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
5839	. Les valeurs mobilières acquises par suite de l'exercice d'un droit 5839 de souscription	Rouge	Décision heureuse	Adéquation discontinuée
5847	. Les valeurs mobilières acquises par suite de l'exercice d'un droit 5839 de souscription ou de préemption ou	Rouge	Décision heureuse	Adéquation discontinuée

Occur.	Contexte	Lcbx	Expert humain	Explication
	autre droit 5847 semblable que confèrent des valeurs propres à l'un des époux lui restent également propres, sauf récompense s'il y a lieu.			
5941	. Les droits 5941 <u>de propriété</u> intellectuelle et industrielle sont propres, mais sont acquêts tous les fruits et revenus qui en proviennent et qui sont perçus ou échus au cours du régime.	Rouge	Décision heureuse	Adéquation continue
6135	. La restriction au droit 6135 de disposer ne limite pas le droit 6142 <u>d'un époux de désigner un tiers comme bénéficiaire ou titulaire subrogé d'une assurance de personnes,</u>	Rouge	Décision heureuse	Adéquation disloquée
6142	(...) ne limite pas le droit 6142 <u>d'un époux de désigner un tiers comme bénéficiaire ou titulaire subrogé d'une assurance de personnes,</u> d'une pension de retraite ou autre rente, sous réserve de l'application des règles relatives au patrimoine familial.	Rouge	Décision heureuse	Adéquation disloquée
6557	. La renonciation doit être inscrite au registre des droits 6557 personnels et réels mobiliers; à défaut d'inscription dans un délai d'un an à compter du jour de la dissolution, l'époux est réputé avoir accepté.	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
6597	. Si l'époux renonce, la part à laquelle il aurait eu droit 6597 dans les acquêts de son conjoint reste acquise à ce dernier.	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
6621	Toutefois, les créanciers de l'époux qui renonce au préjudice de leurs droits 6621 peuvent demander au tribunal de déclarer que la renonciation leur est inopposable et accepter la part des acquêts du conjoint de leur débiteur au lieu et place de ce dernier.	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
6872	. Lorsqu' un époux décède alors qu'il	Noir	Décision	Inadéquation fautive:

Occur.	Contexte	Lcbx	Expert humain	Explication
	était encore en droit 6872 de renoncer, ses héritiers ont, à compter du décès, un nouveau délai d'un an pour faire inscrire leur renonciation.		malheureuse	Carence de la base.
7342	. La dissolution du régime ne peut préjudicier, avant le partage, aux droits 7342 des créanciers antérieurs sur l'intégralité du patrimoine de leur débiteur.	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
7420	. Chaque époux conserve alors un recours contre son conjoint pour les sommes auxquelles il aurait eu droit 7420 si la créance avait été payée avant le partage.	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
7531	Le bien sur lequel aucun des époux ne peut justifier de son droit 7531 exclusif de propriété	Rouge	Décision heureuse	Adéquation discontinue
7686	Ils peuvent aussi se pourvoir contre la séparation de biens prononcée ou exécutée en fraude de leurs droits. 7686	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
7704	. La dissolution du régime matrimonial opérée par la séparation de biens ne donne pas ouverture aux droits 7704 de survie, sauf stipulation contraire dans le contrat de mariage.	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
8015	Il entre dans la mission du tribunal de conseiller les époux de favoriser leur conciliation et de veiller aux intérêts des enfants et au respect de leurs droits. 8015 2.-De l	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
8441	droits 8441 de survie, sauf stipulation contraire dans le contrat de mariage. 510. La séparation de corps ne rend pas caduques les donations	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
8677	. Elle laisse subsister les droits 8677 et les devoirs des père et mère à l'égard de leurs enfants.	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
8726	. Au moment où il prononce la séparation de corps ou postérieurement, le tribunal statue sur la garde, l'entretien et l'éducation	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.

Occur.	Contexte	Lcix	Expert humain	Explication
	des enfants, dans l'intérêt de ceux-ci et le respect de leurs droits , 8726 en tenant compte, s'il y a lieu, des accords conclus entre les époux.			
9035	. Tous les enfants dont la filiation est établie ont les mêmes droits 9035 <u>et</u> les mêmes obligations , quelles que soient les circonstances de leur naissance. CHAPITRE PREMIER DE LA FILIATION PAR LE SANG SECTION I DES PREUVES DE LA FILIATION 1.- Du titre et de la	Rouge	Décision heureuse	Adéquation discontinuée
9912	. Le décès du père présumé ou de la mère avant l'expiration du délai prévu pour le désaveu ou la contestation d'état n'éteint pas le droit 9912 d'action	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
9917	. Toutefois, ce droit 9917 doit être exercé par les héritiers dans l'année qui suit le décès.	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
10168	. L'un des descendants de cette personne peut également se prévaloir de ce droit 10168 si le fait d'être privé des renseignements qu'il requiert risque de causer un préjudice grave à sa santé ou à celle de l'un de ses proches.	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
10719	. Le consentement à l'adoption entraîne de plein droit , 10719 jusqu'à l'ordonnance de placement, délégation de l'autorité parentale à la <u>personne</u> à qui l'enfant est remis.	Rouge	Décision heureuse	Adéquation disloquée
11472	; elle permet à l'enfant, pendant la durée du placement, d'exercer ses droits 11472 civils sous les nom et prénoms choisis par l'adoptant, lesquels sont constatés dans l'ordonnance.	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
11972	. L'adoption fait naître les mêmes droits 11972 <u>et</u> obligations que la filiation par le sang.	Rouge	Décision heureuse	Adéquation continue
12027	le tuteur, s'il en existe, perd ses droits 12027 <u>et</u> est libéré de ses	Rouge	Décision heureuse	Adéquation disloquée

Occur.	Contexte	Léix	Expert humain	Explication
	devoirs à l'endroit de l'adopté, sauf l' <u>obligation</u> de rendre compte.			
12235	. L'adopté majeur ou l'adopté mineur de quatorze ans et plus a le droit 12235 <u>d'obtenir</u> les renseignements lui permettant de retrouver ses parents, si ces derniers y ont préalablement consenti.	Rouge	Décision heureuse	Adéquation continue
12284	. L'adopté mineur de moins de quatorze ans a également le droit 12284 <u>d'obtenir</u> les renseignements lui permettant de retrouver ses parents, si ces derniers, ainsi que ses parents adoptifs, y ont préalablement consenti.	Rouge	Décision heureuse	Adéquation continue
12392	. L'un des proches parents de l'adopté peut également se prévaloir de ce droit 12392 si le fait d'être privé des renseignements qu'il requiert risque de causer un préjudice grave à sa santé ou à celle de l'un de ses proches.	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
12626	. Afin de maintenir la valeur monétaire réelle de la créance qui résulte du jugement accordant des aliments, ceux-ci, s'ils sont payables sous forme de pension, sont indexés de plein droit , 12626 au 1er janvier de chaque année, suivant l'indice annuel des rentes établi conformément à l'article 119 de la Loi sur le régime de rentes du Québec.	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
13051	. Les père et mère ont, à l'égard de leur enfant, le droit 13051 et le devoir de garde, de surveillance et d'éducation.	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
13252	. Que la garde de l'enfant ait été confiée à l'un des parents ou à une tierce personne, quelles qu'en soient les raisons, les père et mère conservent le droit 13252 de surveiller son	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.

b) Condition

Occur.	Contexte	Lclx	Expert humain	Explication
763	. Le mariage qui n'est pas célébré suivant les prescriptions du présent titre et suivant les conditions 763 nécessaires à sa fonction peut être frappé de nullité à la demande de toute personne intéressée, sauf au tribunal à juger suivant les circonstances.	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
2039	. Si une déclaration de résidence familiale a été préalablement inscrite contre l'immeuble, le conjoint qui n'a pas donné son consentement à l'acte d'aliénation peut exiger de l'acquéreur qu'il lui consente un bail des lieux déjà occupés à des fins d'habitation, aux conditions 2039 régissant le bail d'un logement;	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
2049	sous la même condition , 2049 celui qui n'a pas donné son consentement à l'acte de location peut, s'il ne l'a pas ratifié, en demander la nullité.	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
2371	. L'attribution du droit d'usage ou de propriété se fait, à défaut d'accord entre les parties, aux conditions 2371 que le tribunal détermine et notamment, s'il y a lieu, moyennant une soulte payable au comptant ou par versements.	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
5018	. Il fixe les modalités et les conditions 5018 d'exercice des pouvoirs conférés.	Rouge	Décision heureuse	Adéquation discontinue
9627	. Si l'enfant a déjà une autre filiation établie soit par un titre, soit par la possession d'état, soit par l'effet de la présomption de paternité, l'action en réclamation d'état ne peut être exercée qu'à la condition 9627 d'être jointe à une action en contestation de l'état ainsi établi.	Rouge	Décision heureuse	Adéquation disloquée
10206	. CHAPITRE DEUXIEME DE L'ADOPTION SECTION I DES CONDITIONS 10206 DE L'ADOPTION 1.-Dispositions	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.

Occur.	Contexte	Lcix	Expert humain	Explication
	générales 543.			
10228	L'adoption ne peut avoir lieu que dans l'intérêt de l'enfant et aux conditions 10228 prévues par la loi.	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
11050	.-Des conditions 11050 particulières à l'adoption d'un enfant domicilié hors du Québec 563.	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
11089	Toute personne domiciliée au Québec qui veut adopter un enfant domicilié hors du Québec doit préalablement faire l'objet d'une évaluation psychosociale effectuée dans les conditions prévues par la Loi sur la protection de la jeunesse.	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
11116	. Les démarches en vue de l'adoption sont effectuées soit par l'adoptant, dans les conditions 11116 prévues par la Loi sur la protection de la jeunesse, soit, à la demande de l'adoptant, par le ministre de la Santé et des Services sociaux ou par un organisme agréé en vertu de la même loi. -	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
11330	. Avant de prononcer l'ordonnance de placement, le tribunal s'assure que les conditions 11330 de l'adoption ont été remplies et, notamment, que les consentements requis ont été valablement donnés.	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.
12936	Le débiteur de qui on réclame des arrérages peut opposer un changement de sa condition 12936 ou celle de son créancier survenu depuis le jugement et être libéré de tout ou partie de leur paiement.	Noir	Décision malheureuse	Inadéquation fautive: Carence de la base.

E. STRUCTURE DES FICHIERS

Numéro	Nom du champ	Type	Dimension
1	CODE MOT	Numérique	7
2	MOT	Caractère	30
3	AMBIGU	Logique	1

Tableau 43 FORMES.DIC

Numéro	Nom du champ	Type	Dimension
1	CODE FORME	Numérique	7
2	FORME	Caractère	30
3	CODE LEMME	Numérique	7
4	LEMME	Caractère	30

Tableau 44 LIENS.LIE

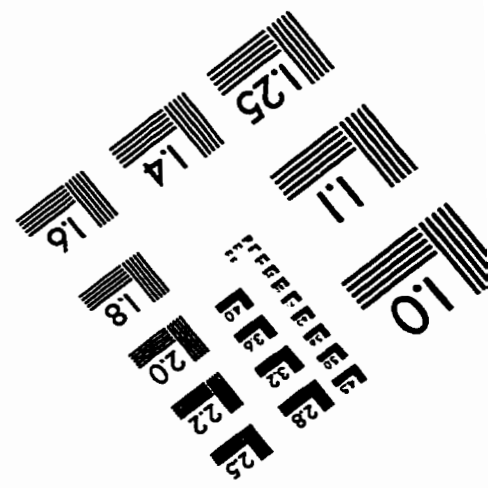
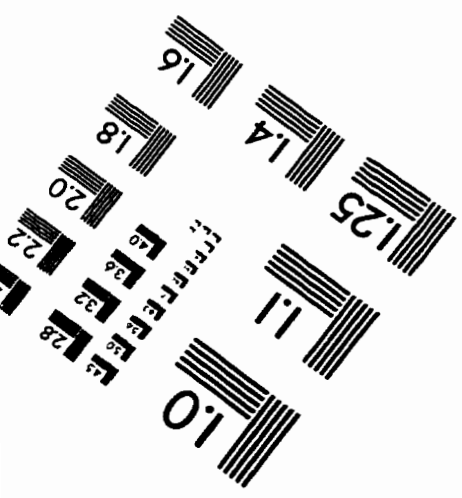
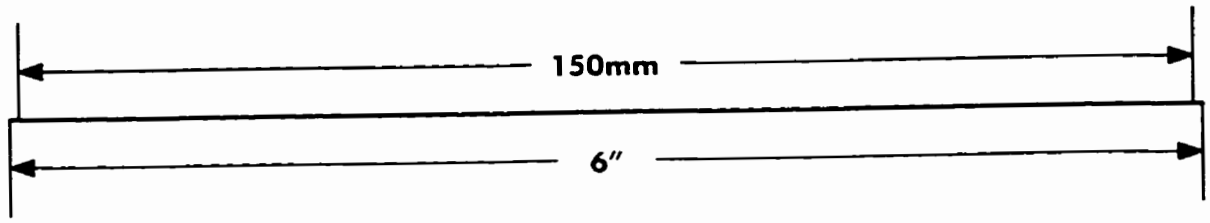
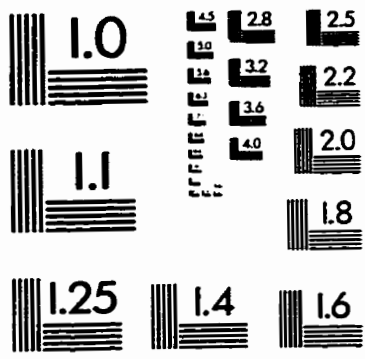
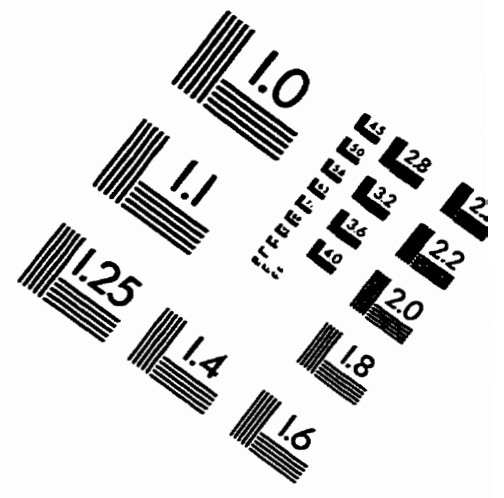
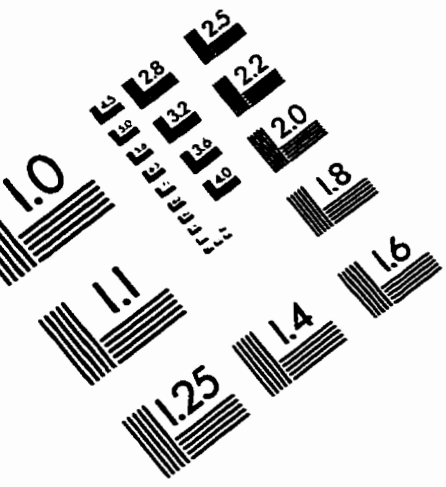
Numéro	Nom du champ	Type	Dimension
1	CODE LEM	Numérique	7
2	LEMME	Caractère	30
3	CD CATEG	Numérique	2
4	CATEGORIE	Caractère	20
5	NOTES	Caractère	125
6	ETYMOLOGIE	Caractère	40
7	DOMAINE	Caractère	10

Tableau 45 LEMMES.DC2

Numéro	Nom du champ	Type	Dimension
1	MOT DIC	Caractère	30
2	CODE LEMME	Numérique	7
3	NUM SENS	Numérique	7
4	SENS	Caractère	80
5	DEFINITION	Mémo	10

Tableau 46 ACCEPT.ACC

IMAGE EVALUATION TEST TARGET (QA-3)



APPLIED IMAGE, Inc
 1653 East Main Street
 Rochester, NY 14609 USA
 Phone: 716/482-0300
 Fax: 716/288-5989

© 1993, Applied Image, Inc., All Rights Reserved